NATIONAL QUALITY FORUM

Memo

TO: Consensus Standards Approval Committee (CSAC)
FR: Pediatric Performance Measures Project Team
RE: Pediatric Performance Measures, Phase II
DA: June 30, 2017

CSAC Action Required

The CSAC will review recommendations from the Pediatric Performance Measures project at its July 11-12, 2017, meeting and vote whether to uphold the recommendations from the Committee. This memo includes a summary of the project, recommended measures, and themes identified from and responses to the member and public. NQF member voting on these recommended measures closed on June 26, 2017.

Accompanying this memo are the following documents:

1. **Pediatric Performance Measures 2016-2017 Draft Report.** The draft report has been updated to reflect the changes made following the Standing Committee discussion of member and public comments. The complete draft report and supplemental materials are available on the project page.
2. **Comment table.** Staff has identified themes within the comments received. This table lists all 11 comments received and the NQF, measure developer, and Standing Committee responses.

Background

Approximately 74 million children under 18 years of age live in the United States, representing 23.3 percent of the population.[i] Understanding the health-related needs of children and adolescents is essential for developing measures to improve the quality of care for the pediatric population. Currently, the NQF portfolio includes 117 NQF-endorsed measures that include the pediatric population: 55 NQF-endorsed measures specific to the pediatric population and 62 NQF-endorsed measures including the pediatric and adult population. The measures pertain to a range of clinical and crosscutting areas, including cardiovascular surgery, pulmonary care, cancer, perinatal care, health and well-being, and safety. Currently, many of these measures are used in public and/or private accountability and quality improvement programs, such as the Center for Medicare and Medicaid Services (CMS) Child Core Set.

For this project, the 23-member Standing Committee evaluated 11 newly submitted measures against NQF's standard evaluation criteria. The Committee recommended four measures for endorsement and did not recommend seven measures.

Draft Report

The Pediatric Performance Measures 2016-2017 Draft Report presents the results of the evaluation of the 11 measures considered under the Consensus Development Process (CDP). Four are recommended for endorsement and seven were not recommended.

The measures were evaluated against the 2015 version of the measure evaluation criteria.

	New	Total
Measures under consideration	12	12
Measures recommended for endorsement	4	4
Measures where consensus is not yet reached	0	0
Measures not recommended for endorsement	7	7
Measures withdrawn from consideration	1	1
Reasons for not recommending	Importance – 4 Scientific Acceptability – 3 Overall – 0 Competing Measure – 0	

CSAC Action Required

Pursuant to the CDP, the CSAC is asked to consider endorsement of four candidate consensus standards.

Pediatric Performance Measures Recommended for Endorsement:

- 3136: GAPPS: Rate of preventable adverse events per 1,000 (Center of Excellence for Pediatric Quality Measurement, Boston Children's Hospital)
    - Overall Suitability for Endorsement: Y-14; N-8
- 3153: Continuity of Primary Care for Children with Medical Complexity (Seattle Children's Research Institute)
    - Overall Suitability for Endorsement: Y-17; N-5
- 3154: Informed Coverage (Children's Hospital of Philadelphia)
    - Standing Committee Recommendation for Endorsement: Y-13, N-4
- 3166: Antibiotic Prophylaxis Among Children with Sickle Cell Anemia (QMETRIC, University of Michigan)
    - Overall Suitability for Endorsement: Y-23; N-0

Pediatric Performance Measures Not Recommended (See Appendix A for the Committee's votes and rationale):

- 2816: Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma (Collaboration for the Advancement of Pediatric Quality Measures, University Hospitals Cleveland Medical Center)
- 3189: Rate of Emergency Department Visit Use for Children Managed for Identifiable Asthma - Visits per 100 Child years (Collaboration for the Advancement of Pediatric Quality Measures, University Hospitals Cleveland Medical Center)

- 3219: Anticipatory Guidance and Parental Education (Child and Adolescent Health Measurement Initiative, Johns Hopkins Bloomberg School of Public Health (CAHMI))
- 3220: Ask About Parental Concerns (CAHMI)
- 3221: Family Centered Care (CAHMI)
- 3222: Assessment of Family Alcohol Use, Substance Abuse and Safety (CAHMI)
- 3223: Assessment of Family Psychosocial Screening (CAHMI)

Comments and Their Disposition

NQF received 11 comments from four member organizations pertaining to the measures under consideration.

A table of comments submitted during the comment period, with the responses to each comment and the actions taken by the Standing Committee and measure developers, is posted to the Pediatric Performance Measures project page under the Member and Public Comment section along with the measure submission forms.

Comment Themes and Committee Responses

Two major themes were identified in the post-evaluation comments, as follow:

1. Support for Committee recommendations
2. Gaps for future measure development

In addition, one measure, *#3136: GAPPS: Rate of preventable adverse events per 1,000 patient-days among pediatric inpatients*, received specific comments requiring a developer response and Committee discussion.

Theme 1 - Support for Committee recommendations

Five comments offered support for the Committee's endorsement recommendations, both for decisions to recommend endorsement and not to recommend endorsement. These comments provided support for the Committee's recommendations on measures #3153, #3166, #3220, and #3221. Commenters agreed with the Committee's decision not to recommend *#3220: Ask About Parental Concerns* and *#3221: Family Centered Care*, noting that despite the clear importance of these topics, there is "difficulty in attributing outcomes within these areas to specific providers and experiences." Two commenters supported the Committee's decision to recommend *#3166: Antibiotic Prophylaxis Among Children with Sickle Cell Anemia*, and one commenter supported the Committee's recommendation to endorse measure *#3153: Continuity of Primary Care for Children with Medical Complexity*.

> **Committee Response:** Thank you for providing these comments.

Theme 2 – Gaps for future measure development

Commenters identified several measure gap areas for consideration by the Committee. Specifically, they suggested gaps could be addressed by the following clinic-/system-level measure concepts:

- The identification of a team to work together to plan and test improvements in providers' ability to elicit parental strengths and needs within a practice site.
- Defining parental strengths and needs within a practice site.
- Integrating tools such as process flows, prompts, and reminders into practice flow to support the engagement of parents.
- Clinic-/systems-level measures that offer more specificity about appropriate antibiotic prophylaxis in children with sickle cell anemia.

Committee Response: Thank you for providing these comments. These gaps have been added to the measure gaps list.

Measure-Specific Comments

#3136: GAPPS: Rate of preventable adverse events per 1,000 patient-days among pediatric inpatients

Two comments focused on measure *#3136: GAPPS: Rate of preventable adverse events per 1,000 patient-days among pediatric inpatients*. The developer addressed each concern separately.

One commenter submitted questions and suggested updates intended to clarify automated triggers to increase the specificity and clarity of the measure specifications. A second commenter did not support the endorsement of this measure. The commenter noted that implementing the trigger tool might be difficult and require excessive resources, and suggested that the tool lacks validity in identifying adverse events.

Developer Response to the Academy of Pediatrics:

- Trigger: Consider rewording to "Hepatotoxic medications and RISING liver enzymes (AST, ALT)"

  Thank you for the suggestion. A consideration here is that if there were not previous hepatic enzyme measurements and the first measurement showed elevated enzymes, this would need to be investigated. If this were written to only include those that are rising (therefore requiring a previous lower value), the process might miss a possible hepatoxic injury. Therefore our preference is to retain the language as "elevated."

- Please explain how "Physician orders: Abrupt medication stop" is defined in the automated trigger tool? Most medication stops are abrupt (with rare exceptions like steroid weans or PCA infusions)

  The definition in our Manual of Operations reads as follows, "An abrupt medication stop is best described as an unexpected stop or deviation from typical ordering practice (e.g., discontinuation of a recently started medication)." Since this type of clinical decision making may be challenging to automate, it is not recommended for electronic trigger review.

- Please define “Transfer to higher level of care” more specifically. Many hospitals have observation units where most patients go home but some patients are admitted to the floor (higher level of care) after a specified time.

  The definition from our Manual of Operations reads as follows, “All transfers from an acute care area to an intensive care unit or intermediate care unit (“step-up unit”) should be considered a trigger.” Therefore the scenario presented in the comment would not meet the defined criteria.

- Consider changing Pressure ulcer documentation to “>= Stage 2” instead of just stage 2.

  Thank you for identifying this discrepancy. This was a typographical error and should read as you suggest. During our testing, the reviewers were instructed to investigate exactly as is suggested by the comment, meaning all pressure injuries Stage 2 and higher and unstageable. We will edit the relevant documents to reflect this change.

- Many places will start patients on laxatives simultaneously with opioids, but patients will still get constipated. Would this qualify as a trigger, or is it only a trigger if laxatives are started after (e.g. >=24 hours after) opioids are started? Latter would be more specific, less sensitive.

  Thank you for pointing out this ambiguity. We agree that excluding cases where laxatives are introduced concurrently (<24 hours after) with opioids is reasonable. The trigger is looking at cases where laxatives were given subsequent to the initial prescription of opioids (>=24 hours after). We will edit the relevant documents to reflect this change.

- Consider adding “positive coagulase-negative staphylococcus species blood culture” as a trigger for review; per algorithm, it should have a higher than 10% rate of being a true contaminant (i.e., an adverse event).

  Thank you for this comment. Since we currently look at a more broadly based trigger (positive blood culture 48 hours after admission), all of the occurrences of the suggested trigger would be included in the trigger as written. We hesitate to insert a new trigger into the recently reviewed tool at this stage.

- Please clarify the denominator of whether a partial day counts as a day. For example, is 1.5 days = 2 days or 1.5 days? What is the start and stop time for determining LOS duration (e.g. start of: time of arrival to floor, time of admission from ED; end of: time of discharge order, time of leaving hospital?)

  Length of stay is calculated as the number of days (discharge date minus admit date). For example, a patient who arrives at 4am on May 17th and is discharged at 4pm on May 18th has a length of stay of 1.0 day. However, a patient who arrives at 10pm on May 17th and is discharged at 10am on May 19th has a length of stay of 2.0 days. Start and stop times were not used to determine length of stay duration, only admit and discharge date.

- Step 2: Line 4. Please describe whether the unit of study (whether entire hospital, division, etc.) should remain stable over time.

  Thank you for the opportunity to clarify. We would suggest that the unit of interest remain stable over time.

One commenter did not support the endorsement of this measure. The commenter noted that implementing the trigger tool might be difficult and require excessive resources, and suggested that the tool lacks validity in identifying adverse events.

Developer Response to Dr. Austin of Armstrong Institute for Patient Safety and Quality at Johns Hopkins University:

We would like to thank Dr. Austin for his comments. As the measure is implemented, the resource burden, while not trivial, should be manageable while providing a great deal of benefit in terms of increased safety events identified. The primary reviewer, typically an experienced nurse, is asked to perform chart review quarterly on 60 patient records per quarter with a limit of 30 minutes per chart. This would represent a total of 30 hours per quarter or 10 hours per month or 2.5 hours a week. The secondary reviewer, typically a physician, reviews the primary reviewer's findings. Assuming a high rate of harm or 33 events per 100 admissions, this would be 20 events to review each quarter. During validation testing, our physician team required on average 4 minutes per chart to review. Therefore, the typical time burden on the secondary reviewer would be approximately 80 minutes per quarter. Based on the frequency of events and the resources required, it is our view that the benefit of this modest resource requirement would far outweigh the burden.

In regards to validity, we developed the draft trigger tool used in the GAPPS measure through the RAND/UCLA Appropriateness Method, which is a modified Delphi process.(1–3) We first compiled a set of 78 candidate triggers from a literature review of existing pediatric and adult trigger tools and input from trigger tool experts.(4–6) We then recruited nine panelists from national pediatric and patient safety organizations and asked them to rate separately the validity and feasibility of the candidate triggers on a nine-point scale (where 1 is the least valid/feasible and 9 is the most valid/feasible). A trigger was considered valid if it was judged to be reasonably likely to identify an underlying AE, indicating that harm potentially occurred. A trigger was considered feasible if it was judged likely to be accurately and consistently documented in either paper or electronic medical records as part of patient care at a wide range of hospitals, from smaller community sites to larger tertiary care centers. Applying the RAND/UCLA Appropriateness Method, we accepted triggers that had both median validity and feasibility ratings greater than or equal to seven. This approach resulted in inclusion of 54 of the initial 78 candidate triggers in the draft GAPPS trigger list.

We focused our validity testing on evaluation of how accurately and completely "typical reviewers" (i.e., clinicians who are trained in GAPPS methodology but not necessarily trigger tool experts) were able to identify preventable AEs using the measure as compared to expert reviewers. The expert reviewers had extensive experience with using trigger tools for preventable AE identification and consequently were most likely to identify preventable AEs accurately and completely. To evaluate the validity of the GAPPS measure, we assessed the performance of the National Field Test hospitals' internal reviewers relative to the performance of external expert reviewers in applying the measure (as documented in our NQF submission materials).

REFERENCES

1. Fitch K, Bernstein S, Aguilar MD, Burnand B, LaCalle JR, Lázaro P, et al. The RAND/UCLA Appropriateness Method User's Manual. Santa Monica, CA: RAND; 2001.

2. Brown B. DELPHI PROCESS: A Methodology Used for the Elicitation of Opinions of Experts. Rand Corp. 1968 Sep;1–14.

3. Sweidan M, Williamson M, Reeve JF, Harvey K, O'Neill JA, Schattner P, et al. Identification of features of electronic prescribing systems to support quality and safety in primary care using a modified Delphi process. BMC Med Inform Decis Mak. 2010 Apr 15;10(1):21.

4. Stockwell D, Bisarya H, Classen D, Kirkendall E, Landrigan C, Lemon V, et al. A trigger tool to detect harm in pediatric inpatient settings. Pediatrics. 2015;

5. Griffin FA, Resar RK. IHI Global Trigger Tool for Measuring Adverse Events (Second Edition). Institute for Healthcare Improvement; 2009. (IHI Innovation Series white paper).

6. Kirkendall ES, Kloppenborg E, Papp J, White D, Frese C, Hacker D, et al. Measuring adverse events and levels of harm in pediatric inpatients with the Global Trigger Tool. Pediatrics. 2012 Nov;130(5):e1206-1214.

Committee Response:

Thank you for providing these comments on measure #3136. The Committee discussed the measure specifications and validity during the in-person meeting. The Committee did note that that the highest possible score for reliability was a moderate, since the measure is tested at the data-element level only; the highest possible score for validity also is moderate, since validity testing is patient-level data element. Overall, the Committee determined that the measure, as specified and tested, offered sufficient validity for endorsement and did not wish to reconsider the measure.

Meeting Follow-Up Issues

Consensus Not Reached Measure

3154: Informed Coverage

During the in-person meeting, Committee members agreed this was an important outcome to assess, but they were concerned about the measure's ability to discern differences among states due to the overlap of the 95% confidence intervals of the performance scores provided for score-level reliability testing. The Committee believed that the measure would be a useful self-assessment tool for states to improve their coverage rates, but questioned whether this measure could be used for accountability purposes. The Committee did not reach consensus on Reliability (1-H; 11-M; 9-L; 3-I).

A memo from the developer responding to the issue of overlap, as well as other questions brought up by the Committee during the in-person meeting, was provided to the Committee prior to the post-comment call. With respect to the issue of the overlapping performance scores, the developer summarized the graph (previously provided) as follows:

- 24 of 43 states (55.8%) can be distinguished from more than 1/2 of the other states;
- 11 (25.6%) states can be distinguished from more than 2/3 of the other states;
- At each end of the spectrum (high and low performers), 3 of 43 states (7.0%) and 3 of 43 states (7.0%), respectively, can be distinguished from 3/4 of the other states.

One NQF member, the American Academy of Pediatrics (AAP), commented on #3154. AAP agrees with the intent of the measure to more accurately capture the continuity of coverage in the Medicaid program, but recommended that this measure be further validated and re-evaluated for endorsement in the future.

Developer Response:

We appreciate that the AAP agrees with the intent of our measure to more accurately capture the continuity of coverage in the Medicaid program so that states can improve coverage. The AAP suggested that our measure "requires assumptions that may not be universally accepted," without telling us which assumptions are objectionable. We would point out that with our assumptions, our results were carefully validated against the gold standard ACS (American Community Survey). Our results, in both development and validation, were superior to the current metrics of Continuity Ratio (Ku et al.) and Duration (currently used by CMS). Informed Coverage had better correlation with the ACS and less error deviation than the other metrics. See Validity Testing, Section 2b2.3, Table 2: Pearson Correlations. Also, see Validity Testing, Section 2b2.3, Table 3: Median Absolute Errors.

Committee Response:

During the call, Committee members discussed their previous concerns about children who were on the cusp of income eligibility. The developer explained that while it did not have access to the incomes of individuals, an analysis was conducted using the average income on a state level. The developer noted that the analysis showed that the metric is stable across income levels across states. The developer further noted that rates of reenrollment are largely policy-driven (i.e., how easy or difficult it is to reenroll) rather than by income, and that improvements in performance can be made by making it easier to re-enroll.

Also during the post-comment call, a Committee member inquired about the intended use of the measure. The developer noted that this could be used by states to measure improvement after implementing initiatives. The measure also helps to identify states that report rates much lower or much higher than other states for closer examination. Currently, the measure relies on presumptive eligibility for Medicaid; the Committee and developer agreed that if eligibility changes, the measure would need revision. After its discussion, the Committee re-voted on the Reliability criterion, which had not achieved consensus during the in-person meeting; the measure passed this criterion. Following that vote, Committee members briefly discussed unintended consequences should the measure be used for rewards or penalties. They ultimately agreed this measure is no different than any other measure that has intrinsic error, and that as long as the range of error is clearly defined, the measure can be useful. The Committee voted on an overall recommendation for endorsement, and agreed the measure should be recommended.

Requests for Re-consideration

3189: Rate of Emergency Department Visit Use for Children Managed for Identifiable Asthma: Visits per 100 Child-years

During the in-person meeting, the Committee concluded that the testing information was insufficient to meet NQF's minimum standards, and the measure did not pass Reliability (N/A-H; 1-M; 4-L; 18-I). No comments were received specific to this measure during the post-meeting commenting period. The developer submitted a request for reconsideration of #3189.

Developer Rationale for Reconsideration:

"At the in-person meeting, measure #3189 passed on Evidence and Gap, and was voted insufficient for Reliability. In general, the sense of the group [the Committee] at the in-person meeting was that measure #3189 is a very viable measure, but having to conform to the NQF procedure, the group required a little bit more data, which is provided herein:

1) Reliability
2) Inclusion/Exclusion
3) Pharmacy Data
4) Race Disparities
5) Data Element Validity"

Committee Response:

The Committee reviewed the new materials prior to the call, and after discussion on the call, agreed that the new information was not sufficient to address its concerns about Reliability. The Committee voted not to reconsider the measure.

2816: Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma

During the in-person meeting, the Committee raised a number of significant concerns with the testing and construction of the measure and the appropriateness criteria. Specifically, the measure was tested in only one hospital, making it difficult to discern meaningful differences among institutions, and not all critical data elements related to appropriateness were tested. The Committee also noted that the measure specifications permit variable use of pharmacy data, as available. Measure #2816 did not achieve consensus on Evidence (12-Pass; 10-No Pass) and did not pass Validity (N/A-H; 1-M; 17-L; 5-I); therefore it was not recommended for endorsement. The developer used data element level validity testing, which is accepted under NQF guidance to assess both Reliability and Validity; therefore, the Committee did not vote separately on Reliability.

One comment was received for this measure from the American Academy of Allergy, Asthma and Immunology (AAAI), which supported concerns about the lack of risk adjustment brought up by the NQF Pulmonary and Critical Care Standing Committee during a previous review; the Pediatric Committee did not discuss this issue since the discussion did not progress to that aspect of validity, given the other concerns.

Developer Rationale for Reconsideration:

At the in person meeting, for measure #2816, consensus was not reached for Evidence, the measure passed on Gap, and did not pass on Reliability. While the developer requested reconsideration for both measures, it did not provide a separate, specific rationale for this measure. The developer noted that the data provided for #3819 also informs this measure; no additional information was provided specifically related to appropriateness.

Developer Response to Comment:

"We have submitted this to the Pediatric Committee in part because of its greater sensitivities to the issues specific to children and in this case asthma in children. Nearly half of US children are covered by public health insurance programs. Equity of outcomes across race and social class is a preeminent concern and value in child health, especially for asthma. As the internationally accepted NHLBI guidelines states, "As a general rule, patients with well-controlled asthma should have:

- Few, if any, asthma symptoms.
- Few, if any, awakenings during the night caused by asthma symptoms.
- No need to take time off from school or work due to asthma.
- Few or no limits on full participation in physical activities.
- No emergency department visits.
- No hospital stays.
- Few or no side effects from asthma medicines."

Further it is not clear whether those stressors that increase asthma burden are likely to increase or to decrease the level of appropriateness of ED use for asthma. Cogent arguments can be made in either direction, or for not at all.

Measure 2816, Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma, is stratified by age, specifically the measure is reported for children ages 2-5,6-11, 12-18, and optionally, 19-21. This is because asthma and its management are related both to the child's age and stage of development. Hence comparing performance in young children is very different from performance in adolescents.

Further, this measure of appropriateness is best interpreted in light of other measures, such as the rate of ED use for asthma. High levels of appropriateness may suggest a highly functional primary care and outpatient approach to asthma. However high levels of appropriateness and high levels of utilization may together suggest that asthma outcomes form outpatient management are not as desired. Hence, this measure informs but is not dispositive without other data.

Our formal RAND style panel of national experts did not recommend risk adjustment by race or social class. They recommend stratification by the age groups indicated below. Environmental differences may produce unequal burdens on various health plans, but the

field's capacity to discriminate and risk adjust in that manner is of uncertain value and such data for adjustment are neither readily available, nor is there a consensus on what and when and how to adjust for such exposure.

Establishment of asthma control should occur from an early age. Because of challenges in identifying asthma before the age of 2, we have not included this age group in our specification.

For purchasers who are interested in stratification beyond race and age we provide OPTIONAL specifications that allow them to ask health plans to incorporate additional stratification in the measure (e.g. insurance status, county rates of poverty, and rurality/urbanicity). Contracting health plans can negotiate with purchasers and other accountability agencies to demonstrate stratified performance if they so desire.

This measure requires stratification by the following age groups:

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

These age strata are to be reported distinctly and not combined for reasons noted above.

This measure has optional stratifications for the following that can be determined by the reporting agency to use all or none, as appropriate:

- Race/Ethnicity: Hispanic, Non-Hispanic Black, Non-Hispanic White; Non- Hispanic Asian/Pacific Islander, other Non-Hispanic
- Insurance type (Public, Commercial, Uninsured)
- Benefit type (if insured): HMO, PPO, Medicaid Primary Care Case Management (PCCM) Plan, Fee for Service (FFS), other relevant enrollment categories (e.g., TANF, SSI)
- Urban influence codes: Identify the Urban Influence Code or UIC. (2013 urban influence codes available at: http://www.ers.usda.gov/data-products/urban- influence-codes.aspx#.UZUvG2cVoj8). Use parent or primary caregiver's place of residence to determine UIC. State and county names can be linked or looked up directly or zip codes can be linked to county indirectly, using the Missouri Census Data Center (http://mcdc.missouri.edu/). These data will link to county or county equivalents as used in various states.
- Urban Influence Codes (UIC) have been developed by the USDA to describe levels of urbanicity and rurality. While each UIC has its own meaningful definition, some researchers choose to aggregate various codes. Well regarded schemas for aggregation of codes include Bennett and colleagues at the South Carolina Rural Research Center. Their aggregation scheme brings together Codes 1 & 2 as Urban; 3,5, & 8 as micropolitan rural; 4,6, & 7 as rural adjacent to a metro area; and 9, 10, 11, & 12 as remote rural. We acknowledge that UIC 5 (adjacent rural area) may appropriately be aggregated with 4,6,&7 as rural. Frontier health care may be approximated by analysis of the remote rural

categories (UIC 9, 11 and 12). Alternatively, Gary Hart, Director of the Center for Rural Health at the University of North Dakota School of Medicine & Health Science suggests that UIC 9-12 is the best overall approach to using county level data to study frontier health. Inclusion of UIC 8 would make the analysis more sensitive to including frontier areas but at a meaningful cost in specificity.
- Those interested in care specific to large cities may wish to aggregate the rural area and analyze UIC 1 and 2 separately.
- When stratifying by urbanicity or UIC, the reporting and accountability entities should specify clearly what if any aggregating schema was used.
- Identify the Level of Poverty in the parent or primary caregiver's county of residence. The percent of all residents in poverty by county or county equivalent are available from the US Department of Agriculture at http://www.ers.usda.gov/data-products/county-level-data-sets/download- data.aspx. Our stratification standards are based on 2011 US population data that we have analyzed with SAS 9.3. Using parent or primary caregiver's state and county of residence (or equivalent) or FIPS code, use the variable PCTPOVALL_2011 to categorize into one of 5 Strata:
 o Lowest Quartile of Poverty if percent in poverty is <=12.5%
 o Second Quartile of Poverty if percent in poverty is >12.5% and <=16.5%
 o Third Quartile of poverty if percent in poverty is >16.5% and <=20.7%
 o First Upper Quartile (75th-90th) if percent in poverty is >20.7% and <=25.7%
 o Second Upper Quartile (>90th percentile)

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

To summarize:

Appropriateness of ED visits is a new construct for pediatric asthma. As such, there are no pre-existing data to suggest a disparate burden of either appropriate or inappropriate ED visits by socioeconomic class or by health plans caring for them. The NHLBI guideline is clear in articulating the expectation that outcomes should be equally good across the general population of individuals regardless of who they are and even how severe their asthma is (obviously there are true exceptions here, but they would not be well accounted for in any risk adjustment or stratification schema that we have ever seen). The Pediatric Committee is in a better position to understand and appreciate the implications of all of this for children and to incorporate such insights into their evaluation of this measure (and the similar rate measure).

The lack of required stratifications by risk does not lead to misinterpretation of results as a potential unintended consequence if the measure is implemented. In fact, this measure is specified to give flexibility to plans and to purchasers to respond to local conditions and needs by using stratification as needed and desired to compare performance within specified strata. These are desirable attributes for child health quality measures."

Committee Response:

The Committee reviewed the new material prior to the call, and after discussion, agreed that the new information was not sufficient to reconsider the measure. Specifically, the Committee agreed the new information did not address the issues raised previously with the validity of the numerator construct (i.e., the measure was still only tested at one institution); the Committee also agreed its concerns about the evidence were not adequately addressed. The Committee voted not to reconsider the measure.

NQF Member Voting Results

The four recommended measures were approved with 75% approval or higher. Representatives of five member organizations voted; no votes were received from Provider Organizations, Public/Community Health Agency, QMRI, and Supplier/Industry Councils. Results for each measure are provided below. Results for each measure are provided in Appendix B.

Removal of Endorsement

Ten measures previously endorsed by NQF were not re-submitted for maintenance of endorsement or were withdrawn during the endorsement evaluation process. Endorsement for these measures will be removed.

Measure	Measure Description	Reason for withdrawal
0010: Young Adult Health Care Survey (YAHCS)	The Young Adult Health Care Survey (YAHCS) is a survey of adolescents 14-18 years of age that assesses how well the health care system provides adolescents with recommended preventive care. The YAHCS assesses the provision of private and confidential care, experience of care, helpfulness of care provided, and the following aspects of preventive care: • Preventive screening and counseling on risky behaviors. • Preventive screening and counseling on sexual activity and sexually transmitted diseases (STDs). • Preventive screening and counseling on weight, healthy diet, and exercise. • Preventive screening and counseling on emotional health and relationship issues. • Private and confidential care. • Helpfulness of counseling. • Communication and experience of care. • Health information. The YAHCS has been used to assess health care quality at the national, State, geographic, county, and health plan levels. English and Spanish versions of the YAHCS are available free of charge on CAHMI´s web site (http://www.cahmi.org), and additional information is available at the Child Healthcare	Developer is no longer able to support the measure.

Measure	Measure Description	Reason for withdrawal
	Quality Toolbox: www.ahrq.gov/chtoolbx/measure7.htm	
0011: Promoting Healthy Development Survey (PHDS)	The Promoting Healthy Development Survey (PHDS) is a 43-item parent survey that can be used by health care providers, health systems, Medicaid agencies, and other stakeholders to measure and improve the quality of preventive and developmental care for children ages 0-48 months. The survey is designed to measure parent's experience with care and the extent to which they received preventative and developmental services in accordance with nationally recommended guidelines put forth by the American Academy of Pediatrics and Bright Futures practice guidelines (3rd edition).1 The PHDS contains 11 modules. The first six items represent measures. Taken together, the six measures also make up a composite PHDS Comprehensive Care Measure. These measures are the focus of this application. Items #7-11 are, respectively, an individual quality measure submitted separately by another steward (Standardized developmental and behavioral screening; NQF measure number 1448), not a quality of care measure (Access to care), used for stratification (Follow-up for children at risk and the CHSCN screener), or provide demographic and background information. Taken individually or as a whole, the PHDS provides valid measures for system, plan, and provider-level assessments.2 1. Anticipatory guidance 2. Parenting information, resources in community 3. Family centered care 4. Ask about and address parental concerns 5. Assessment of family safety, alcohol use and substance abuse 6. Assessment of family psychosocial screening 7. Standardized developmental and behavioral screening 8. Access to care and care coordination 9. Follow-up for children at risk 10. CSHCN screener; 11. Parent and family behaviors and respondent health;	Measure was submitted during this cycle as five new measures: NQF #2219, #3220, #3221, #3222, and #3223
0718: Children Who Had Problems Obtaining Referrals When Needed	The measure aims to ascertain the perceived difficulty in obtaining referrals for children when needed for optimum health.	Developer is no longer able to support the measure.

Measure	Measure Description	Reason for withdrawal
0723: Children Who Have Inadequate Insurance Coverage For Optimal Health	The measure is designed to ascertain whether or not current insurance program coverage is adequate for the child´s health needs--whether the out of pocket expenses are reasonable; whether the child is limited or not in choice of doctors; and whether the benefits meet child´s healthcare needs.	Developer is no longer able to support the measure.
1330: Children With a Usual Source for Care When Sick	Whether child has a source of care that is known and continuous (categorized as a doctor´s office, hospital outpatient department, clinic or health center, school, friend or relative, some other place, or a telephone advice line)	Developer is no longer able to support the measure.
1332: Children Who Receive Preventive Medical Visits	Assesses how many medical preventive visits in a 12 month period, such as a physical exam or well-child check-up (does not include visits related to specific illnesses)	Developer is no longer able to support the measure.
1334: Children Who Received Preventive Dental Care	Assesses how many preventive dental visits during the previous 12 months	Developer is no longer able to support the measure.
1335: Children Who Have Dental Decay or Cavities	Assesses if children age 1-17 years have had a toothache, tooth decay or cavities in the past 6 months	Developer is no longer able to support the measure.
1337: Children With Inconsistent Health Insurance Coverage in the Past 12 Months	Measures whether children are uninsured at the time of the survey or if currently insured children experienced periods of no insurance during past 12 months	Developer is no longer able to support the measure.
1448: Developmental Screening in the First Three Years of Life	The percentage of children screened for risk of developmental, behavioral and social delays using a standardized screening tool in the first three years of life. This is a measure of screening in the first three years of life that includes three, age-specific indicators assessing whether children are screened by 12 months of age, by 24 months of age and by 36 months of age.	The developer is unable to maintain the measure to NQF's current requirements for reliability and validity.

Appendix A – Measures Not Recommended for Endorsement

The table below lists the Committee's vote and rationale for measures not recommended for endorsement.

LEGEND: Y = Yes; N = No; H = High; M = Moderate; L = Low; I = Insufficient

Measure	Voting Results	Standing Committee Rationale
2816 Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma	**Evidence** Pass-12; No-10 **Gap** H-2; M-18; L-2; I-0 **Reliability** N/A **Validity** H-N/A; M-1; L-17; I-5 **Feasibility** N/A **Usability and Use** N/A **Post Comment Call Vote on whether to reconsider the measure:** Y-0; N-17	The Committee did not recommend the measure because it did not pass Validity due to insufficient testing. The Committee voted not to reconsider the measure during the post-comment call.

Measure	Voting Results	Standing Committee Rationale
3189 Rate of Emergency Department Visit Use for Children Managed for Identifiable Asthma: Visits per 100 Child-years	**Evidence** Y-21; N-1 **Gap** H-9; M-13; L-1; I-0 **Reliability** H-N/A; M-1; L-4; I-18 **Validity** N/A **Feasibility** N/A **Usability and Use** N/A **Post Comment Call Vote on whether to reconsider the measure:** Y-9; N-8	The Committee did not recommend the measure because it did not pass Reliability due to insufficient testing. The Committee voted not to reconsider the measure during the post-comment call.

Measure	Voting Results	Standing Committee Rationale
3219 Anticipatory Guidance and Parental Education	**Evidence** First Vote Pass-15; No Pass-7 Second Vote Pass-8; No Pass-14 **Gap** H-1; M-17; L-1; I-0 (void by revote) **Reliability** N/A **Validity** N/A **Feasibility** N/A **Usability and Use** N/A	The measure did not pass Evidence due to concerns about the measure's construct of the applicable timeframe and subsequent attribution to a single provider, so it was not recommended.

Measure	Voting Results	Standing Committee Rationale
3220 Ask About Parental Concerns	**Evidence** Y-10; N-12 **Gap** H-1; M-18; L-3; I-0 **Reliability** H-0; M-7; L-13; I-1 **Validity** N/A **Feasibility** N/A **Usability and Use** N/A	The measure did not pass Reliability due to concerns about attribution to a single provider, the lack of standardization in the survey administration, and the lack of a cut-off for responding to the survey, so it was not recommended.

Measure	Voting Results	Standing Committee Rationale
3221 Family Centered Care	**Evidence** Y-6; N-16 **Gap** N/A **Reliability** N/A **Validity** N/A **Feasibility** N/A	The measure did not pass Evidence due to concerns about the applicable timeframe and subsequent attribution to a single provider, so it was not recommended.

	Usability and Use N/A	

Measure	Voting Results	Standing Committee Rationale
3222 Assessment of Family Alcohol Use, Substance Abuse and Safety	**Evidence** Y-7; N-15 **Gap** N/A **Reliability** N/A **Validity** N/A **Feasibility** N/A **Usability and Use** N/A	The measure did not pass Evidence due to concerns about the applicable timeframe and subsequent attribution to a single provider, so it was not recommended.

Measure	Voting Results	Standing Committee Rationale
3223 Assessment of Family Psychosocial Screening	**Evidence** Y-8; N-14 **Gap** N/A **Reliability** N/A **Validity** N/A **Feasibility** N/A **Usability and Use** N/A	The measure did not pass Evidence due to concerns about the applicable timeframe and subsequent attribution to a single provider, so it was not recommended.

Appendix B – NQF Member Voting Results

NQF MEMBER VOTING RESULTS

The four recommended measures were approved with 75% approval or higher. Representatives of five member organizations voted; no votes were received from Provider Organizations, Public/Community Health Agency, QMRI, and Supplier/Industry Councils. Results for each measure are provided below.

NQF Member Council	Voting Organizations	Eligible to Vote	Rate
Consumer	1	38	3%
Health Plan	1	21	5%
Health Professional	2	104	2%
Provider Organizations		110	0%
Public/Community Health Agency		15	0%
Purchaser	1	22	5%
QMRI		74	0%
Supplier/Industry		35	0%
All Councils	**5**	**419**	**2%**

Measure #3136 GAPPS: Rate of preventable adverse events per 1,000 patient-days among pediatric inpatients

Member Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	1			1	100%
Health Plan			1	1	
Health Professional	2			2	100%
Provider Organizations				0	
Public/Community Health Agency				0	
Purchaser	1			1	100%
QMRI				0	
Supplier/Industry				0	
All Councils	**4**	**0**	**1**	**5**	**100%**
Percentage of councils approving (>60%)					100%
Average council percentage approval					100%

*equation: Yes/ (Total - Abstain)

Measure #3153 Continuity of Primary Care for Children with Medical Complexity

Member Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	1			1	100%
Health Plan		1		1	0%
Health Professional	2			2	100%
Provider Organizations				0	
Public/Community Health Agency				0	
Purchaser	1			1	100%
QMRI				0	
Supplier/Industry				0	
All Councils	**4**	**1**	**0**	**5**	**80%**
Percentage of councils approving (>60%)					75%
Average council percentage approval					75%

*equation: Yes/ (Total - Abstain)

Voting Comments

America's Health Insurance Plan: Based on review of the measure testing attachment, testing for this measure appears to have been performed at the state level, not at the health plan level as described in the draft report. We also have concerns about the actionability of the measure from a performance improvement standpoint. Because of state regulations, plans must ensure members have their choice of provider. Lastly, we are concerned about health plans being penalized for a clinician leaving the network, a family appropriately seeing multiple providers until they find one that meets their needs, and/or similar scenarios.

American Academy of Pediatrics: The AAP agrees with the focus and intent of this measure, thus will vote to recommend for endorsement. Evidence suggests that a higher continuity of care is associated with better outcomes and the use of this measure aligns with the AAP's vision to ensure that every child receives the right care every time.

Measure #3154 Informed Coverage

Member Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer			1	1	
Health Plan			1	1	
Health Professional	2			2	100%
Provider Organizations				0	
Public/Community Health Agency				0	
Purchaser	1			1	100%
QMRI				0	
Supplier/Industry				0	
All Councils	**3**	**0**	**2**	**5**	**100%**
Percentage of councils approving (>60%)					100%

Average council percentage approval	100%

*equation: Yes/ (Total - Abstain)

Measure #3166 Antibiotic Prophylaxis Among Children with Sickle Cell Anemia

Member Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	1			1	100%
Health Plan		1		1	0%
Health Professional	2			2	100%
Provider Organizations				0	
Public/Community Health Agency				0	
Purchaser	1			1	100%
QMRI				0	
Supplier/Industry				0	
All Councils	**4**	**1**	**0**	**5**	**80%**
Percentage of councils approving (>60%)					75%
Average council percentage approval					75%

*equation: Yes/ (Total - Abstain)

Voting Comments

America's Health Insurance Plan: Based on review of the measure testing attachment, testing for this measure appears to have been performed at the state level, not at the health plan level as described in the draft report. We are concerned about the limitations that may exist with availability of pharmacy data because of carve-out. We also feel that the denominator population seems too small for reliable plan-level measurement. Additionally, it is not clear from the report whether a year or more of continuous enrollment is required for a member to be counted in the denominator. We feel this is an important prerequisite for achieving the 300 days of coverage.

Appendix C – Measure Evaluation Summary Tables

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

Measures Recommended

3136 GAPPS: Rate of preventable adverse events per 1,000 patient-days among pediatric inpatients

Submission | Specifications

Description: GAPPS is a measure of the number of preventable adverse events per 1,000 patient-days among pediatric inpatients. It is designed to compare rates across institutions and over time. The GAPPS measure utilizes the GAPPS trigger tool to identify adverse events.

Numerator Statement: The number of preventable adverse events found in a patient sample.

Denominator Statement: The denominator is 1,000 patient-days for all sampled pediatric patients who meet inclusion, but not exclusion, criteria.

Exclusions: N/A

Adjustment/Stratification: Statistical risk model Stratification is not required within institutions. However, if desired, quality improvement teams may choose to stratify preventable adverse event rates. Variables commonly used to stratify outcome measures include service (e.g., medical versus surgical), department (e.g., cardiology, neurology, etc.), and patient safety focus area (e.g., healthcare-associated infections).

For comparisons between institutions, preventable adverse event rates should be stratified by teaching versus community hospitals due to differences in types (e.g., complexity) of patient populations

Level of Analysis: Facility

Setting of Care: Hospital : Acute Care Facility

Type of Measure: Outcome

Data Source: Electronic Health Record (Only), Paper Records

Measure Steward: Center of Excellence for Pediatric Quality Measurement

STANDING COMMITTEE MEETING [03/02/2017]

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

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

1a. Evidence: **21-Pass; 1-No Pass;** 1b. Performance Gap: **16-H; 6-M; 0-L; 0-I**

Rationale:

- The developer noted the intent of the measure is to provide a method to identify events so that hospitals can use the results to prioritize patient safety improvement efforts.
- While there is limited evidence directly pertaining to the pediatric population, the Committee took note of evidence that a similar adult measure for the adult population effectively identified and reduced preventable adverse events.
- For medication-related events, the Committee noted there is a greater risk of dose-related errors in the pediatric population, so the ideal evidence would be from studies completed in the pediatric population; the Committee recognized, however, the limitations of the available evidence.
- The developer agreed with the Committee member in noting that the measure focuses on in-patient admissions, as a way to keep implementation of the measure consistent across care settings.
- The Committee agreed there is a gap: No pediatric tool or measure currently exists to measure preventable adverse events, and there are significant numbers of such events, which vary by measured entity. In the testing sample of 16 hospitals, 414 adverse events were identified and

3136 GAPPS: Rate of preventable adverse events per 1,000 patient-days among pediatric inpatients

214 (50.7%) were preventable. Compared to community hospitals, academic hospitals had higher preventable harm rates (13.1 [CI 11.4-15.2] vs. 2.4 [CI 1.5-3.8] AEs/1,000 patient days, p<0.001).

- The developer identified a disparities gap in the occurrence of an adverse event based on race/ethnicity, the number of chronic conditions, and insurance status.

2. Scientific Acceptability of Measure Properties: <u>The measure meets the Scientific Acceptability criteria</u>

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

2a. Reliability: **N/A-H; 16-M; 6-L; 0-I** 2b. Validity: **N/A-H; 16-M; 6-L; 0-I**

<u>Rationale</u>:

- From a literature review and an appropriateness panel using the RAND approach, the developer created a list of 54 draft triggers. The developer noted that no gold standard exists. The panel assessed the relationship between each trigger and risk of an adverse event, the feasibility to extract data from medical record, and if panel members believed it was a valid trigger. The developer then tested the list in the national field test and found the list to be appropriate. Based on the testing, the developer recommends a sampling 60 charts per quarter/20 per month across the institution and then using a random number generator to select 25 charts.
- The Committee also raised concerns about diagnostic errors, since the trigger tool would not be able to identify the diagnostic error. The developer agreed with the Committee and noted that the measure is more appropriate for analyzing errors of commission.
- A Committee member inquired about the exclusion of patients who are in the hospital less than 24 hours and noted it is possible for a patient to have a high-risk procedure and be discharged within that time frame. The developer agreed with the importance of identifying adverse events for short term stays, but explained that in order to define the measure and keep it consistent across institutions, it decided to focus this measure on identifying adverse events for inpatients.
- The Committee discussed the threat to reliability due to the learning curve involved in training a reviewer. Validity testing found low sensitivity of new reviewers in comparison to expert reviewers, but their scores improved over time. Based on the testing experience, the developer expanded and extended the educational materials for reviewers; the developer also suggests a ramp-up period for implementation as reviewers learn to review charts.
- In response to a question, the developer clarified that the specifications include urinary tract infections and other hospital-acquired infections, as well as severe mental health conditions, but not admissions to less-than-acute care (such as newborn nurseries, rehab, or chronic psychiatric care). The developer agreed that emergency department patients are an important safety population, but were not within the scope of this measure.
- In response to a question, the developer noted that other voluntary reporting systems only identify 2-10 percent of what the GAPPS tool identifies.
- A Committee member asked if minor events were equated to major events. The developer responded that the measure uses the National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP) scale to rate events to distinguish the severity of events in reporting results.
- The Committee noted the potential issues when using the measure to compare academic versus community hospitals. The developer recommended stratification to account for patient population differences in academic versus community hospitals.
- In response to a question, the developer clarified that patients who are transferred from another institution are not included in the measure to avoid potential mis-assignment of an adverse event and resultant unintended public reporting consequences: The measure score of

3136 GAPPS: Rate of preventable adverse events per 1,000 patient-days among pediatric inpatients
hospitals with many transferred patients could be negatively impacted if transfer patients were included. • NQF staff clarified for the Committee that the highest possible score for reliability was a moderate, since the measure is tested at the data-element level only; the highest possible score for validity is also moderate, since validity testing is patient-level data element. • A Committee member asked for clarification for differences in testing between community hospitals and academic hospitals. The developer responded that testing had been done in the two different settings to test the functionality of the measure and the measure functioned well in both. • The developer noted that the automated trigger list contained more triggers than the manual trigger list. However, it explained that these triggers are less frequent or are based on laboratory values, which would be too burdensome to screen for manually. Testing was conducted at institutions using the manual approach. • The developer cited a 2007 study examining hospitals without fully integrated electronic health record (EHR) system and found the function of the measure was equivalent to a hospital with a fully integrated EHR.
3. Feasibility: 1-H; 15-M; 5-L; 1-I *(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c.Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)* Rationale: • The Committee raised concerns about the number of physicians involved in the review and the time burden. The developer clarified the specifications require that registered nurses conduct the first level of review and physicians the second level, which assesses preventability of the event. The developer estimates physicians spend a minimum of four hours a year; Committee members emphasized the time spent reviewing the adverse events was a valuable opportunity for physicians to learn more about the safety vulnerabilities at their facility. • The developer noted it provides training materials without cost and opined that the actual time commitment was low. The training program consists of five one-hour videos and a small amount of homework/case practices. In response to a question, the developer stated it does not currently have specific follow-up or additional training for poor reviewers. • The Committee raised concerns about the use of the tool in hospitals with no electronic system, or electronic systems where automated trigger screening is not yet possible. The developer responded that an integrated electronic health record system made the process easier, but the measure was created to be, and is still, feasible with manual review. The developer noted that the measure is designed to be manually administered and was tested as such. Hospitals without electronic health records systems will be able to implement the measure. • In response to questions, the developer explained the ability to automate screening helps to improve efficiency, but it is possible to do the needed chart review manually.
4. Usability and Use: 0-H; 10-M; 11-L; 1-I *(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)* Rationale: • The Committee noted the measure is meaningful for internal quality improvement, but could have several unintended consequences if used in hospital-to-hospital comparisons and public reporting. Specifically, facilities that do a good job of documenting and reporting are penalized,

whereas under-reporting (intentionally or due to poor training) rewards facilities. It also noted the difficulties involved in accounting for variation in patient populations.

- The Committee did not reach consensus on usability and use, but this is not a must-pass criterion.

5. Related and Competing Measures

- No related or competing measures noted.

Standing Committee Recommendation for Endorsement: **14-Y; 8-N**

Rationale

- The Committee agreed this measure meets the NQF criteria for endorsement.

6. Public and Member Comment: April 12-May 11, 2017

One commenter provided specific questions regarding several of the measure's specifications and suggested updates intended to clarify automated triggers to increase the specificity and clarity of the measure specifications. Another commenter did not support the endorsement of this measure, noting that implementing the trigger might be difficult and require excessive resources, and suggested that the tool lacks validity in identifying adverse events. The developer responded to each comment and question separately.

Developer Response 1:

- Trigger: Consider rewording to "Hepatotoxic medications and RISING liver enzymes (AST, ALT)"
  - Thank you for the suggestion. A consideration here is that if there were not previous hepatic enzyme measurements and the first measurement showed elevated enzymes, this would need to be investigated. If this were written to only include those that are rising (therefore requiring a previous lower value), the process might miss a possible hepatoxic injury. Therefore our preference is to retain the language as "elevated."
- Please explain how "Physician orders: Abrupt medication stop" is defined in the automated trigger tool? Most medication stops are abrupt (with rare exceptions like steroid weans or PCA infusions)
  - The definition in our Manual of Operations reads as follows, "An abrupt medication stop is best described as an unexpected stop or deviation from typical ordering practice (e.g., discontinuation of a recently started medication)." Since this type of clinical decision making may be challenging to automate, it is not recommended for electronic trigger review.
- Please define "Transfer to higher level of care" more specifically. Many hospitals have observation units where most patients go home but some patients are admitted to the floor (higher level of care) after a specified time.
  - The definition from our Manual of Operations reads as follows, "All transfers from an acute care area to an intensive care unit or intermediate care unit ("step-up unit") should be considered a trigger." Therefore the scenario presented in the comment would not meet the defined criteria.
- Consider changing Pressure ulcer documentation to ">= Stage 2" instead of just stage 2.
  - Thank you for identifying this discrepancy. This was a typographical error and should read as you suggest. During our testing, the reviewers were instructed to investigate exactly as is suggested by the comment, meaning all pressure injuries Stage 2 and higher and unstageable. We will edit the relevant documents to reflect this change.
- Many places will start patients on laxatives simultaneously with opioids, but patients will still get constipated. Would this qualify as a trigger, or is it only a trigger if laxatives are started after (e.g. >=24 hours after) opioids are started? Latter would be more specific, less sensitive.

    - Thank you for pointing out this ambiguity. We agree that excluding cases where laxatives are introduced concurrently (<24 hours after) with opioids is reasonable. The trigger is looking at cases where laxatives were given subsequent to the initial prescription of opioids (>=24 hours after). We will edit the relevant documents to reflect this change.
- Consider adding "positive coagulase-negative staphylococcus species blood culture" as a trigger for review; per algorithm, it should have a higher than 10% rate of being a true contaminant (i.e., an adverse event).
    - Thank you for this comment. Since we currently look at a more broadly based trigger (positive blood culture 48 hours after admission), all of the occurrences of the suggested trigger would be included in the trigger as written. We hesitate to insert a new trigger into the recently reviewed tool at this stage.
- Please clarify the denominator of whether a partial day counts as a day. For example, is 1.5 days = 2 days or 1.5 days? What is the start and stop time for determining LOS duration (e.g. start of: time of arrival to floor, time of admission from ED; end of: time of discharge order, time of leaving hospital?)
    - Length of stay is calculated as the number of days (discharge date minus admit date). For example, a patient who arrives at 4am on May 17th and is discharged at 4pm on May 18th has a length of stay of 1.0 day. However, a patient who arrives at 10pm on May 17th and is discharged at 10am on May 19th has a length of stay of 2.0 days. Start and stop times were not used to determine length of stay duration, only admit and discharge date.
- Step 2: Line 4. Please describe whether the unit of study (whether entire hospital, division, etc.) should remain stable over time.
    - Thank you for the opportunity to clarify. We would suggest that the unit of interest remain stable over time.

Developer response 2:

- We would like to thank Dr. Austin for his comments. As the measure is implemented, the resource burden, while not trivial, should be manageable while providing a great deal of benefit in terms of increased safety events identified. The primary reviewer, typically an experienced nurse, is asked to perform chart review quarterly on 60 patient records per quarter with a limit of 30 minutes per chart. This would represent a total of 30 hours per quarter or 10 hours per month or 2.5 hours a week. The secondary reviewer, typically a physician, reviews the primary reviewer's findings. Assuming a high rate of harm or 33 events per 100 admissions, this would be 20 events to review each quarter. During validation testing, our physician team required on average 4 minutes per chart to review. Therefore, the typical time burden on the secondary reviewer would be approximately 80 minutes per quarter. Based on the frequency of events and the resources required, it is our view that the benefit of this modest resource requirement would far outweigh the burden.

  In regards to validity, we developed the draft trigger tool used in the GAPPS measure through the RAND/UCLA Appropriateness Method, which is a modified Delphi process.(1–3) We first compiled a set of 78 candidate triggers from a literature review of existing pediatric and adult trigger tools and input from trigger tool experts.(4–6) We then recruited nine panelists from national pediatric and patient safety organizations and asked them to rate separately the validity and feasibility of the candidate triggers on a nine-point scale (where 1 is the least valid/feasible and 9 is the most valid/feasible). A trigger was considered valid if it was judged to

be reasonably likely to identify an underlying AE, indicating that harm potentially occurred. A trigger was considered feasible if it was judged likely to be accurately and consistently documented in either paper or electronic medical records as part of patient care at a wide range of hospitals, from smaller community sites to larger tertiary care centers. Applying the RAND/UCLA Appropriateness Method, we accepted triggers that had both median validity and feasibility ratings greater than or equal to seven. This approach resulted in inclusion of 54 of the initial 78 candidate triggers in the draft GAPPS trigger list.

We focused our validity testing on evaluation of how accurately and completely "typical reviewers" (i.e., clinicians who are trained in GAPPS methodology but not necessarily trigger tool experts) were able to identify preventable AEs using the measure as compared to expert reviewers. The expert reviewers had extensive experience with using trigger tools for preventable AE identification and consequently were most likely to identify preventable AEs accurately and completely. To evaluate the validity of the GAPPS measure, we assessed the performance of the National Field Test hospitals' internal reviewers relative to the performance of external expert reviewers in applying the measure (as documented in our NQF submission materials).

REFERENCES

1. Fitch K, Bernstein S, Aguilar MD, Burnand B, LaCalle JR, Lázaro P, et al. The RAND/UCLA Appropriateness Method User's Manual. Santa Monica, CA: RAND; 2001.
2. Brown B. DELPHI PROCESS: A Methodology Used for the Elicitation of Opinions of Experts. Rand Corp. 1968 Sep;1–14.
3. Sweidan M, Williamson M, Reeve JF, Harvey K, O'Neill JA, Schattner P, et al. Identification of features of electronic prescribing systems to support quality and safety in primary care using a modified Delphi process. BMC Med Inform Decis Mak. 2010 Apr 15;10(1):21.
4. Stockwell D, Bisarya H, Classen D, Kirkendall E, Landrigan C, Lemon V, et al. A trigger tool to detect harm in pediatric inpatient settings. Pediatrics. 2015;
5. Griffin FA, Resar RK. IHI Global Trigger Tool for Measuring Adverse Events (Second Edition). Institute for Healthcare Improvement; 2009. (IHI Innovation Series white paper).
6. Kirkendall ES, Kloppenborg E, Papp J, White D, Frese C, Hacker D, et al. Measuring adverse events and levels of harm in pediatric inpatients with the Global Trigger Tool. Pediatrics. 2012 Nov;130(5):e1206-1214.

Committee Response:

- Thank you for providing these comments on measure #3136. The Committee discussed the measure specifications and validity during the in-person meeting. The Committee did note that that the highest possible score for reliability was a moderate, since the measure is tested at the data-element level only; the highest possible score for validity also is moderate, since validity testing is patient-level data element. Overall, the Committee determined that the measure as specified and tested offered sufficient validity for endorsement and did not wish to reconsider the measure.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

3153 Continuity of Primary Care for Children with Medical Complexity

Submission | Specifications

Description: This measure assesses the percentage of children with medical complexity age 1 to 17 years old who have a Bice-Boxerman continuity of care index (hereafter referred to as Bice-Boxerman COC index) of >=0.5 in the primary care setting over a 12-month period.

Numerator Statement: Number of eligible children(1) who have a Bice-Boxerman COC index >=0.50 in the primary care setting during the measurement year.

1. Eligible children are defined as children who are continuously enrolled for 12 months with no more than a 30-day gap in enrollment. Children with a gap greater than 30 days are excluded because of the potential for them to be enrolled in a different health plan at that time. In such cases, the child's administrative data for the health plan being measured would be incomplete and thus might not reflect the health plan's true performance on the measure. The timeframe of 30 days as the length of the gap was chosen to be consistent with the month-to-month eligibility assessments used by many Medicaid health plans.

Denominator Statement: Children with medical complexity(1) who are 1-17 years old(2) and who have had >= 4 primary care visits(3) during the measurement year.

1. Children with medical complexity are defined as children who are classified by the Pediatric Medical Complexity algorithm, Version 2 (PMCA-V2) as having no chronic illness or non-complex chronic illness.

2. Children must be >=1 year and <=17 years of age on the last day of the measurement year.

3. Research has shown that stability of the COC index increases as the number of visits increases (ie. less subject to significant change as a result of minor variations in care dispersion).(1) We therefore established a minimum of four visits as has been done in previous studies.(1-3)

References

1. Christakis DA, Wright JA, Koepsell TD, Emerson S, Connell FA. Is greater continuity of care associated with less emergency department utilization? Pediatrics. 1999;103(4 Pt 1):738-742.

2. Christakis DA, Mell L, Koepsell TD, Zimmerman FJ, Connell FA. Association of lower continuity of care with greater risk of emergency department use and hospitalization in children. Pediatrics. 2001;107(3):524-529.

3. Tom JO, Tseng C-W, Davis J, Solomon C, Zhou C, Mangione-Smith R. Missed well-child care visits, low continuity of care, and risk of ambulatory care-sensitive hospitalizations in young children. Arch Pediatr Adolesc Med. 2010; 11:1052-1058.

Exclusions:

Adjustment/Stratification: No risk adjustment or risk stratification N/A, no stratification is recommended.

Level of Analysis: Health Plan

Setting of Care: Clinician Office/Clinic

Type of Measure: Structure

Data Source: Claims (Only)

Measure Steward: Seattle Children's Research Institute

STANDING COMMITTEE MEETING [03/02/2017]

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

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

1a. Evidence: **3-H; 17-M; 2-L; 0-I**; 1b. Performance Gap: **9-H; 13-M; 0-L; 0-I**

Rationale:

3153 Continuity of Primary Care for Children with Medical Complexity

- The developer conducted a literature review of continuity of care and found an association between continuity of care and better outcomes in multiple pediatric studies. The Committee noted that the tool was developed in 1977 and the evidence cited only uses this one tool.
- Since the measure relies on administrative claims data, a Committee member raised concerns about the use of the Bice-Boxerman Index in a collaborative care setting. The developer noted that the index accounts for patients seeing the same primary care clinician (physician, nurse practitioner, physician's assistant) several times, in addition to seeing two primary care clinicians multiple times.
- The specifications require a minimum of four visits for a patient to be included. The developer noted that this is due to evidence demonstrating that the Bice-Boxerman Index is less stable if less than four visits are used.
- The Committee agreed there was evidence to support the measure.
- The developer reported that performance across 17 state Medicaid plans varied from 23% to 96%. Fee-for-service states and combination fee-for-service and managed care states were accounted for in the gap analysis and testing. Since MAX data are often incomplete with respect to managed care data, the developer acknowledged that it is possible that the variation in state scores could be attributed to the lack of information in states with higher managed care utilization. Nevertheless, the Committee concluded that a gap in care exists.

2. Scientific Acceptability of Measure Properties: <u>The measure meets the Scientific Acceptability criteria</u>

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

2a. Reliability: **5-H; 14-M; 2-L; 1-I**; 2b. Validity: **1-H; 17-M; 4-L; 0-I**

<u>Rationale</u>:

- Validity and reliability testing used a 2008 Medicaid analytic extract for 17 Medicaid plans from MAX data. Testing was performed using the ICD-9 specifications, but the developer stated an ICD-10 version will be available in the near future.
- In response to a question, the developer explained that the measure was tested for state Medicaid health plans; testing using commercial health plan data was not conducted.
- A Committee member raised validity concerns about the categorization of the measure's result as either pass or fail for continuity of care. The Committee member inquired if the complexities involved in continuity of care could be captured using a binary result. The developer noted that the evidence indicated a Bice-Boxerman index score of >=0.5 resulted in better outcomes and that in its study, the mean score was 0.65. The developer stated it had not identified issues with the pass/fail construct.
- In response to a question, the developer responded that, due to the incompleteness of 2008 MAX data and poor definition of pediatric ambulatory care-sensitive hospitalizations, it did not look at hospitalizations.

3. Feasibility: 8-H; 14-M; 0-L; 0-I

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

<u>Rationale</u>:

- The Committee concluded that implementation is feasible: Electronic claims data are readily accessible and the developer makes the SAS code to compute the measure publicly available.
- In response to a question from a Committee member, the developer stated that the provided SAS code was applicable to commercial claims.

4. Usability and Use: **1-H; 16-M; 5-L; 0-I**

3153 Continuity of Primary Care for Children with Medical Complexity

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

Rationale:

- A Committee member noted the measure lacks the ability to measure the quality of the visits.
- A Committee member noted this measure will help encourage continuity at the organizational and plan levels.
- It also was noted that the goal of the measure is not to incentivize patients to have more visits, but to identify if individuals who are high utilizers have continuity in their care.

5. Related and Competing Measures

- No related or competing measures noted.

Standing Committee Recommendation for Endorsement: **17-Y; 5-N**

Rationale

- The Committee agreed this measure meets the NQF criteria for endorsement.

6. Public and Member Comment: April 12-May 11, 2017

- One comment was received supporting the endorsement of this measure.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

3154 Informed Coverage

Submission | **Specifications**

Description: Improved measurement of the continuity of insurance coverage in the Medicaid and CHIP population is needed to help maximize insurance continuity and coverage for vulnerable children. To further this goal, the AHRQ-CMS CHIPRA PQMP Center of Excellence at the Children's Hospital of Philadelphia developed the metric Informed Coverage. The metric is designed to more accurately measure coverage among children enrolled in Medicaid or CHIP at the state level and overcome the current inability in the Medicaid Analytic eXtract (MAX) dataset to determine whether a child disenrolled from Medicaid and CHIP due to loss of eligibility (such as due to parental income increase or the acquisition of employer-sponsored insurance, a "good" reason) or failure to appropriately re-enroll (a "bad" reason). This measure can help federal and state programs develop strategies to retain children eligible for coverage and minimize gaps that can occur during the renewal process. Informed Coverage assesses the continuity of enrollment of children in publicly financed insurance programs (Medicaid and CHIP), as defined by the ratio of enrolled month to eligible months over an 18 month observation window. Informed Coverage uses a natural experiment based on the random event of appendicitis to "inform" the estimate of coverage in a given state, bounded by two extreme assumptions regarding unknown eligibility information: Coverage Presumed Eligible (PE) and Coverage Presumed Ineligible (PI).

Numerator Statement: The numerator for Informed Coverage represents the sum (within a state) of months enrolled in Medicaid/CHIP for all children over an 18-month window.

Denominator Statement: The sum (within a state) of months eligible for Medicaid/CHIP for all children (0-18 years) over an 18-month window. In addition, months that could be defined as "eligible" are based on known events recorded in the MAX data that would affect eligibility (birth or ageing out).

3154 Informed Coverage

Exclusions: For the appendicitis calculation, the population is limited to children between the ages of 2 to 16 years old. To determine what is the best assumption to use (either the Appendectomy Coverage Rate (or ACR), PI, or PE) inside each state, we compare the observed appendectomy coverage rate in a state, to the estimated coverage rate that would be calculated in that state with either PI, or PE assumptions.

Adjustment/Stratification: No risk adjustment or risk stratification Stratification can be performed for Informed Coverage using any desired strata that policymakers choose to study. For example, stratification can be performed within states based on the type of Medicaid and CHIP programs, or by race.

Level of Analysis: Population : Regional and State

Setting of Care: No Applicable Care Setting

Type of Measure: Outcome

Data Source: Claims (Only)

Measure Steward: The Children's Hospital of Philadelphia

STANDING COMMITTEE MEETING [03/02/2017]

1. Importance to Measure and Report: <u>The measure meets the Importance criteria</u>

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

1a. Evidence: **23 -Pass; 0-No Pass**; 1b. Performance Gap: **10-H; 11-M; 1-L; 0-I**

<u>Rationale</u>:

- This measure provides states with information about participation of children in state insurance programs over an 18-month period. Using this information, states can take action to improve continuous enrollment, including presumptive eligibility. Research demonstrates that continuous enrollment improves access to healthcare services and health outcomes.
- Committee members also noted that, from a health plan perspective, the disenrollment and re-enrollment of Medicaid-eligible children is a costly process.
- The measure uses the random event of appendicitis – which is unrelated to any care received or not received – to assess whether children have insurance coverage at a given point in time (exactly four months before the date they are diagnosed, to cover instances of back enrollment).
- The Committee agreed that the use of appendicitis as a tracer resulted in a creative new measure that addresses the issue of assessing access and continuous enrollment.
- A Committee member suggested renaming the terms used in the measure as coverage presumed maximally eligible and coverage presumed minimally eligible to clarify the measure.
- The developer presented evidence that variation exists in continuous enrollment for 43 states, and that disparities occur among different racial and ethnic groups. The Committee agreed that a gap exists.

2. Scientific Acceptability of Measure Properties: <u>The measure meets the Scientific Acceptability criteria</u>

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

2a. Reliability: Original vote: **1-H; 11-M; 9-L; 3-I (consensus not reached);**

Re-vote on post-comment call: **1-H; 14-M; 3-L; 0-I**

2b. Validity: **0-H; 17-M; 6-L; 0-I**

<u>Rationale</u>:

- The measure is based on administrative claims data, and the level of analysis is state or region.
- The Committee expressed concern about whether the measure could be used for accountability purposes, given the difficulty in differentiating scores among states, most of which appeared to have overlapping confidence intervals (CIs) in the score-level reliability testing results—i.e., there were concerns in a significant number of cases that the measure scored could not

distinguish whether one state performed better than another. The developer responded that while some states had large confidence intervals (small states like Hawaii and Delaware), at least half of the states did not overlap. One Committee member noted the strength of demonstrating the CIs in this manner, stating that it makes the uncertainty and margins of error clear and explicit, where as other measures do not clearly document levels of uncertainty.

- Committee members also questioned the potential impacts of differences in incomes and burdens of chronic illnesses among states. The developer noted that it did not include risk adjustment because coverage should be independent of the health conditions of the states' populations.
- Committee members questioned whether informed coverage was related to better healthcare quality. They noted that states may have high coverage rates, but sicker children. The developer clarified that the measure is intended to provide states with information regarding their performance on coverage, not necessarily to address eligibility criteria or quality of received care.
- There was significant discussion about the randomness factor specified by the measure: The measure is calculated at a random point in time, which the developer argued limited the opportunity for bias. Additionally, the measure would be unaffected by temporal churns in coverage. The developer noted that no states offer more than four months of retroactive coverage, so coverage rate at that time before the appendectomy should not be significantly affected by temporal changes in coverage.
- The Committee noted that children whose families move in and out of eligibility for coverage may be missed in the measure. The developer clarified that changes in coverage would not adversely bias the metric, because the enrollment at a certain point in time must be dichotomous and would be unrelated to the chance event of appendicitis.
- With respect to validity, the measure developer chose the American Community Survey (ACS), which asks patients and families if they are covered by insurance or not, as the gold standard for comparison with its measure. Committee members noted that the informed coverage metric had a high correlation with ACS.
- The Committee questioned the effect of excluding those over age 16 and under 2 years old, and queried whether these exclusions were consistent with the evidence; ultimately Committee members agreed it was not an issue, noting the vast majority of the excluded children were excluded because they were about to age out. In addition, the Committee identified missing data as a concern, since missing data led to 12 percent of states being excluded from the analysis.
- The Committee did not reach consensus on the Reliability criterion, but the measure did pass Validity. The developer will provide additional reliability information during the comment period and the Committee will revote during the post-comment call.

3. Feasibility: 1-H; 22-M; 0-L; 0-I

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

Rationale:

- Committee members noted that the database used for the measure, the MAXX database, needs improvement and that six states were removed because of missing data issues. Following discussion, Committee members concluded that this issue was outside the developer's control, so they did not feel it was relevant to their vote on this criterion.
- The Committee expressed concern about whether states could accurately calculate the measure, given the complicated nature of the specifications. The developer noted there are two ways to calculate the measure, through bootstrapping or with SAS. The developer stated that

using SAS produced the same results and was more user-friendly for states to use when calculating the measure.

4. Usability and Use: 1-H; 17-M; 5-L; 0-I

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

Rationale:

- Committee members noted that the measure would be a useful tool for Medicaid leadership, and a useful self-assessment tool for states to improve their coverage rates, especially given the importance of the issue of continuous churn. However, the Committee did question whether this measure could be used in a traditional accountability fashion to reward or penalize states based on performance, given the overlapping scores reported for reliability testing. Committee members also noted the specifications are complex and difficult to understand, which may limit its usability by some audiences.
- Overall, Committee members viewed this measure as valuable for internal purposes and for measuring the quality of coordination in maintaining enrollment.

5. Related and Competing Measures

- No related or competing measures noted.

Standing Committee Recommendation for Endorsement: Y-13, N-4

Rationale

-
- The Standing Committee did not conduct a vote for Overall Suitability for Endorsement during the March 2, 2017, in-person meeting because consensus was not reached on the Reliability criterion. The Standing Committee discussed and re-voted on the Reliability criterion during the post-comment call on May 31, 2017. Based on new information submitted by the developer, the Standing Committee agreed the measure meets the Reliability criterion, and then voted Yes on Overall Suitability for Endorsement.

6. Public and Member Comment: April 12-May 11, 2017

- One commenter agreed with the intent of the measure to more accurately capture the continuity of coverage in the Medicaid program so that states can improve continuous coverage, an important measure to support children's health. However, while recognizing this metric used readily available datasets, the commenter felt assumptions that may not be universally accepted are used. As such, the commenter recommended that this measure be further validated and re-evaluated for inclusion in the future.

Developer response:

- We appreciate that the AAP agrees with the intent of our measure to more accurately capture the continuity of coverage in the Medicaid program so that states can improve coverage. The AAP suggested that our measure "requires assumptions that may not be universally accepted," without telling us which assumptions are objectionable. We would point out that with our assumptions, our results were carefully validated against the gold standard ACS (American Community Survey). Our results, in both development and validation, were superior to the current metrics of Continuity Ratio (Ku et al.) and Duration (currently used by CMS). Informed Coverage had better correlation with the ACS and less error deviation than the other metrics. See Validity Testing, Section 2b2.3, Table 2: Pearson Correlations. Also, see Validity Testing, Section 2b2.3, Table 3: Median Absolute Errors.

3154 Informed Coverage

Committee discussion and response:

- The developer provided a memo (included in Appendix A of the voting memo) with additional information to address a comment raised at the in-person meeting. The new analysis compared the look back period of four months versus five months, and did not find a change in results.
- Also in response to Committee members' concern about clustering, the developer noted that the states were intentionally rank ordered by informed coverage rate and purposely appeared to be similar. If the states were ranked by poverty level, the apparent similarities were not present.
- During the call, Committee members discussed their previous concerns about children who were on the cusp of income eligibility. The developer explained that while it did not have access to the incomes of individuals, an analyses was conducted using the average income on a state level. The developer noted that analyses showed that the metric is stable across income levels across states. The developer noted that rates of reenrollment are largely policy-driven (i.e., how easy or difficult it is to reenroll) rather than by income, and that improvements in performance can be made by making it easier to re-enroll.
- Also during the post-comment call, a Committee member inquired about the intended use of the measure. The developer noted that this could be used by states to measure improvement after implementing initiatives. The measure also helps to identify states that report rates much lower or much higher than other states for closer examination.
- Currently, the measure relies on presumptive eligibility for Medicaid; the Committee and developer agreed that if eligibility changes, the measure would need revision.
- After its discussion, the Committee re-voted on the Reliability criterion, which had not achieved consensus during the in-person meeting; the measure passed this criterion.
- Following that vote, Committee members briefly discussed unintended consequences should the measure be used for rewards or penalties; but ultimately agreed this measure is no different than any other measure that has intrinsic errors, and that as long as the range of error is clearly defined, the measure can be useful. The Committee voted on an overall recommendation for endorsement, and agreed the measure should be recommended.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

3166 Antibiotic Prophylaxis Among Children with Sickle Cell Anemia

Submission | Specifications

Description: The percentage of children ages 3 months to 5 years old with sickle cell anemia (SCA, hemoglobin [Hb] SS) who were dispensed appropriate antibiotic prophylaxis for at least 300 days within the measurement year.

Numerator Statement: The numerator is the number of children ages 3 months to 5 years old with SCA (Hb SS) who were dispensed appropriate antibiotic prophylaxis for at least 300 days within the measurement year.

Denominator Statement: The denominator is the number of children ages 3 months to 5 years with SCA (Hb SS) within the measurement year.

Exclusions: There are no denominator exclusions.

Adjustment/Stratification: No risk adjustment or risk stratification/NA

3166 Antibiotic Prophylaxis Among Children with Sickle Cell Anemia

Level of Analysis: Health Plan

Setting of Care: Other

Type of Measure: Process

Data Source: Claims (Only)

Measure Steward: QMETRIC - University of Michigan

STANDING COMMITTEE MEETING [03/02/2017]

1. Importance to Measure and Report: <u>The measure meets the Importance criteria</u>

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

1a. Evidence: **9-H; 13-M; 1-L; 0**-I; 1b. Performance Gap: **18-H; 4-M; 1-L; 0-I**

<u>Rationale</u>:

- The Committee acknowledged that the measure addresses a medical condition affecting a relatively small population, children with sickle cell anemia (SCA), but recognized that its focus has significant effects on their health outcomes, including high mortality rates.
- The Committee noted that several of the studies used for evidence were relatively old and might not include the impact of pneumococcal vaccination, which has become more prevalent. It was further noted, however, there is still infection risk for this population despite the vaccines developed since the studies.
- The Committee found that there was significant variation in performance between states (the developer reports a range of 5.7-36%), although significant racial or ethnic disparities do not exist. It also was noted there has not been an increase in antibiotic prophylaxis over time, which the developer and Committee agreed reflected a lack of work on improvement in this area. The Committee agreed there is significant room for improvement.

2. Scientific Acceptability of Measure Properties: <u>The measure meets the Scientific Acceptability criteria</u>

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

2a. Reliability: **7-H; 16-M; 0-L; 0-I** 2b. Validity: **9-H; 14-M; 0-L; 0-I**

<u>Rationale</u>:

- The measure developer provided empirical testing for reliability at the performance data element and measure score levels and face validity and empirical testing of the measure score for validity.
- The developers conducted its analysis at the score level using Medicaid claims reported to CMS for Medicaid enrollees within the state of Michigan (2007-2011), as well as MAX data for all Medicaid claims reported to CMS for Medicaid enrollees within six state Medicaid programs with moderate to high prevalence of sickle cell anemia: Florida, Illinois, Louisiana, Michigan, South Carolina and Texas (2005-2010). Committee members felt the testing sample size was sufficient.
- Committee members supported the aspect of the specifications capturing patients on antibiotics for 10 out of 12 months in order to account for the time it takes to visit the doctor and have a prescription filled.
- One Committee member noted concern about how the measure would capture additional antibiotics needed to treat breakthrough infections. In addition, Committee members noted that most of the evidence provided was related to penicillin and not the full spectrum of available antibiotics. The developer explained the measure includes oral antibiotics only, not antibiotics delivered via injection.
- The Committee discussed whether the benefits of the measure outweigh the potential risk of increasing antibiotic resistance. The developer clarified that its analysis showed that the patients were no more likely to develop antibiotic resistant infections. In addition, the developer stated it

considered the possibility that the child could be allergic to penicillin, and accounted for this in the measure specifications by including erythromycin. The developer noted that it did not account for complications due to constant use of antibiotics, such as post-secondary infections with *C. Difficile*.

- Some Committee members expressed concern that the measure excludes patients with SCA who have fewer than three claims, potentially missing a significant opportunity to address a gap in care for an underserved population. Other Committee members noted that the reliability increased significantly by specifying the inclusion of patients with three claims and not those with a single SCA-related claims. One Committee member, however, noted that diagnosis is often fairly straightforward and could possibly be identifiable through a single visit to a clinician.
    - The developer confirmed the specification for three claims was necessitated to ensure the measure's validity.
    - The developer emphasized the need for three claims by explaining changes in sensitivity and specificity of three different case definitions the developer used to identify children. The developer also noted that the measure uses three claims that were broadly associated with sickle cell anemia, which could include claims for durable medical equipment or a laboratory visit, not thee claims just for visits to a primary care provider or hematologist.
    - The Committee expressed concerns that pharmacy claims data may not be complete. The developer stated that the data for days of medication supplied was populated relatively well, with an average of 186 days (counting refills) through the entire 12-month study period. The developer also noted that very few children had less than 50 days' supply.
    - The Committee ultimately concurred that using three claims resulted in a significant reduction in false positives and only a small exclusion of true positives.
- Committee members suggested that the developer consider exclusions for patients with co-morbid conditions (organ transplant, cancer, or other immunosuppressive medications such as steroids). Committee members also noted that once a month shots for Bicillin (listed under NHLBI guidelines as an approved method of prophylaxis) is not a method captured in the measure.
- A Committee member noted the high correlation between the prescription and dispensation of antibiotics based on the administrative data.
- One Committee member was concerned that there may be data issues with care provided across state lines, which might affect the reliability of the measure—i.e., patients may need to travel across state lines to see the closest specialist or children's hospital. Another Committee member clarified, however, that Medicaid programs pay across state lines, and related claims would go to the state where the child is covered. The Committee also discussed how some children will see their primary care physician and others will go to a hematologist for their SCA-related care; it wanted to ensure this scenario would not cause data issues. The Committee was assured all appropriate claims would be included.

3. Feasibility: 17-H; 6-M; 0-L; 0-I

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

Rationale:

- The Committee agreed that this measure was highly feasible, since the data elements required for the measure are routinely generated and used during care delivery, and all data elements used in the measure are defined fields in electronic claims.

4. Usability and Use: 12-H; 11-M; 0-L; 0-I

3166 Antibiotic Prophylaxis Among Children with Sickle Cell Anemia

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

Rationale:

- Committee members believed this measure would be a useful addition for evaluating Medicaid programs, as this condition is a significant issue for Medicaid enrollees.
- Committee members supported the use of the measure to assess the ability of organizational structures to ensure that patients have the medicines they need.
- This is a new measure, so it is not currently in use. However, the developer states it is working with the New York Medicaid program to implement the measure, and Committee members agreed it would be a good health plan- or state-level measure.

5. Related and Competing Measures

- No related or competing measures noted.

Standing Committee Recommendation for Endorsement: **23-Y; 0-N**

Rationale

- The Committee agreed this measure meets the NQF criteria for endorsement.

6. Public and Member Comment: April 12-May 11, 2017

- One commenter wrote to support endorsement of this measure. A second commenter acknowledged the importance of assessing antibiotic prophylaxis among children with Sickle Cell Anemia, but also noted difficulty in attributing outcomes within these areas to specific providers and experiences. The commenter encouraged NQF to further examine and identify measures at the clinic/systems level that offer more specificity about appropriate "antibiotic prophylaxis."

Committee response:

- Thank you for providing this comment. This measure is specified at the plan level. The Committee will consider measures at the clinic/system levels as a gap for future measure development and has added the suggestion to the list of gaps in the report.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

Measures Not Recommended

2816 Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma

Submission

Description: This measure estimates the proportion of emergency department (ED) visits that meet criteria for the ED being the appropriate level of care, among all ED visits for identifiable asthma in children and adolescents.

Numerator Statement: The numerator is the number of eligible asthma ED visits in the random sample that also satisfy at least one of the explicit criteria to indicate that the ED is an appropriate level of care. Distinct numerators are reported for children ages 2-5, 6-11, 12-18, and optionally, 19 - 21.

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

2816 Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma

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

Exclusions: ED visits that are already in the sample OR Children that fall outside of specified age range of 2-21 OR who do not meet time enrollment criteria OR do not meet identifiable asthma prior to the ED visit, OR children with concurrent or pre-existing COPD, Cystic Fibrosis or Emphysema.

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

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

Stratify by age group (reporting entity should specify whether to use age at month of qualifying event or age on first day of reporting year):

•Age 2-5 years (second birthday to the day before the 6th birthday);

•Age 6-11 years (sixth birthday to the day before the 12th birthday);

•Age 12-18 years (twelfth birthday to the day before the 18th birthday); and

•Age 19-21 years (nineteenth birthday to the day before the 21st birthday).

Age strata are to be reported distinctly and not combined.

Optional stratifications require data elements such as:

•Race/Ethnicity

•Insurance type (Public, Commercial, Uninsured)

•Benefit type (if insured): HMO, PPO, Medicaid Primary Care Case Management

(PCCM) Plan, Fee for Service (FFS), other

•Zip code, state and county or equivalent area of parent/caregiver's residence. Record FIPS if available

Stratification variables details

•Race/Ethnicity: Hispanic, Non-Hispanic Black, Non-Hispanic White; Non- Hispanic Asian/Pacific Islander, other Non-Hispanic

•Public vs Commercial (Private Insurance).

•HMO vs PPO vs FFS vs PCCM vs other; Within Medicaid, States may ask for reporting of FFS vs Managed Care or other relevant enrollment categories (e.g., TANF, SSI).

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

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

2816 Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma

level data to study frontier health. Inclusion of UIC 8 would make the analysis more sensitive to including frontier areas but at a meaningful cost in specificity.

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

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

•Identify the Level of Poverty in the parent or primary caregiver's county of residence. The percent of all residents in poverty by county or county equivalent are available from the US Department of Agriculture at http://www.ers.usda.gov/data-products/county-level-data-sets/download- data.aspx. Our stratification standards are based on 2011 US population data that we have analyzed with SAS 9.3. Using parent or primary caregiver's state and county of residence (or equivalent) or FIPS code, use the variable PCTPOVALL_2011 to categorize into one of 5 Strata:

o Lowest Quartile of Poverty if percent in poverty is <=12.5%

o Second Quartile of Poverty if percent in poverty is >12.5% and <=16.5%

o Third Quartile of poverty if percent in poverty is >16.5% and <=20.7%

o First Upper Quartile (75th-90th) if percent in poverty is >20.7% and <=25.7%

o Second Upper Quartile (>90th percentile)

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

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

Level of Analysis: Facility, Health Plan

Setting of Care: Emergency Department, Hospital

Type of Measure: Outcome

Data Source: Claims (Only), EHRs Hybrid, Paper Records

Measure Steward: University Hospitals Cleveland Medical Center

STANDING COMMITTEE MEETING [03/02/2017]

1. Importance to Measure and Report: <u>The measure did not reach consensus on the Importance criteria</u> (1a. Evidence, 1b. Performance Gap)

1a. Evidence: **12-Pass; 10-No Pass** (consensus not reached); 1b. Performance Gap: **2-H; 18-M; 2-L; 0-I**

<u>Rationale</u>:

- This measure was originally reviewed by the Pulmonary Committee as a process measure. The measure did not pass Evidence during that review, and that Committee suggested it be resubmitted as an outcome measure. Accordingly, for this project the developer revised and resubmitted the measure as an outcome measure.
- The developer explained this measure is an attempt to recognize there are various reasons for use of the emergency department (ED) for asthma treatment, some of which are appropriate and others that are a failure to manage the patient's asthma.
- Committee members noted that the measure does not account for factors outside the control of the facility or plan, such as medication compliance. The Committee also flagged confounding factors that can influence rates, but that are actually about access to care—e.g., shorter waits in the ED than in primary care clinics for Medicaid patients.

2816 Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma

- Committee members noted the measure is an appropriate use measure, but felt it was more of an overuse measure, since it cannot assess patients who should have gone to the ED, but did not; they felt this was potentially a far more dangerous outcome. The developer noted that it is important that going to the ED not be seen as overuse, automatically, as there absolutely are cases when it is the appropriate level of care. Members agreed a better score on the measure could be attributed to children visiting the ED instead of their PCP or specialist. Both the developer and Committee agreed that performance on this measure could be better assessed by pairing it with a measure estimating how much the ED is used.
- Committee members noted that the measure's construction makes it a problem for the ED to solve, but the problem is actually a system function measure: EDs accept all arrivals and the prevention of visits should fall to other providers (PCPs or specialists). Despite this observation, Committee members noted this is a plan-level measure and there are things a plan can do to reduce inappropriate ED visits, e.g., increasing access and compliance to asthma medications and improving access to primacy care. Ultimately, the Committee agreed that there are modifiable processes that can influence the outcome (appropriate ED visits for asthma), such as ensuring children are assigned to a specialist.
- Committee members agreed there is a gap in care and disparities; Hispanic children and uninsured children showed higher rates of questionable ED use, and Hispanic and African American children have higher rates of asthma (12.4% and 15.8% respectively).

2. Scientific Acceptability of Measure Properties: <u>The measure does not meet the Scientific Acceptability criteria</u>

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

2a. Reliability: **X-H; X-M; X-L; X-I** 2b. Validity: **N/A-H; 1-M; 17-L; 5-I**

<u>Rationale</u>:

- The developer used data element level validity testing, which may be used under NQF guidance to assess both Reliability and Validity, so the Committee did not vote separately on Reliability.
- Committee members had concerns about the appropriateness criteria, noting that some needed clarification (e.g. ,when oxygen saturation should be recorded, on presentation or after treatment) and that others could be easily gameable (e.g., ordering an ABG on every patient). They also noted that referral by a PCP would be considered acceptable under the criteria for the measure, but this could actually indicate poor quality/inappropriate care provided by the PCP. Committee members found the list was subjective and noted that some of the clinical indicators (such as labored breathing) are subjective as well; they also noted that some of the appropriateness indicators are common and others rare.
- Committee members questioned whether the developer could have used this denominator and the numerator for #3189 to create a measure of severity; they also suggested this could be useful as a population measure. A Committee member suggested that a measure that looks at the rate of unnecessary ED use per 100 child years of children with asthma may be more effective than using ED visits as the denominator.
- A Committee member noted that the measure is specified similar to a HEDIS measure, so it should be able to be collected reliably.
- Overall, Committee members expressed a number of concerns about scientific acceptability, including: the measure has only been tested in one hospital (testing cannot demonstrate meaningful differences among institutions); all appropriateness criteria items were not tested at the single institution, since the data elements for every item were not used in the ED at the testing institution; and the specifications permit variability in the use of pharmacy data due to

availability differences. Based on these concerns, the measure did not pass Validity due to insufficient testing.

- Committee members also noted that, while the measure is scored at the patient level, it was tested at the item level; it acknowledged that, after training, the kappas were generally good.

3. Feasibility: X-H; X-M; X-L; X-I

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

Rationale:

-

4. Usability and Use: X-H; X-M; X-L; X-I

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

Rationale:

-

5. Related and Competing Measures

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

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

Since the measure was not recommended, none of these were discussed.

Standing Committee Recommendation for Endorsement: Did not pass Validity

Rationale

- The Committee did not recommend the measure because it did not pass Validity due to insufficient testing.

6. Public and Member Comment: April 12-May 11, 2017

- One commenter noted that this and similar measures proposed by this measure steward were critically appraised by the NQF Pulmonary and Critical Care Standing Committee and agreed with the concerns raised by the NQF Pulmonary and Critical Care Standing Committee; the Pediatric Committee did not discuss this issue since the discussion did not progress to that aspect of validity, given the other concerns. The Pulmonary Committee's comments, which were accessed at the NQF website (http://www.qualityforum.org/Publications/2016/10/Pulmonary_and_Critical_Care_2015-2016_Final_Report.aspx) are:

  "The Committee raised concern about the lack of stratification by risk. While the developer stratified by age, the Committee expressed concern about clinical differences across the age spectra, especially in the first six years of life, which are not accounted for by the measure. The Committee also noted that while the developer provided for stratification by race, it did not address demographic and environmental factors that impact race (e.g., location), which can affect patient risk and quality of care… The Committee discussed the lack of stratification by risk leading to misinterpretation of results as a potential unintended consequence if the measure is implemented… Noting differences in rates, the Committee was concerned with the lack of adjustment for sociodemographic factors (SDS)."

2816 Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma

- **:** We have submitted this to the Pediatric Committee in part because of its greater sensitivities to the issues specific to children and in this case asthma in children. Nearly half of U.S. children are covered by public health insurance programs. Equity of outcomes across race and social class is a preeminent concern and value in child health, especially for asthma. As the internationally accepted NHLBI guidelines states, "As a general rule, patients with well-controlled asthma should have:
  - Few, if any, asthma symptoms.
  - Few, if any, awakenings during the night caused by asthma symptoms.
  - No need to take time off from school or work due to asthma.
  - Few or no limits on full participation in physical activities.
  - No emergency department visits.
  - No hospital stays.
  - Few or no side effects from asthma medicines."

- Further, it is not clear whether those stressors that increase asthma burden are likely to increase or to decrease the level of appropriateness of ED use for asthma. Cogent arguments can be made in either direction, or for not at all.

  Measure 2816, Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma, is stratified by age, specifically the measure is reported for children ages 2-5,6-11, 12-18, and optionally, 19-21. This is because asthma and its management are related both to the child's age and stage of development. Hence comparing performance in young children is very different from performance in adolescents.

  Further, this measure of appropriateness is best interpreted in light of other measures, such as the rate of ED use for asthma. High levels of appropriateness may suggest a highly functional primary care and outpatient approach to asthma. However high levels of appropriateness and high levels of utilization may together suggest that asthma outcomes form outpatient management are not as desired. Hence, this measure informs but is not dispositive without other data.

  Our formal RAND style panel of national experts did not recommend risk adjustment by race or social class. They recommend stratification by the age groups indicated below. Environmental differences may produce unequal burdens on various health plans, but the field's capacity to discriminate and risk adjust in that manner is of uncertain value and such data for adjustment are neither readily available, nor is there a consensus on what and when and how to adjust for such exposure.

  Establishment of asthma control should occur from an early age. Because of challenges in identifying asthma before the age of 2, we have not included this age group in our specification.

  For purchasers who are interested in stratification beyond race and age we provide OPTIONAL specifications that allow them to ask health plans to incorporate additional stratification in the measure (e.g. insurance status, county rates of poverty, and rurality/urbanicity). Contracting health plans can negotiate with purchasers and other accountability agencies to demonstrate stratified performance if they so desire.

  This measure requires stratification by the following age groups:

  - Age 2-5 years (second birthday to the day before the 6th birthday);

2816 Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma

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

These age strata are to be reported distinctly and not combined for reasons noted above.

This measure has optional stratifications for the following that can be determined by the reporting agency to use all or none, as appropriate:

- Race/Ethnicity: Hispanic, Non-Hispanic Black, Non-Hispanic White; Non- Hispanic Asian/Pacific Islander, other Non-Hispanic
- Insurance type (Public, Commercial, Uninsured)
- Benefit type (if insured): HMO, PPO, Medicaid Primary Care Case Management (PCCM) Plan, Fee for Service (FFS), other relevant enrollment categories (e.g., TANF, SSI)
- Urban influence codes: Identify the Urban Influence Code or UIC. (2013 urban influence codes available at: http://www.ers.usda.gov/data-products/urban- influence-codes.aspx#.UZUvG2cVoj8). Use parent or primary caregiver's place of residence to determine UIC. State and county names can be linked or looked up directly or zip codes can be linked to county indirectly, using the Missouri Census Data Center (http://mcdc.missouri.edu/). These data will link to county or county equivalents as used in various states.
- Urban Influence Codes (UIC) have been developed by the USDA to describe levels of urbanicity and rurality. While each UIC has its own meaningful definition, some researchers choose to aggregate various codes. Well regarded schemas for aggregation of codes include Bennett and colleagues at the South Carolina Rural Research Center. Their aggregation scheme brings together Codes 1 & 2 as Urban; 3,5, & 8 as micropolitan rural; 4,6, & 7 as rural adjacent to a metro area; and 9, 10, 11, & 12 as remote rural. We acknowledge that UIC 5 (adjacent rural area) may appropriately be aggregated with 4,6,&7 as rural. Frontier health care may be approximated by analysis of the remote rural categories (UIC 9, 11 and 12). Alternatively, Gary Hart, Director of the Center for Rural Health at the University of North Dakota School of Medicine & Health Science suggests that UIC 9-12 is the best overall approach to using county level data to study frontier health. Inclusion of UIC 8 would make the analysis more sensitive to including frontier areas but at a meaningful cost in specificity.
- Those interested in care specific to large cities may wish to aggregate the rural area and analyze UIC 1 and 2 separately.
- When stratifying by urbanicity or UIC, the reporting and accountability entities should specify clearly what if any aggregating schema was used.
- Identify the Level of Poverty in the parent or primary caregiver's county of residence. The percent of all residents in poverty by county or county equivalent are available from the US Department of Agriculture at http://www.ers.usda.gov/data-products/county-level-data-sets/download- data.aspx. Our stratification standards are based on 2011 US population data that we have analyzed with SAS 9.3. Using parent or primary caregiver's state and county of residence (or equivalent) or FIPS code, use the variable PCTPOVALL_2011 to categorize into one of 5 Strata:
  - o Lowest Quartile of Poverty if percent in poverty is <=12.5%
  - o Second Quartile of Poverty if percent in poverty is >12.5% and <=16.5%
  - o Third Quartile of poverty if percent in poverty is >16.5% and <=20.7%
  - o First Upper Quartile (75th-90th) if percent in poverty is >20.7% and <=25.7%
  - o Second Upper Quartile (>90th percentile)

2816 Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma

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

To summarize:
Appropriateness of ED visits is a new construct for pediatric asthma. As such, there are no pre-existing data to suggest a disparate burden of either appropriate or inappropriate ED visits by socioeconomic class or by health plans caring for them. The NHLBI guideline is clear in articulating the expectation that outcomes should be equally good across the general population of individuals regardless of who they are and even how severe their asthma is (obviously there are true exceptions here, but they would not be well accounted for in any risk adjustment or stratification schema that we have ever seen). The Pediatric Committee is in a better position to understand and appreciate the implications of all of this for children and to incorporate such insights into their evaluation of this measure (and the similar rate measure).

The lack of required stratifications by risk does not lead to misinterpretation of results as a potential unintended consequence if the measure is implemented. In fact, this measure is specified to give flexibility to plans and to purchasers to respond to local conditions and needs by using stratification as needed and desired to compare performance within specified strata. These are desirable attributes for child health quality measures.

Developer Request for Reconsideration:

- At the in-person meeting, for measure #2816, the Committee did not reach consensus on Evidence, the measure passed on Gap, and did not pass on Reliability. The developer requested reconsideration of this measure on the grounds of additional testing information provided. The additional materials are in Appendix B of the voting memo. While the developer has requested reconsideration for both this measure and #3189, it did not provide a separate, specific rationale for this measure. The developer noted that the data provided for #3819 also informs this measure; no additional information was provided specifically related to evidence or appropriateness.

Committee response:

- The Committee reviewed the new material prior to the call, and after discussion, agreed that the new information was not sufficient to reconsider the measure. Specifically, the Committee agreed the new information did not address the issues raised previously with the validity of the numerator construct (i.e., the measure was still only tested at one institution); the Committee also agreed its concerns about the evidence were not adequately addressed.
- Requests for reconsideration require greater than 60% of the Committee voting for reconsideration. The Committee unanimously voted not to move forward with the request (17 no votes) so the measure remains not recommended.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

3189 Rate of Emergency Department Visit Use for Children Managed for Identifiable Asthma: Visits per 100 Child-years

Submission

Description: This measure estimates the rate of emergency department visits for children ages 2 – 21 who are being managed for identifiable asthma, using specified definitions. The measure is reported in visits per 100 child-years.

Numerator Statement: The numerator estimates the number of emergency department (ED) visits for asthma among children being managed for asthma. To enhance validity, a numerator event may be identified either as an ED visit or as a hospitalization.

Denominator Statement: The denominator represents the person time experience among eligible children with identifiable asthma. Assessment of eligibility is determined for each child monthly. The total number of child months experienced is summed and divided by 1200 to achieve the units of 100 child years for the denominator.

Assessing eligibility for the denominator requires 2 years of data, the reporting year and the 12 month period before the reporting year. (See Appendix 1, Figure 1)

Exclusions: Children with specified concurrent or pre-existing diagnosis and children who have not been consecutively enrolled in the reporting plan for at least three months, as specified in the details section.

Adjustment/Stratification: Other In order to allow for more granular comparisons this measure is specified to be stratified. Stratification for risk adjustment of this measure would not be justified by the literature. Although epidemiological findings support our stratification schema, no biological evidence exists to support intrinsic correlation of ED rates with stratification variables. This measure calls for stratification by age group, by race/ethnicity, and by age group and race/ethnicity. Several additional stratifications are recommended but optional. These may be required by the accountability entity or reported by the reporting entity. These variables include rurality/urbanicity and county level of poverty.

Age groups are 2-5, 6-11, 12-18, and 19-20, each inclusive. (reporting entity should specify whether to use age at month of qualifying event or age on first day of reporting year)

Race/ethnicity should incude White non-Hispanic, Black non-Hispanic, and Hispanic as well as other groups as requested by the accountability entity and consistent with current HHS usage.

For social demographic stratification: identify County equivalent of child's residence. If County and State or FIPS code are not in the administrative data, the zip codes can be linked to County indirectly, using the Missouri Census Data Center (http://mcdc.missouri.edu/). These data will link to County or County equivalents as used in various states.

i.Identify the Urban Influence Code (1) or UIC for the county of child's residence. (2013 urban influence codes available at: http://www.ers.usda.gov/data-products/urban-influence- codes.aspx#.UZUvG2cVoj8).

ii.Identify the Level of Poverty in the child's county of residence. The percent of all residents in poverty by county or county equivalent are available from the US Department of Agriculture at http://www.ers.usda.gov/data-products/county-level-data-sets/download- data.aspx. Our stratification standards are based on 2011 US population data that we have analyzed with SAS 9.3. Using child's state and county of residence (or equivalent) or FIPS code, use the variable PCTPOVALL_2011 to categorize into one of 5 Strata:

a.Lowest Quartile of Poverty if percent in poverty is <=12.5%

b.Second Quartile of Poverty if percent in poverty is >12.5% and <=16.5%

c.Third Quartile of poverty if percent in poverty is >16.5% and <=20.7%

d.First Upper Quartile (75th-90th) if percent in poverty is >20.7% and <=25.7%

e.Second Upper Quartile (>90th percentile)

3189 Rate of Emergency Department Visit Use for Children Managed for Identifiable Asthma: Visits per 100 Child-years

iii.Categorize age by age at the last day of the month that ends the assessment period. Aggregate into age categories 2-4, ages 5 through 11, ages 12-18, ages 19-21.

iv.Categorize Race/Ethnicity as Hispanic, Non-Hispanic White, Non- Hispanic Black, Non-Hispanic Asian/Pacific Islander, and Non-Hispanic Other

v.Categorize Insurance Type as Private (Commercial), Public, None or Other

vi.Categorize benefit type as HMO, PPO, FFS, PCCM, or Other

Level of Analysis: Population : Community, County or City, Health Plan, Population : Regional and State

Setting of Care: Hospital : Acute Care Facility, Emergency Department, Hospital, Other

Type of Measure: Outcome

Data Source: Claims (Only), Claims (Other)

Measure Steward: University Hospitals Cleveland Medical Center

STANDING COMMITTEE MEETING [03/02/2017]

1. Importance to Measure and Report: <u>The measure meets the Importance criteria</u>

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

1a. Evidence: **21-Pass; 1-No Pass;** 1b. Performance Gap: **9-H; 13-M; 1-L; 0-I**

<u>Rationale</u>:

- This measure was originally submitted to the Pulmonary Committee and was not recommended. It has been revised to address issues raised by that Committee and resubmitted. It is an outcome measure based on the rationale that accessible, high-quality primary care reduces the need for ED visits for persistent asthma, which are an undesirable outcome.
- The Committee agreed asthma is a serious condition, many ED visits should be preventable, and the link to the evidence for the measure is strong.
- Committee members raised concerns, however, about the specifications' lower age limit of two years, noting that accurate diagnoses of asthma (versus persistent wheezing due to viral infections) are challenging at that age; they suggested ages three or four years would be a better lower limit.
- Committee members also noted that asthma is strongly influenced by environmental and social factors out of the control of providers or plans, and that some ED visits would only be preventable with social interventions—i.e., asthma outcomes cannot solely be attributed to the care provided. The developer explained that it constructed and tested the measure using chart review data, so does not have data available on environmental triggers, etc., but that the National Heart, Lung, and Blood Institute guidelines indicate that children in more challenging circumstances need to be managed more aggressively to prevent ED visits (a goal for all children).
- Committee members generally agreed there is evidence of disparities in care, and the developer's testing found differences in performance by race, urbanity, and quartile of poverty.

2. Scientific Acceptability of Measure Properties: <u>The measure does not meet the Scientific Acceptability criteria</u>

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

2a. Reliability: **N/A-H; 1-M; 4-L; 18-I** 2b. Validity: **X-H; X-M; X-L; X-I**

<u>Rationale</u>:

- This is a claims-based measure. The developer submitted additional reliability testing the morning of the Committee's in-person meeting, since the original submission had not met NQF's minimum standards. The Committee was unable to review the data prior to its discussion. The

3189 Rate of Emergency Department Visit Use for Children Managed for Identifiable Asthma: Visits per 100 Child-years
developer explained it had examined reliability a number of ways, with different plans as an index, and the measure scores differentiated performance. • In addition to the insufficiency of the original data and the lack of time to review the new data, the Committee requested additional information and/or testing, such as confirmation that the claims data would match a chart review. The Committee agreed the reliability testing was insufficient. • The Committee also remained concerned about the denominator time interval specified by the measure. The denominator for this measure is, "The person time experience among eligible children with identifiable asthma. Assessment of eligibility is determined for each child monthly. The total number of child months experienced is summed and divided by 1200 to achieve the units of 100 child years." The Committee requested additional clarification, noting that an ED visit in February would include 13 months to be diagnosed with asthma, but the time interval for a visit in December would be 20 months. It requested additional information on whether this might bias the results, especially given the seasonality of asthma. • Since the measure did not pass Reliability, the Committee did not discuss the remaining criteria. The Committee agreed, however, to review and re-discuss the measure on the post-comment call. • The Committee also provided other high-level feedback to the developer about the specifications, including requests to confirm whether including or excluding bronchitis affects the measure, and the impact of excluding short-acting beta agonists.
3. Feasibility: X-H; X-M; X-L; X-I *(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c.Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)* Rationale: •
4. Usability and Use: X-H; X-M; X-L; X-I *(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)* Rationale: •
5. Related and Competing Measures The Developer did not include information on any of the related or competing measures. However, NQF staff identified the following measures that may be related and/or competing. • 0047: Asthma: Pharmacologic Therapy for Persistent Asthma • 0728: Asthma Admission Rate (PDI 14) • 1800: Asthma Medication Ratio • 2414: Pediatric Lower Respiratory Infection Readmission Measure • 2816: Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma (*submitted by the same Developer for review in this project)* Since the measure was not recommended, none of these were discussed.
Standing Committee Recommendation for Endorsement: **Did not pass Reliability** Rationale • The Committee did not recommend the measure because it did not pass Reliability due to insufficient testing.
6. Public and Member Comment: April 12-May 11, 2017

3189 Rate of Emergency Department Visit Use for Children Managed for Identifiable Asthma: Visits per 100 Child-years

- During the comment period, no NQF member or public comments were received, but the developer did submit a request for reconsideration. The supplementary materials are in Appendix B of the voting memo.

  Developer Rationale for Reconsideration: "At the in-person meeting, measure #3189 passed onEvidence and Gap, and was voted insufficient for Reliability. In general, the sense of the group [the Committee] at the in-person meeting was that measure #3189 is a very viable measure, but having to conform to the NQF procedure, the group required a little bit more data, which is provided herein:

  1) Reliability
  2) Inclusion/Exclusion
  3) Pharmacy Data
  4) Race Disparities
  5) Data Element Validity"

- **Committee response:** The Committee reviewed the new material prior to the call. During the call, after the developer presented its request for reconsideration, the Committee questioned the look-back period, noting that for a measurement month in January, the look-back is 12 months, but for November, the look-back period is 23 months. The developer explained that longer look-back periods are more reliable. Committee members did not have additional questions, and voted on whether to reconsider the measure; the vote was nine Yes to reconsider and eight No, against reconsideration. Because the threshold for reconsideration is greater than 60% voting yes, the Committee did not reconsider the measure.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

3219 Anticipatory Guidance and Parental Education

Submission

Description: This measure is used to assess the degree to which pediatric clinicians discussed key recommended anticipatory guidance and parental education (AGPE) topics. Necessarily, anticipatory guidance questions vary by child age. Anticipatory guidance for children ages 0-9 months include 15 questions. Anticipatory guidance for children ages 10-18 months includes 16 questions; and anticipatory guidance for children ages 19-48 months includes 16 questions.

Numerator Statement: The numerator is the number of parents who had a well child visit within the last 12 months and who indicated that they received anticipatory guidance and education, that their questions were answered or that they already had the information and did not require anticipatory guidance on that topic.

Denominator Statement: Parents whose children ages 0-48 months who received a well-child visit in the last 12 months and who responded to at least half of the AGPE items (see Attachment A-2 pages 8-10) on the Promoting Healthy Development Survey (PHDS: www.wellvisitsurvey.org)

Exclusions: Unknown and missing values (responses coded missing) are excluded in the data analysis. Approximately 2.6% of parents who started the Online PHDS did not complete the survey (range 0.0-

3219 Anticipatory Guidance and Parental Education

3.3% for top 5 providers with highest number of surveys; see Testing form, pages 23-24 for more detailed information on missing data).

Adjustment/Stratification: No risk adjustment or risk stratification Although no stratification is required, the Promoting Healthy Development Survey (PHDS) includes a number of variables that allow for stratification of the findings by possible vulnerability, should any individual provide have sufficient data (parent responses) to do so. Potential variables for stratification include:

(1) Child demographic characteristics (e.g., the child's age, race);

(2) Child health and descriptive characteristics (e.g., children at high risk for developmental, behavioral or social delays, special health care needs); and/or

(3) Parent health characteristics (e.g., children whose parents are experiencing symptoms of depression)

Level of Analysis: Clinician : Individual

Setting of Care: Clinician Office/Clinic

Type of Measure: Outcome: PRO

Data Source: Other

Measure Steward: Child and Adolescent Health Measurement Initiative

STANDING COMMITTEE MEETING [03/02/2017]

1. Importance to Measure and Report: <u>The measure does not meet the Importance criteria</u>

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

1a. Evidence: (First Vote: 15-Pass; 7-No Pass) **Second Vote: 8-Pass; 14-No Pass**; 1b. Performance Gap: **1-H; 17-M; 1-L; 3-I** (made void by revote)

<u>Rationale</u>:

- This patient-reported outcome performance measure (PRO-PM) assesses, at the clinician level, whether providers gave anticipatory guidance to parents on a number of age-appropriate topics within three domains (physical health, behavior/language/learning, and injury prevention). The score is based on the percent of parents who said either that they received anticipatory guidance, or that they did not need that information, for all of the topics.
- The evidence base, the American Academy of Pediatrics' Bright Futures guidelines, for this measure suggests that a perfect score (discussion of all items) actually leads to better outcomes than a lower score (discussion of some items). Providers receive a report noting which areas were discussed and where they are falling short. However, Bright Futures suggests that just 3-5 items should be covered in each visit, with the premise that the full set is covered over time. Committee members also noted that research shows behavior change is more likely when a few items are focused on rather than a laundry list.
- The Committee raised concerns about whether checklists and surveys are actually helpful for parents, or if they are increasing burden. Committee members noted that parents may find some of the questions challenging or invasive, and that while parents have to put in time answering the survey, they do not get any benefits other than a list of topics they should ask about at the next visit – which may be a year away. Some Committee members, however, noted that a parental perception of what was discussed may be more important than what was actually discussed, as it informs providers that they may not be discussing topics in a meaningful way.
- While some of the Committee agreed there is an action providers can take to affect the outcome, there also were members who wondered if it is actionable by a pediatrician. In particular, Committee members expressed concern that the discussion may not have been with the "final" provider, but could have occurred with another provider. In such a case, actionability to affect the score cannot be linked to the provider sending the survey and being judged by the results.

3219 Anticipatory Guidance and Parental Education

- As a PRO-PM, NQF's algorithm uses the outcome pathway for evidence (pass/no pass)—whether there is a relationship between the measured health outcome and at least one healthcare option. The Committee found this a challenging question and discussed it in detail. Some Committee members were uncomfortable with the measure as an outcome, thinking of it instead as a process measure. Other Committee members noted that even if viewed as a process measure, the measure is based on the Bright Futures guidelines—i.e., the developer had supported the survey instrument and subsequent PRO-PM with the best evidence currently available.
- Committee members discussed the confounding factor of multiple visits, and whether using the tool is improving scores over time versus the relationship building and increase in rapport that happens over multiple visits. The developer explained that, while developing the measure, the focus groups of parents reported that they really liked giving feedback to the providers, and that they (the parents) saw improvement. The developer agreed, however, that the available data could not differentiate whether it was the use of the tool or the relationship building over time that improved performance.
- Committee members agreed there was a performance gap, nothing that there was a large range of performance in the testing data submitted by the developer--the proportion of parents who reported discussion of all anticipatory guidance and parental education topics or reported no need of discussion among unaddressed topics ranged 46.8-84.8% across the top five observed providers; all children averaged 60 percent. The Committee also agreed that there are disparities in performance across race/ethnicity, for some socioeconomic factors, and when there were language barriers.
- Committee members felt additional information on the gap at the provider level would be useful, noting that the raw data for subgroups is less informative than data showing whether providers have differential quality for their higher or lower income patients, etc.
- This measure passed Evidence during the first vote taken, but after additional discussion on Scientific Acceptability, a second vote on Evidence was taken because questions became apparent about the timing of the survey/PRO-PM's administration and the "look-back" period for the measure construct beyond the last visit, which called into question for some whether the measured entity (provider) could take an action to change the score if he/she had not been the provider. Of concern is that questions are constructed, "since your child was born" or "in the last 12 months," and the results sent to the provider may not actually assess care provided, nor actions taken, by the individual being measured. The measure did not pass Evidence during the second vote.

2. Scientific Acceptability of Measure Properties:

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

2a. Reliability: **X-H; X-M; X-L; X-I** 2b. Validity: **X-H; X-M; X-L; X-I**

Rationale:

- Committee members noted discrepancies in the timeline for the questions from the survey, ranging from "since your child was born" to "during the last 12 months". Committee members felt that it would be difficult to hold one provider accountable for the care provided, since the wording could include many providers (including those outside the practice of the provider who sent the survey), and that parents may be asked different questions by different providers during a visit. Committee members also were troubled by the assumption a child would always see the same provider or would even be at the same practice during the entire measurement reference period. They also noted that the more a child visits providers (i.e., for many sick visits), the higher the chance they will receive anticipatory guidance, but it then becomes more difficult to attribute to one provider, as this PRO-PM does. They felt that the results may not

3219 Anticipatory Guidance and Parental Education

echo the care provided by the person receiving the results, and hence that provider could not take action to influence the score. On a related note, they were concerned that the developer had presented the measure as giving feedback to the providers to improve their own care, but with NQF endorsement comes the possibility of measures being used for physician-level accountability, and the way the questions were worded make it challenging to use the measure for individual provider accountability.

- The developer noted that this is the way several CAHPS measures are structured, and the measure is intended to match both CAHPS and questions on the National Survey of Children's Health. One Committee member responded that the CAHPS Patient Centered Medical Home (PCMH) survey is constructed in a way that makes it clear the primary care provider is responsible for all care delivered to a patient (even if he/she is not directly providing the care), and the primary care physician is expected to identify and solve gaps in care. It also was noted that many medical homes are considered at a facility level, rather than an individual clinician (although solo practitioners can also be considered medical homes).
- Committee members noted that originally the survey from which this PRO-PM was derived was endorsed at the state level, and that the survey had been used to compare providers within a practice and practices within plans. It is now intended for scoring at the individual provider level, and concerns were expressed about testing to support application and attribution to an individual provider.
- Committee members also raised a number of additional concerns, noting that it relies on parental recall, and that recall can be inaccurate, especially as time elapses from the visit; there may be differences in responses one day after a visit versus three days, but no information in this regard was provided by the developer nor was data collection standardized to include only responses within a specified number of days. Some Committee members noted, however, that patient-reporting might be better than EHRs in terms of accurately reporting what was discussed during a visit.
- The developer explained that it does not control when the survey is sent, as that is up to the provider—it could go out immediately following a visit or not for some period of time. The Committee felt that the lack of a clear timeframe for sending the surveys and accepting responses was a significant issue with the reliability and validity of the measure.
- Committee members agreed the methods of reliability testing were acceptable, noting that the developer used three different studies, each with an adequate sample and variability in patient populations and acceptable Cronbach's alphas. They noted that test/re-test testing also would have been useful, especially given their concerns around survey timing.
- After extensive discussion on the details of the timing of when the care being asked about was provided, and the ability of the measure to attribute care to the one provider who is receiving the results and, in turn, the ability of that provider to undertake an action to influence the outcome (the threshold for Evidence for a PRO-PM), the Committee requested a revote on Evidence. During the second vote, it did not pass Evidence.
- Committee members were concerned about health literacy and language issues, noting that the survey is only available in English, which they felt was a major issue, and that parents and providers may not be speaking the same language. They also noted the survey instrument is set at an 8th-9th grade reading level, which may present challenges for the parents taking the survey.
- Since the Committee elected to revote on Evidence, and the measure did not pass, no votes were taken on Reliability or Validity.

3. Feasibility: X-H; X-M; X-L; X-I

3219 Anticipatory Guidance and Parental Education

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

Rationale:

-

4. Usability and Use: X-H; X-M; X-L; X-I

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

Rationale:

-

5. Related and Competing Measures

- This measure is related to a set of measures submitted by the same developer for review in this project; all are harmonized.

Standing Committee Recommendation for Endorsement: **Y-X; N-X**

Rationale

- The measure did not pass Evidence due to concerns about the measure's construct of the applicable timeframe and subsequent attribution to a single provider, so it was not recommended.

6. Public and Member Comment: April 12-May 11, 2017

- There were no comments received for this measure.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

3220 Ask About Parental Concerns

Submission

Description: This measure is used to assess the proportion of children whose parents were asked by their child's health care provider if they have concerns about their child's learning, development and behavior.

Numerator Statement: The numerator measures the number of parents who had a well child visit within the last 12 months and who indicated that they were asked about their concerns about their child

Denominator Statement: Children age 3 months to 48 months who received a well-child visit in the last 12 months and whose parents responded to the items Ask About Parental Concerns (see Attachment A-2, page 14) on the Promoting Healthy Development Survey (PHDS: www.wellvisitsurvey.org)

Exclusions: Missing data for the Ask About Parental Concerns questions are excluded from analysis

Adjustment/Stratification: No risk adjustment or risk stratification Although no stratification is required, the Promoting Healthy Development Survey (PHDS) includes a number of variables that allow for stratification of the findings by possible vulnerability, should any individual provide have sufficient data (parent responses) to do so. Potential variables for stratification include:

(1) Child demographic characteristics (e.g., the child´s age, race);

(2) Child health and descriptive characteristics (e.g., children at high risk for developmental, behavioral or social delays, special health care needs); and/or

(3) Parent health characteristics (e.g., children whose parents are experiencing symptoms of depression)

Level of Analysis: Clinician : Individual

3220 Ask About Parental Concerns

Setting of Care: Clinician Office/Clinic

Type of Measure: Outcome: PRO

Data Source: Other

Measure Steward: Child and Adolescent Health Measurement Initiative

STANDING COMMITTEE MEETING [03/02/2017]

1. Importance to Measure and Report: <u>The measure did not reach consensus on the Importance criteria</u>

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

1a. Evidence: **10-Pass; 12-No Pass** (consensus not reached); 1b. Performance Gap: **1-H; 18-M; 3-L; 0-I**

<u>Rationale</u>:

- Committee members noted this question should be asked at every visit, and that a primary care provider is responsible for ensuring that someone is asking this question during every well-child visit. The Committee agreed that evidence exists that the outcome can be influenced by a provider, but as with #3219, expressed concerns with the timing and attribution issues. While the survey may be sent after the 15-month visit, the wording of the question refers, again, to the last 12 months and any provider seen: "In the last 12 months, did your child's doctor or other health provider (could be a general doctor, a specialist, a pediatrician, a nurse practitioner, a physician assistant, a nurse or any one else your child would see for health care) ask if you have concerns about your child's learning, development or behavior?"
- The Committee did not reach consensus on Evidence.
- The Committee agreed there was a gap in performance: The developer's testing data indicated nearly half of parents do not report being asked this question, and there are variations by child's age, race/ethnicity, level of risk for developmental, behavioral, or social delays, respondent education level, birth order, and children's special health care needs status.

2. Scientific Acceptability of Measure Properties: <u>The measure does not meet the Scientific Acceptability criteria</u>

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

2a. Reliability: **0-H; 7-M; 13-L; 2-I** 2b. Validity: **X-H; X-M; X-L; X-I**

<u>Rationale</u>:

- As with #3219, the Committee raised significant reliability issues, specifically around the timing of the survey and lack of standardization in the timeframe to administer the survey/PRO-PM and timing for response completion. The developer noted that in its studies the survey was sent soon after a visit, but the Committee felt the measure could not be used for accountability purposes without more specificity.
- Committee members noted that additional validity testing that would demonstrate parents were actually answering about what happened in a particular practice – as opposed to being asked by a WIC nurse, a school nurse, or ED doctor – they would feel more comfortable with the measure. Committee members felt the current wording confounded the question of which practice the parent may be referring to in his or her response.
- Due to concerns about the timeframe of the questions, when the survey is to be sent, a cut-off time for returned responses, and the inability of the measure to attribute care to one provider, the measure did not pass Reliability and did not move forward.

3. Feasibility: X-H; X-M; X-L; X-I

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

<u>Rationale</u>:

-

3220 Ask About Parental Concerns

4. Usability and Use: X-H; X-M; X-L; X-I

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

Rationale:

-

5. Related and Competing Measures

- This measure is related to a set of measures submitted by the same developer for review in this project; all are harmonized.

Standing Committee Recommendation for Endorsement: **Y-X, N-X**

Rationale

- The measure did not pass Reliability due to concerns about attribution to a single provider, the lack of standardization in the survey administration, and the lack of a cut-off for responding to of the survey, so it was not recommended.

6. Public and Member Comment: April 12-May 11, 2017

- A commenter noted that eliciting parental strengths and needs promotes this learning and affirms and strengthens the role of the family as primary partner in health promotion. Since families most often are responsible for implementing next steps and recommendations, it is important that healthcare professionals listen to and learn from their perspectives. The commenter encouraged NQF to consider this continuum of health in the context of a partnership between families, physicians, and payors. The commenter acknowledged the importance of eliciting parental concerns, and stated it understands the difficulty in attributing outcomes within these areas to specific providers and experiences. The commenter disagreed with assessing parental concerns at the individual level and instead recommends that NQF measure this concept at a clinic/systems level which recognizes team based care. Finally, the commenter proposed a list of measure concepts that could address this measurement gap area.

Committee response:

- Thank you for providing this comment. The issue has been added to the list of measure gaps in this report.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

3221 Family Centered Care

Submission

Description: This measure is used to assess the average percentage of recommended of aspects of family-centered care (FCC) regularly received by the parent from the pediatric clinician. Topics specifically focus on the following components of FCC:

(1) whether the health care provider understands specific needs of child and concerns of parent;

(2) builds confidence in the parent;

(3) explains things in a way that the parent can understand; and

(4) shows respect for a family's values, customs, and how they prefer to raise their child.

Numerator Statement: The numerator measures the number of parents who had a well child visit within the last 12 months and who experienced family centered care in 7 specific areas.

3221 Family Centered Care

Denominator Statement: The denominator is the number of parents with children ages 0-48 months who have completed a well child visit within the last 12 months who answered the Family Centered Care questions on the Promoting Healthy Development Survey (see Attachment A-2, page 12).

Exclusions: Missing data for the Family Centered Care questions excluded from analysis.

Adjustment/Stratification: No risk adjustment or risk stratification Although no stratification is required, the Promoting Healthy Development Survey (PHDS) includes a number of variables that allow for stratification of the findings by possible vulnerability, should any individual provide have sufficient data (parent responses) to do so. Potential variables for stratification include:

(1) Child demographic characteristics (e.g., the child´s age, race);

(2) Child health and descriptive characteristics (e.g., children at high risk for developmental, behavioral or social delays, special health care needs); and/or

(3) Parent health characteristics (e.g., children whose parents are experiencing symptoms of depression)

Level of Analysis: Clinician : Individual

Setting of Care: Clinician Office/Clinic

Type of Measure: Outcome: PRO

Data Source: Other

Measure Steward: Child and Adolescent Health Measurement Initiative

STANDING COMMITTEE MEETING [03/02/2017]

1. Importance to Measure and Report: <u>The measure does not meet the Importance criteria</u>

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

1a. Evidence: **6-Pass; 16-No Pass**; 1b. Performance Gap: **X-H; X-M; X-L; X-I**

<u>Rationale</u>:

- As with the two previous measures, #3219 and #3220, the Committee raised questions about the limited ability of this measure to attribute results to a single provider and, given this, the ability of that provider to influence the score. The Committee further noted that for this particular set of questions, it would be difficult for a physician to receive the results and understand how to improve on the measure. The Committee also expressed concerned about the developer's data showing no improvement in these questions over time among providers who participated in the cited studies
- The Committee questioned the title of the measure, since family-centered care is an approach to care and encompasses much more than what is included in the items included in the measure.
- The Committee also noted that the questions presented were conceptually similar to the previous survey questions discussed in measure #3220: Ask About Parental Concerns.
- A Committee member asked if the automated reporting system and website had launched in February 2017, as planned. The developer stated that the website had not launched yet, but was expected to launch in March 2017.
- Due to the concerns about the measure's ability to attribute care to one provider, it did not pass Evidence and did not move forward.

2. Scientific Acceptability of Measure Properties:

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

2a. Reliability: **X-H; X-M; X-L; X-I** 2b. Validity: **X-H; X-M; X-L; X-I**

<u>Rationale</u>:

-

3. Feasibility: X-H; X-M; X-L; X-I

3221 Family Centered Care
(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c.Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented) Rationale: •
4. Usability and Use: X-H; X-M; X-L; X-I *(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)* Rationale: •
5. Related and Competing Measures •
Standing Committee Recommendation for Endorsement: Y-X; N-X Rationale • The measure did not pass Evidence due to concerns about the inability of the measure to attribute care to one provider, so it was not recommended.
6. Public and Member Comment: April 12-May 11, 2017 • A commenter noted the importance of family experiences and care, but also noted the difficulty in attributing outcomes within these areas to specific providers and experiences. The commenter encouraged NQF to consider mechanisms for family-centered care delivery when examining and testing methods to measure family-centered care given its importance to pediatrics, and encouraged NQF to consider additional measures that assess family-centered care at the clinic/systems level. **Committee Response:** • Thank you for providing this comment. The issue has been added to the list of measure gaps in this report.
7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X
8. Board of Directors Vote: Y-X; N-X
9. Appeals

3222 Assessment of Family Alcohol Use, Substance Abuse and Safety
Submission
Description: This measure is used to evaluate the proportion of children whose parents reported being assessed for one or more of the recommended topics regarding alcohol use, substance abuse, safety, and firearms in the home. **Numerator Statement**: The numerator measures the number of parents who had a well child visit within the last 12 months and who were asked about alcohol use, substance abuse, safety and firearms in the house. **Denominator Statement**: The denominator is the number of parents with children ages 0-48 months who have completed a well child visit within the last 12 months and answered all of the Family Alcohol Use, Substance Abuse and Safety questions on the Promoting Healthy Development Survey(PHDS, see Attachment A-2, page 17). **Exclusions**: Missing data were excluded from the analysis.

3222 Assessment of Family Alcohol Use, Substance Abuse and Safety

Adjustment/Stratification: No risk adjustment or risk stratification Although no stratification is required, the Promoting Healthy Development Survey (PHDS) includes a number of variables that allow for stratification of the findings by possible vulnerability, should any individual provide have sufficient data (parent responses) to do so. Potential variables for stratificationners include:

(1) Child demographic characteristics (e.g., the child´s age, race);

(2) Child health and descriptive characteristics (e.g., children at high risk for developmental, behavioral or social delays, special health care needs); and/or

(3) Parent health characteristics (e.g., children whose parents are experiencing symptoms of depression)

Level of Analysis: Clinician : Individual

Setting of Care: Clinician Office/Clinic

Type of Measure: Outcome: PRO

Data Source: Other

Measure Steward: Child and Adolescent Health Measurement Initiative

STANDING COMMITTEE MEETING [03/02/2017]

1. Importance to Measure and Report: <u>The measure does not meet the Importance criteria</u>

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

1a. Evidence: **7-Pass; 15-No Pass**; 1b. Performance Gap: **X-H; X-M; X-L; X-I**

<u>Rationale</u>:

- The Committee agreed this measure was similar in content and structure to the previous measures presented (#3219, #3220, #3221, #3222). This measure evaluates the proportion of children whose parents report being assessed for three items: alcohol use, substance abuse, and firearms in the home. It can be used by providers to determine the level at which they discuss these issues with the parents.
- The developer indicated that the American Academy of Pediatrics and U.S. Maternal and Child Health Bureau Bright Futures guidelines include recommendations related to assessments of alcohol and drug use, the presence of guns, family violence, and other safety issues in the family.
- The Committee questioned why #3222 and #3223: Family Psychosocial Screening were split into different measures, because these kinds of questions are typically asked together in clinical practice. The developer clarified that #3223 was intended to focus on psychosocial screening and emotional well-being versus other environmental risk factors.
- The Committee again had concerns regarding attribution of performance, so the measure did not pass Evidence and did not move forward.

2. Scientific Acceptability of Measure Properties:

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

2a. Reliability: **X-H; X-M; X-L; X-I** 2b. Validity: **X-H; X-M; X-L; X-I**

<u>Rationale</u>:

-

3. Feasibility: X-H; X-M; X-L; X-I

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

<u>Rationale</u>:

-

4. Usability and Use: X-H; X-M; X-L; X-I

3222 Assessment of Family Alcohol Use, Substance Abuse and Safety

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

Rationale:

-

5. Related and Competing Measures

- This measure is related to a set of measures submitted by the same developer for review in this project; all are harmonized.

Standing Committee Recommendation for Endorsement: **Y-X, N-X**

Rationale

- The measure did not pass Evidence due to concerns about the inability of the measure to attribute care to one provider, so it was not recommended.

6. Public and Member Comment: April 12-May 11, 2017

- There were no comments received on this measure.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

3223 Assessment of Family Psychosocial Screening

Submission

Description: This measure is used to assess the proportion of children whose parents were assessed by a health provider on one or more of the recommended psychosocial well-being topics, including depression, emotional support, changes or stressors in the home, and how parenting is working.

Numerator Statement: The numerator is the number of parents who had a well child visit within the last 12 months and who were asked about psychosocial well-being.

Denominator Statement: The number of parents with children ages 0-48 months who have completed a well child visit within the last 12 months and all answered questions related to the family psychosocial screening scale (see Attachment A-2, page 18).

Exclusions: Missing data are excluded from the analysis.

Adjustment/Stratification: No risk adjustment or risk stratification Although no stratification is required, the Promoting Healthy Development Survey (PHDS) includes a number of variables that allow for stratification of the findings by possible vulnerability, should any individual provide have sufficient data (parent responses) to do so. Potential variables for stratification include:

(1) Child demographic characteristics (e.g., the child´s age, race);

(2) Child health and descriptive characteristics (e.g., children at high risk for developmental, behavioral or social delays, special health care needs); and/or

(3) Parent health characteristics (e.g., children whose parents are experiencing symptoms of depression)

Level of Analysis: Clinician : Individual

Setting of Care: Clinician Office/Clinic

Type of Measure: Outcome: PRO

Data Source: Other

Measure Steward: Child and Adolescent Health Measurement Initiative

STANDING COMMITTEE MEETING [03/02/2017]

3223 Assessment of Family Psychosocial Screening

1. Importance to Measure and Report: <u>The measure does not meet the Importance criteria</u>

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

1a. Evidence: **8-Pass; 14-No Pass**; 1b. Performance Gap: **X-H; X-M; X-L; X-I**

<u>Rationale</u>:

- The Committee found that this measure was similar in construct and evidence to the other related measures previously discussed (#3219, #3220, #3221, #3222). The Committee noted that the questions are part of Bright Futures, and have been endorsed by the American Academy of Pediatrics and the Maternal and Child Health Bureau.
- One Committee member raised some concerns about the impact on parents of being asked about potentially difficult mental health issues, noting that it could be upsetting, and it is not clear what benefit, if any, parents receive from completing the survey.
- Overall, however, the Committee supported the importance of parents being involved in the development of this type of measure.
- As with the other measures, attribution to a single provider continued to be a concern, given the construct of the measure/questions. The measure did not pass Evidence and did not move forward.

2. Scientific Acceptability of Measure Properties:

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

2a. Reliability: **X-H; X-M; X-L; X-I** 2b. Validity: **X-H; X-M; X-L; X-I**

<u>Rationale</u>:

-

3. Feasibility: X-H; X-M; X-L; X-I

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

<u>Rationale</u>:

-

4. Usability and Use: X-H; X-M; X-L; X-I

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

<u>Rationale</u>:

-

5. Related and Competing Measures

- This measure is related to a set of measures submitted by the same developer for review in this project; all are harmonized.

Standing Committee Recommendation for Endorsement: **Y-X; N-X**

<u>Rationale</u>

- The measure did not pass Evidence due to concerns about the inability of the measure to attribute care to one provider, so it was not recommended.

6. Public and Member Comment: April 12-May 11, 2017

- There were no comments received on this measure.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

National Consensus Standards for Pediatric Performance Measures

Consensus Standards Approval Committee Review and Recommendations

July 11-12th, 2017

Co-Chairs:
John Brookey, MD, FAAP
Jeffrey Susman, MD

NQF Staff:
Suzanne Theberge, Senior Project Manager
Kate McQueston, Project Manager
Madison Jung, Project Analyst
Robyn Y. Nishimi, PhD, Senior Consultant

NATIONAL QUALITY FORUM

Pediatric Performance Measures

- For this project, the 23-member Standing Committee evaluated 11 newly submitted measures against NQF's standard evaluation criteria.
  - Measure topics included patient safety, care for children with special health care needs, asthma care, insurance coverage, and a set of measures based on the Promoting Healthy Development Survey.
  - The Committee recommended four measures for endorsement and did not recommend seven measures.

Pediatric Performance Measures

- Measures Recommended for Endorsement at the in-person meeting:
  - #3136: GAPPS: Rate of preventable adverse events per 1,000 (Center of Excellence for Pediatric Quality Measurement, Boston Children's Hospital)
  - #3153: Continuity of Primary Care for Children with Medical Complexity (Seattle Children's Research Institute)
  - #3166: Antibiotic Prophylaxis Among Children with Sickle Cell Anemia (QMETRIC, University of Michigan)

Consensus Not Reached Measure

- Consensus Not Reached at the in-person meeting:
  - 3154: Informed Participation [*formerly Informed Coverage*] (Children's Hospital of Philadelphia)
    - Consensus not reached on Reliability

Pediatric Performance Measures Before Member and Public Commenting

	New	Total
Measures submitted	12	12
Measures recommended for endorsement	3	3
Measures where consensus is not yet reached	1	1
Measures not recommended for endorsement	7	7
Measures withdrawn from consideration	1	1
Reasons for not recommending	Importance – 4 Scientific Acceptability – 3 Overall – 0 Competing Measure – 0	

Pediatric Performance Measures

- Recommended measures:
  - **Patient Safety – 1 measure**
    - Type of measures: 0 intermediate outcome; 0 process; 0 composite; 1 outcome
  - **Continuity of Care – 1 measure**
    - Type of measure: 1 structure; 0 intermediate outcome; 0 process; 0 composite; 0 outcome
  - **Sickle Cell Anemia – 1 measure**
    - Type of measures: 0 intermediate outcome; 1 process; 0 composite; 0 outcome
  - **Insurance coverage– 1 measure**
    - Type of measures: 0 intermediate outcome; 0 process; 0 composite; 1 outcome

Overarching Issues:

- Measures for Accountability vs. Quality Improvement (QI)
  - Committee noted some measures are useful and important as internal QI, but not accountability; suggested NQF consider endorsing these types of measures in the future.
- Patient-Reported Outcome Performance Measures (PRO-PMs)
  - Different types of measures are held to different standards of evidence (outcome vs. process).
  - Which measures qualify as PRO-PMs?
    - » Some measures collected by patient reporting are not necessarily PRO-PMs (e.g., a measure asking a patient to report on a process of care).

Comments Received

Comments Received

- NQF received 11 comments from four member organizations
- Two major themes identified:
  - Support for Committee recommendations
  - Gaps for future measure development
- Measure #3136*: GAPPS: Rate of preventable adverse events per 1,000 patient-days among pediatric inpatients*, received specific comments focusing on usability and questions about the definitions.
- The developer responded individually to each question regarding the specifications and definitions
- Regarding usability the developer stated, “As the measure is implemented, the resource burden, while not trivial, should be manageable while providing a great deal of benefit in terms of increased safety events identified.”

Consensus Not Reached Follow Up

- Committee reviewed new information provided by developer, and then re-voted on Reliability and Overall Suitability for Endorsement during the post-comment call.
- 3154: Informed Participation [*formerly Informed Coverage*] (Children's Hospital of Philadelphia) was recommended for endorsement during the post-comment call.

Requests for Reconsideration

- The Committee received two requests for reconsideration:
  - #3189: Rate of Emergency Department Visit Use for Children Managed for Identifiable Asthma: Visits per 100 Child-years
  - #2816: Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma
- The Committee reviewed the new materials prior to the call, and after discussion on the call, agreed that the new information was not sufficient to address its concerns.
- The Committee voted not to reconsider these measures.

Pediatric Performance Measures After Member and Public Commenting

	New	Total
Measures under consideration	12	12
Measures recommended for endorsement	4	4
Measures where consensus is not yet reached	0	0
Measures not recommended for endorsement	7	7
Measures withdrawn from consideration	1	1
Reasons for not recommending	Importance – 4 Scientific Acceptability – 3 Overall – 0 Competing Measure – 0	

Project Timeline and Next Steps

Process Step	Timeline
Appeals Period	July 17- August 15, 2017
Adjudication of Appeals	August 16-September 6, 2017
Final Report	September 26, 2017

Questions?

NATIONAL QUALITY FORUM