# Supplemental Document No. 1

# Information Types and Sources Relevant to the 2013 SNAC Measure Retirement Considerations, by Criterion

*Prepared by:*
Denise Dougherty, PhD
Kamila Mistry, PhD, MPH
Olivia Lindly, MA
Maushami Desoto, PhD, MHSA
Karen LLanos, MBA
Francis Chesley, MD

**Agency for Healthcare Research and Quality**
*Advancing Excellence in Health Care • www.ahrq.gov*

# Introduction

This supplemental document provides detailed information on the types and sources of information relevant to the criteria used by the 2013 AHRQ National Advisory Council Subcommittee on Children's Healthcare Quality Measures for Medicaid and CHIP Programs (SNAC) as it considered whether to retire any of the 20 Child Core Set (CCS) measures required by Title IV of the Children's Health Insurance Program (CHIP) Reauthorization Act (P.L. 111-3). The measures were selected by the Centers for Medicare & Medicaid Services (CMS). The criteria were: importance, scientific acceptability, feasibility, and usability.

Appendix A contains the template SNAC used to report information for each criterion. Appendix B provides details on the search for evidence pertaining to the focus of the measures that scored less than a grade B in 2009 based upon the Oxford Centre for Evidence-Based Medicine (CEBM) Levels of Evidence and Grades of Recommendation.

# Information Sources, by Criterion

## Importance

To guide considerations of measure importance, the template identified a number of data types, including: prevalence; incidence; health care utilization and costs; several indicators of State Medicaid and CHIP performance on the measures to date; and information on variations in the data by race/ethnicity, socioeconomic status, and special health care needs, if available. For context, AHRQ provided the 2013 SNAC with summary data on Medicaid/CHIP child enrollment and the most recent available information on overall health care costs to Medicaid. CMS reported that almost 45 million children were enrolled in Medicaid or CHIP during Federal fiscal year (FFY) 2012. Expenditures for FFY 2012 were not available at the time of this publication, but in 2009, Medicaid Analytic eXtract (MAX) analyses show that total Medicaid fee-for-service expenditures for enrollees ages 0 to 20 years were $54,490,503,103.[1] For comparison, Medicaid fee-for-service expenditures were $171,665,179,984 for enrollees younger than age 65 years and $234,110,168,516 for all enrollees.

Definitions for prevalence/incidence and cost/utilization for each measure are available in AHRQ's *Background Report on 2013 Retirement of Measures from the Child Core Set.*[2] The 2009 MAX data were the preferred source for Medicaid/CHIP prevalence, incidence, health care utilization, costs, and their variations for specific measures. Other data sources were used when MAX data were unavailable or not applicable. Such sources included: analyses using AHRQ Healthcare Cost and Utilization Project (HCUP) hospital discharge data via HCUPNet; published statistical briefs based on the AHRQ Medical Expenditure Panel Survey; data from the National Survey of Child Health; published vital statistics data; and selected published studies.

Data on Medicaid/CHIP State performance on the measures came primarily from State program reports to CMS's CHIP Annual Reporting Template System (CARTS). CMS provided data on the number of States using CMS-prescribed technical specifications for the measure across FFYs 2010 through 2012. Additionally, CMS provided the mean and median performance scores across these States only for each reporting year to enhance comparability. For example, for the

chlamydia screening measure, CMS only reported performance data for the 20 States using the National Committee for Quality Assurance (NCQA) Healthcare Effectiveness Data and Information Set (HEDIS) specifications for all 3 years. Fifteen other States reported on this measure, but did not use the HEDIS specifications. Reporting on the three FFYs allowed the 2013 SNAC to assess both the average absolute rate of performance (i.e., to assess if the measures were "topping off") and progress over time. To provide context to some of the CARTS data, NCQA shared data on performance (average, 10th percentile, 90th percentile) by commercial health plans in FFY 2012 for confidential use only by the 2013 SNAC.

Data for measure importance also includes information about "scientific soundness" for the underlying focus of the measure, which can also be articulated as the evidence on the effectiveness of each care delivery process or structure reflected in each relevant measure, as shown by improved health outcomes. In 2009, SNAC identified the levels of scientific soundness by very rapid reviews of the evidence base for each measure topic (e.g., research on whether recommended rates of the timeliness or frequency for prenatal care affect neonatal outcomes) and assessment of the strength of evidence using the Oxford CEBM criteria.[3] Given time and resource constraints, the 2013 SNAC agreed to use the 2009 SNAC ratings for all relevant initial CCS measures that were graded as an A or B. However, to compensate for a lack of formal updates of the evidence, the 2013 SNAC agreed to provide any knowledge they had of studies that would cause previous ratings to fall below a grade B.

For the initial core measures graded below a B in 2009, AHRQ staff conducted new literature searches to determine whether any new evidence published from 2009 to early 2013 would affect the grades. The measures were: 1) timing of prenatal care, 2) frequency of prenatal care, 3) asthma-related emergency department (ED) visits, 4) followup visits for children prescribed medication for attention deficit hyperactivity disorder, and 5) annual HbA1c testing of children with diabetes. As more fully described in Appendix B, the AHRQ staff conducted rapid searches and syntheses. While the parameters of "rapid reviews" are still emerging, they are responsive to situations such as the ones facing the 2013 SNAC, in which decisionmakers need information for emergent choices.[4] AHRQ staff searched Medline/PubMed for reviews and individual studies, and additionally used U.S. Preventive Services Task Force recommendations, the AHRQ Effective Health Care Program database, and the Cochrane Library Database. After conducting searches, the ensuing citations were inventoried. Based on staff examination of abstract reviews and full-text articles as needed, data on study quality and level of evidence were abstracted into a structured reporting template. Two AHRQ reviewers discussed the findings and reached consensus regarding overall evidence grades using the Oxford CEBM levels, as done in 2009. The reviewers provided the evidence grades and comprehensive summaries of the reviews to the 2013 SNAC; copies of key articles were provided upon request.

## Scientific Acceptability

The scientific acceptability criterion components agreed to by the 2013 SNAC were: measure reliability, measure validity, and Medicaid/CHIP program deviation from technical specifications provided by CMS. There are many approaches to assessing measure reliability and validity. As a practical matter, our data came from two sources. For reliability, NCQA provided an analysis of

reliability for most of the HEDIS measures using its own health plan data, including the number of cases necessary to reach a reliability score of 0.70.

Reliability for the NCQA HEDIS measures was based on the beta-binomial model, which is a ratio of signal to noise. The signal is the proportion of the variability in measure performance that can be explained by real differences in performance. A reliability of zero implies that the total variability in a measure is attributable to measurement error. A reliability of one implies that the total variability is attributable to real differences in performance. The higher the reliability score, the greater the confidence with which one can distinguish the performance of one plan from another. Generally, a minimum reliability score of 0.70 is used to indicate sufficient signal strength. Measures with a strong signal-to-noise ratio have sufficient signal strength to discriminate performance between accountable entities.[5]

For validity, AHRQ staff did a quick search of PubMed for studies related to child and perinatal quality measures for each measure topic and overall.

## Feasibility

For the feasibility criterion, CMS provided data from CARTS on the following: number and percent of States, including the District of Columbia, reporting on the measure by fiscal year; number and percent of States reporting the measure at least once during the three fiscal years; number and percent of States reporting the measure for all three fiscal years; number and percent of States reporting by program (Medicaid, CHIP, or combined) in 2012; and State-reported challenges experienced with data collection and/or reporting in 2012. In addition, CMS provided information on States' reporting challenges reflected in the technical assistance (TA) requests submitted to a CHIP TA mailbox. Information on the data source for the measure (e.g., administrative, medical record, electronic health record, survey, vital records) was taken from CMS's *Technical Specifications and Resource Manual for Federal Fiscal Year 2013 Reporting*. From a State perspective, measures using only administrative data are most feasible. Finally, the CMS "State Health Official Letter of January 2013" was used to get information on the alignment of the use of the measures by other Federal programs, including the measure name as used by the other programs and identification of the Federal program using the measure. Use by other Federal programs implies feasibility in two ways: 1) if other programs are using the measure, it suggests ease of implementation and reporting, and 2) if other programs are already using the measure at the State level, the State may gain some efficiency.

## Usability

Two types of data provided information on the usability criterion; both were intended to provide at least a partial assessment of whether State Medicaid/CHIP programs could take action to close quality gaps. The first type of data identified actions being taken by one or more States to improve quality of care. Some information on State action came from a combination of States reporting to CMS on External Quality Review Organization activities and other State-based quality improvement initiatives; these were subsequently reported in the National Strategy for Quality Improvement in Health Care's 2013 report to Congress.[6] In addition, AHRQ staff used four types of Web sites (Google, Google Scholar, PubMed, and State health department sites) to

search for examples of State efforts to improve quality of care for measure topics that are likely to garner the attention of Medicaid/CHIP (prenatal care, immunizations, chlamydia screening, well-child visits, pharyngitis testing, and ED utilization).

The second source of usability data was information from the medical literature on tests of the extent to which quality improvement (QI) strategies can be effective in closing quality gaps for specific measure topics. AHRQ staff used an iterative approach to identify relevant studies. Although PubMed recently added a medical subject heading (or "MeSH") term for "quality improvement," few studies were found using this term in association with measure topic terms (e.g., prenatal care, asthma-related ED visits) and the terms relevant to children (infant, child, adolescent, or pediatric) or perinatal care. To broaden the searches, we added or substituted search strings using terms such as "performance" or "quality" in association with improvement, intervention, and change. We also looked for systematic reviews in the AHRQ Effective Health Care Program's second "Closing the Quality Gap" series.[7]

For the topics with a sufficient number of QI publications, we entered identified reviews and original studies into Excel spreadsheets, and then summarized the results in separate one-page documents. Each QI review summary (and detailed spreadsheet, if applicable) was shared with at least one identified expert on the 2013 SNAC, and subsequently revised in response to suggestions for additional studies or alternative interpretations of a body of literature, prior to circulation to the full 2013 SNAC. During a 2013 SNAC webinar in which the QI studies were reviewed, one SNAC member suggested that payment reforms likely would lead to improved performance on a number of measures. Prior to providing the final QI summaries, AHRQ staff searched for studies using terms such as "payment reform," "pay-for-performance," "patient-centered medical home," and "accountable care," and noted the results of those searches in the measure reports provided to the SNAC.

## References

1. RTI International, Inc. Medicaid CHIP Contextual Stats. CHIPRA Webex. Rockville, MD: RTI International, Inc.; 2013. Available upon request.
2. RTI International, Inc. Background Report on 2013 Retirement of Measures from the Child Core Set. Summary Report (Prepared by RTI International, Inc. under Contract No. HHSA290201100004C). Rockville, MD: Agency for Healthcare Research and Quality; 2014. AHRQ Publication No. 14-0027. Available at: http://www.ahrq.gov/policymakers/chipra/measure_retirement/measure-retirement-2013.pdf. Accessed May 26, 2014.
3. OCEBM Levels of Evidence Working Group. The Oxford 2009 Levels of Evidence. Oxford, England: Oxford Centre for Evidence-Based Medicine; 2009.
4. Khangura S, Konnyu K, Cushman R, Grimshaw J, Moher D. Evidence summaries: the evolution of a rapid review approach. Syst Rev 2012;1:10.
5. Adams JL. The Reliability of Provider Profiling: A Tutorial. Santa Monica, CA: RAND Corporation; 2009. Available at: http://www.rand.org/pubs/technical_reports/TR653.html. Accessed June 18, 2014.
6. U.S. Department of Health and Human Services. 2013 Annual Progress Report to Congress: National Strategy for Quality Improvement in Health Care. Washington, DC: U.S.

Department of Health and Human Services; 2013. Available at: http://www.ahrq.gov/workingforquality/nqs/nqs2013annlrpt.pdf. Accessed June 16, 2014.

7. McDonald KM, Chang C, Schultz E. Through the Quality Kaleidoscope: Reflections on the Science and Practice of Improving Health Care Quality (Prepared by the Stanford-UCSF Evidence-based Practice Center under Contract No. 290-2007-10062-I). Rockville, MD: Agency for Healthcare Research and Quality; 2013. AHRQ Publication No. 13-EHC041-EF. Available at: http://www.ahrq.gov/research/findings/evidence-based-reports/gapkaleidtp.html. Accessed June 3, 2014.

## Appendix A. Measure Information Reporting Template

| Basic Measure Information | | |
|---|---|---|
| 1 | Measure Number | |
| 2 | National Quality Forum Measure Number | |
| 3 | Measure Steward | |
| 4 | Brief Measure Name and Description | |
| 4.a. | Measure Name | |
| 4.b. | Measure Description (brief narrative) | |
| 5 | Numerator | |
| 6 | Numerator Exclusions | |
| 7 | Denominator | |
| 8 | Denominator Exclusions | |

| Section # | Information Category or Criterion | Data Sources Used | Data pertaining to this measure | This column is available for SNAC to add comments (e.g., rationale for preliminary score) |
|---|---|---|---|---|
| 1. | IMPORTANCE | | | |
| 1.a | Prevalence/Incidence for Focus of Measure | | | |
| 1.a.i | Medicaid/CHIP prevalence/incidence for focus of measure | | | |
| 1.a.i.1 | Variation in Prevalence/Incidence within Medicaid/CHIP<br>*Note to SNAC: Whether a difference is a disparity __and__ b) what type and level of disparities/differences are "important" to consider in measure retirement are judgment calls best made by the SNAC. For informational purposes, the HHS National Healthcare Disparities report currently considers a difference a disparity if it is statistically significant and a 10% difference.* | | | |
| 1.a.i.1.a | Variation by race and ethnicity | | | |
| 1.a.i.1.b | Variation by SES | | | |
| 1.a.i.1.c | Variation by special | | | |

**Appendix A. Measure Information Reporting Template**

| Section # | Information Category or Criterion | Data Sources Used | Data pertaining to this measure | This column is available for SNAC to add comments (e.g., rationale for preliminary score) |
|---|---|---|---|---|
| | health care need | | | |
| 1.a.ii | Optional CONTEXTUAL DATA-- Prevalence/Incidence for Focus of Measure for all U.S. children or on average for U.S. children *(Enter if data are readily available (e.g., from source used to estimate prevalence/incidence among Medicaid/CHIP children)* | | | |
| | | | | |
| 1.a.iii | Optional CONTEXTUAL DATA -- Variation in Prevalence/Incidence U.S. Children Overall/on average *(Enter if readily available (e.g., from source used to estimate prevalence/incidence among Medicaid/CHIP children )* | | | |
| 1.a.iii.a | Variation by race and ethnicity | | | |
| 1.a.ii.b | Variation by SES | | . | |
| 1.a.ii.c | Variation by special health care need | | | |
| 1.b | Health care utilization and costs related to the focus of the measure | | | |
| 1.b.i. | Health care utilization specific to Medicaid/CHIP | | | |
| 1.b.i.1 | Variation in Medicaid/CHIP-specific healthcare utilization | | | |
| 1.b.i.1.a | Variation by race and ethnicity | | | |
| 1.b.i.1.b | Variation by SES | | | |
| 1.b.i.1.c | Variation by special health care need | | | |
| 1.b.ii | Health care costs specific to Medicaid/CHIP | | | |

## Appendix A. Measure Information Reporting Template

| Section # | Information Category or Criterion | Data Sources Used | Data pertaining to this measure | This column is available for SNAC to add comments (e.g., rationale for preliminary score) |
|---|---|---|---|---|
| | | | | |
| 1.b.ii.1 | Variation in Medicaid/CHIP-specific healthcare costs | | | |
| 1.b.ii.1.a | Variation by race/ethnicity | | | |
| 1.b.ii.1.b | Variation by SES | | | |
| 1.b.ii.1.c | Variation by special health care need | | | |
| 1.b.iii | OPTIONAL INFORMATION FOR CONTEXT *(Enter only if readily available)* | | | |
| 1.b.iii.1 | Healthcare utilization overall | | | |
| 1.b.iii.2 | Healthcare costs overall | | | |
| 1.c | Summary of Evidence for focus of the measure<br>NOTE TO SNAC: SNAC agreed with AHRQ's recommendation that only those measures with less than a B grade in 2009 would be subject to AHRQ staff's rapid literature review. SNAC members have the option of providing updated studies related to the other measures. These studies could indicate a possible reduction or elevation of a grade. This information should be provided to AHRQ by July 31, 2103 for inclusion in the draft reporting templates. | | | |
| 1.d | Performance on the measure overall and variations<br>*(NOTE TO SNAC: In interpreting this section, SNAC should attend to different numbers of States reporting in different years for this measure and different approaches to measurement among States as shown below in the Feasibility section)* | | | |
| 1.d.i | Average State Medicaid/CHIP performance on the measure *(mean, median; and # of Reporting States)* | | | |
| 1.d.i.1 | FFY 2010 | | | |
| 1.d.i.2 | FFY 2011 | | | |
| 1.d.i.3 | FFY 2012 | | | |
| 1.d.ii | Cross-state Medicaid/CHIP variation in performance on the measure | | | |

**Appendix A. Measure Information Reporting Template**

| Section # | Information Category or Criterion | Data Sources Used | Data pertaining to this measure | This column is available for SNAC to add comments (e.g., rationale for preliminary score) |
|---|---|---|---|---|
| | | | | |
| 1.d.ii.1 | FFY 2010 | | | |
| 1.d.ii.2 | FFY 2011 | | | |
| 1.d.ii.3 | FFY 2012 | | | |
| 1.d.iii | Overall state Medicaid/CHIP performance on the measure – other data sources IF AVAILABLE (E.G., NCQA HEDIS REPORTS) | | | |
| 1.d.iii.1 | OPTIONAL— CONTEXTUAL INFORMATION ON PERFORMANCE (E.G., COMMERCIAL PERFORMANCE DATA ON MEASURE FROM HEDIS/NCQA | | | |
| 1.d.iv | Variation in state Medicaid/CHIP performance on the measure using other data sources IF AVAILABLE. | | | |
| 1.d.iv.1 | Variation by race and ethnicity | | | |
| 1.d.iv.2 | Variation by SES | | | |
| 1.d.iv.3 | Variation by special health care needs | | | |
| 1.d.v. | OPTIONAL – AHRQ STAFF OBSERVATIONS THAT MAY BE OF INTEREST TO SNAC | | | |
| SNAC preliminary member score for this measure on Importance Criterion (scale of 1-9*) | | | [Enter Score Here] | [Enter comments here] |

**Appendix A. Measure Information Reporting Template**

| Section # | Information Category or Criterion | Data Sources Used | Data pertaining to this measure | This column is available for SNAC to add comments (e.g., rationale for preliminary score) |
|---|---|---|---|---|
| 2. | **SCIENTIFIC ACCEPTABILITY (WEIGHT = 0.20)** | | | |
| 2.a | **Reliability** | | | |
| 2.b | **Validity** | | | |
| 2.c | **Medicaid/CHIP Program Deviation from technical specifications provided by CMS – Number (%) of reporting states with a deviation, by federal fiscal year** | | | |
| 2.c.i | **FFY 2010** | | | |
| 2.c.ii | **FFY 2011** | | | |
| 2.c.iii | **FFY 2012** | | | |
| 2.d | **OPTIONAL – AHRQ STAFF OBSERVATIONS THAT MAY BE OF INTEREST TO SNAC:** | | | |
| **SNAC preliminary member score for this measure on Scientific Acceptability (scale of 1-9)** | | | **[Enter Score Here]** | **[Enter comments here]** |
| 3. | **FEASABILITY** | | | |
| 3.a.i | **Number (%) of States reporting, by fiscal year**<br>*CAVEAT: If a State doesn't report, SNAC should not necessarily infer that there is a technical/feasibility problem with the measures. Some States may not wish to participate at all or have to be selective about resources.* | | | |
| 3.a.i.1 | **FFY 2010** | | | |
| 3.a.i.2 | **FFY 2011** | | | |
| 3.a.i.3 | **FFY 2012** | | | |
| 3.a.ii | **Number (%) of State Medicaid/CHIP programs reporting the measure at least once during the 3 fiscal years** | | | |
| 3.a.iii | **Number (%) of State Medicaid/CHIP programs reporting the measure all 3 years** | | | |

**Appendix A. Measure Information Reporting Template**

| Section # | Information Category or Criterion | Data Sources Used | Data pertaining to this measure | This column is available for SNAC to add comments (e.g., rationale for preliminary score) |
|---|---|---|---|---|
| 3.a.iv | Number (% of reporting states) by program | | | |
| 3.a.iv.1 | Medicaid only | CMS, from CARTS | | |
| 3.a.iv.2 | CHIP only | CMS, from CARTS | | |
| 3.a.iv.3 | Combined | CMS, from CARTS | | |
| 3.b | State Medicaid/CHIP program challenges experienced with data collection and/or reporting (state reported) | | | |
| 3.b.i. | Data not available | CHIPRA TA Mailbox | 2 | |
| 3.b.ii | Population not covered | | 0 | |
| 3.b.iii | Sample size too small | | 3 | |
| 3.b.iv | Other | | 3 | |
| 3.b.v | Not specified | | 0 | |
| 3.c | State reporting challenges reflected in the TA requests submitted to the CHIPRA TA mailbox | | | |
| 3.c.i | Number of TA requests re measure for period xx-xx and number of States making requests | CMS | 8 | |
| 3.c.ii | TA topics for the measure | CMS – Note number of requests for each type of request | Calculation of denominator Calculation of numerator Measure exclusions) Data sources used to calculate the measure Use of alternate methodology | |
| 3.d | Data source for the measure | FFY 2013 CMS Resource Manual and Technical Specifications | Administrative claims | |
| 3.e | Alignment with measures for other federal programs (For each measure, report: name of measure and federal program using the measure) | | | |
| 3.e.i | Name of other measure(s) | CMS | | |
| 3.e.ii | Federal program(s) using the measure | CMS | | |
| 3.f | OPTIONAL – AHRQ STAFF OBSERVATIONS THAT MAY | | | |

**Appendix A. Measure Information Reporting Template**

| Section # | Information Category or Criterion | Data Sources Used | Data pertaining to this measure | | This column is available for SNAC to add comments (e.g., rationale for preliminary score) |
|---|---|---|---|---|---|
| | BE OF INTEREST TO SNAC | | | | |
| **SNAC preliminary member score for this measure for Feasibility (scale of 1-9)** | | | **[Enter Score Here]** | | **[Enter comments here]** |
| **4** | **USABILITY** | | | | |
| **4.a** | **Action(s) Taken by States on Measure Topic** | | | | |
| **4.a.i** | **State Medicaid/CHIP efforts** | | | | |
| **4.a.ii** | **Other STATE program initiatives (not Medicaid/CHIP)** | | | | |
| **4.b** | **Improvability (potential to improve) - summary of studies demonstrating that performance CAN be improved** | | | | |
| **4.c** | **OPTIONAL – AHRQ STAFF OBSERVATIONS THAT MAY BE OF INTEREST TO SNAC** | | | | |
| **SNAC preliminary member score for this measure for Usability (scale of 1-9)** | | | **[Enter Score Here]** | | **[Enter comment here]** |
| **SNAC Member COMMENTS:** Observations, concerns, questions about the measures not covered by domains in the template | | | | | |
| Taking all criteria into consideration, should measure by retired from the Child Core Set? (Circle YES or NO) | | **Yes** | **No** | | |

*In the Modified Delphi approach used, the scale of 1-9 is typically broken into three categories: a score of 1-3 is low; a score of 4-6 is considered medium; and a score of 7-9 is considered high. For more information about the use of Delphi scoring to select quality measures, see the following publications:

    a. Schuster MA, Asch SM, McGlynn EA, Kerr EA, Hardy AM, Gifford DS. Development of a quality of care measurement system for children and adolescents. Methodological considerations and comparisons with a system for adult women. Arch Pediatr Adolescent Med 1997;151(11):1085-92.

## Appendix A. Measure Information Reporting Template

b.  Mangione-Smith R, Schiff J, Dougherty D. Identifying children's health care quality measures for Medicaid and CHIP: an evidence-informed, publicly transparent expert process. Acad Pediatr 2011;11(3 Suppl):S11-21.
c.  Jeffs L, Law M, Straus S, Cardoso R, Lyons R, Bell C. Defining quality outcomes for complex-care patients transitioning across the continuum using a structured panel process. BMJ Qual Saf 2013;22(12):1014-24.
d.  Davies S, Romano P, Schmidt E, Schultz E, Geppert J, McDonald K. Assessment of a novel hybrid Delphi and Nominal Groups technique to evaluate quality indicators. Health Serv Res 2011;46(6pt1):2005-18.

**Source:** Centers for Medicare & Medicaid Services. Core Set of Children's Health Care Quality Measures for Medicaid and CHIP (Child Core Set): Technical Specifications and Resource Manual for Federal Fiscal Year 2013 Reporting. Baltimore, MD: Centers for Medicare & Medicaid Services; 2013.

# Summary of Evidence for Focus of the Measure

1. Use of the structure-process-outcome construct for conceptualizing evidence
2. Use of 2009 evidence grades if graded A or B
3. Rapid review of evidence for measures graded below a B in 2009

Based on the evidence grades determined in 2009 for the 20 pediatric health care quality measures included in the Children's Health Insurance Program Reauthorization Act (CHIPRA) initial core set, the following measures were initially identified as having evidence grades below a B, according to the Oxford Centre for Evidence-Based Medicine (CEBM) Levels of Evidence and Grades of Recommendation. The measure numbering below is the convention used by the 2013 AHRQ National Advisory Council Subcommittee on Children's Healthcare Quality Measures for Medicaid and CHIP Programs (SNAC), however, this numbering differs from the numbering provided in the Initial Core Set Technical Specifications document.

- Measure 1: Timeliness of Prenatal Care (2009 grades: B and D)
- Measure 2: Frequency of Prenatal Care (2009 grades: B and D)
- Measure 7: Weight Assessment and Counseling for Nutrition and Physical Activity for Children/Adolescents: Body Mass Index Assessments for Children/Adolescents (2009 grades: D and I)
- Measure 17: Annual Percentage of Asthma Patients Ages 2 to 20 Years with One or More Asthma-Related Emergency Department Visits (2009 grade: C)
- Measure 18: Followup Care for Children Prescribed Attention Deficit and Hyperactivity Disorder Medication (2009 grade: D)
- Measure 19: Annual Pediatric Hemoglobin A1c (HbA1c) Testing (2009 grade: D)

As described in further detail by Mangione-Smith and colleagues,[1] measures received evidence grades of C and D when the studies on the focus of the measure were primarily designed as case control or case series and/or were based on expert consensus opinion, such as some of the clinical practice guidelines published by the American Academy of Pediatrics (AAP). Given that the U.S. Preventive Services Task Force (USPSTF) changed its recommendation for screening for obesity in children and adolescents to a grade B in January 2010, shortly after the 2009 CHIPRA pediatric health care quality measures effort, it was determined by AHRQ that an updated evidence review was not needed for Measure 7.[2] Additionally, it was determined that the rapid evidence reviews on the importance of Measures 1 and 2 should be combined after a preliminary review of the literature published since 2009 failed to identify any studies that separately examined prenatal care (PNC) timing versus frequency.

The approach used to conduct rapid reviews of the evidence on the importance of the measure topical areas and develop draft syntheses for the measures is provided below. The rapid review and synthesis process was concentrated over a 3-month period in the 2013. However, planning and literature searching prior to this timeframe also informed this process.

## 1. Search Strategies

For each of the measure topical areas, identical search strategies were used to obtain evidence on the importance of each care delivery process reflected by the given measurement area. More

**Appendix B. CHIPRA Quality Measures Retirement Process for Summarizing Evidence for the Focus of the Measure**

specifically, this evidence was defined as published literature on the effectiveness of the given care delivery processes in eliciting advantageous or improved health outcomes in women and/or children. To begin, medical subject headings ( or "MeSH") terms and subheadings most closely aligned with the measure topical area(s) or care process (e.g., HbA1c testing for children with diabetes), population(s) of interest, and potential health outcomes were identified by two AHRQ staff members using the U.S. National Library of Medicine MeSH database (Table 1). For certain measure topical areas, such as asthma, it was necessary to select MeSH subheadings in PubMed to limit the searches and yield appropriate results.

**Table 1. MeSH Terms by CHIPRA Core Measure Topical Areas of Interest**

| CHIPRA Measure Topical Area | Maternal or Child Relevant MeSH Terms | Health Care/Condition Relevant Topical Area MeSH Terms | Outcome Relevant MeSH Terms (subheading) |
|---|---|---|---|
| Timeliness of PNC | ▪ Females | ▪ Prenatal care<br>▪ Antenatal care | ▪ Live birth<br>▪ Reproductive history |
| Frequency of PNC | | | |
| Asthma-Related ED Visits | ▪ Children<br>▪ Adolescents | ▪ Asthma<br>▪ Emergency medical care | ▪ Prevention and control (asthma)<br>▪ Therapy (asthma)<br>▪ Therapeutic use (asthma)<br>▪ Primary health care |
| Followup for Children Prescribed ADHD Medication | | ▪ Attention deficit hyperactivity disorder (subheading: drug therapy) | ▪ Prevention and control (ADHD)<br>▪ Rehabilitation (ADHD)<br>▪ Therapy (ADHD) |
| HbA1c Testing for Children with Diabetes | | ▪ Diabetes mellitus<br>▪ Hemoglobin A1c protein, human | ▪ Prevention and control (DM)<br>▪ Therapy (DM)<br>▪ Rehabilitation (DM) |

**Abbreviations:** ADHD = attention deficit hyperactivity disorder; CHIPRA = Children's Health Insurance Program Reauthorization Act; DM = diabetes mellitus; ED = emergency department; HbA1c = glycated hemoglobin; MeSH = medical subject heading; PNC = prenatal care.

After the identification of MeSH terms and subheadings, the process by which evidence regarding each of the measure topical areas was determined. To ensure that evidence reviews and other secondary information sources not available or not easily accessible through the PubMed database were located, AHRQ staff members first electronically searched the USPSTF Recommendations for Children and Adolescents (available at http://www.uspreventiveservicestaskforce.org/recommendations.htm) by the clinical categories relevant to each of the measure topical areas. The AHRQ Effective Health Care Program database (available at http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/) was searched next for relevant reports of comparative effectiveness research or systematic reviews on each of the measure topical areas, according to relevant health conditions. The Cochrane Library Database was searched last for relevant systematic reviews using the MeSH terms and subheadings in the keyword search field.

To find primary literature or any additional secondary literature on the importance of each measure topical area, the PubMed database was searched using the MeSH terms and subheadings identified along with the following filters: age (birth to age 18 years) (this filter was not used

when searching for evidence on the PNC measures, sex [female] filter was used instead); publication dates (custom: January 1, 2008, to August [day when search conducted], 2013); species (human); and languages (English). In addition, MeSH terms for the various study types of greatest interest were specified. These included: systematic reviews; randomized, control trials; cohort studies; case control studies; and case reports. For certain measure topical areas, additional reviewed articles were referred by members of SNAC or the CHIPRA Centers of Excellence. Further, seminal studies published prior to 2008 but cited in recent clinical practice guidelines were located and included in this review. Detailed documents describing the results for the searches conducted in each measure topical area are available upon request.

## 2.  Development of Draft Evidence Syntheses

For each search conducted, the resulting citations were inventoried. After conducting abstract reviews of those articles identified as potentially relevant to each of the measure topical areas, full-text articles were reviewed. If a reviewer was uncertain if the full text of a certain article should be reviewed, another reviewer was consulted, and then both reviewers decided the article's relevancy. After full-text review, data were abstracted into a structured reporting template containing the following information categories: article citation; study location (setting type, geographical location); Medicaid- or CHIP-specific or addressed; study funder/lead; study purpose; patient population/sample; intervention (including duration); study design and comparator; and study quality/level of evidence.

Study quality/level of evidence was ascertained in 2009 based on the Oxford CEBM Levels of Evidence.[3] Two AHRQ reviewers reached consensus about the study quality/level of evidence for each full-text article from which data were abstracted. Using the Oxford CEBM Grades of Recommendation, the two reviewers then determined what grade was most appropriate for the resultant body of evidence. Brief descriptions explaining the rationale for each grade were also provided in the draft evidence syntheses. Table 2 summarizes the key elements that comprised the draft evidence syntheses on the measure topical areas. Citations for relevant clinical practice guidelines published prior to 2008, as well as any other more recently published guidelines, were also provided in the full draft evidence syntheses reviewed by the SNAC members.

**Appendix B. CHIPRA Quality Measures Retirement Process for Summarizing Evidence for the Focus of the Measure**

**Table 2. Overview of Evidence Search Results and Syntheses by Measure Topical Areas**

| CHIPRA Measure Topical Area | Number of Articles Included in the Body of Evidence | Evidence Levels of Relevant Studies | 2013 Evidence Grade and Explanation |
|---|---|---|---|
| **Timeliness and Frequency of Prenatal Care** | 5 | 1a (study type: systematic reviews)<br>2b (study type: cohort studies)<br>2b (study type: cohort studies)<br>2b (study type: cohort studies)<br>3b (study type: individual case control studies) | **Grade B:** Studies found since 2008 provide modest evidence that addresses the focus of the evidence review. Most studies did not specifically address timeliness and/or frequency, but rather the content of prenatal care. Additionally, several studies characterize the "adequacy of prenatal care," which incorporates aspects of timeliness and frequency, but makes it difficult to comment on the individual effect of each on health outcomes. Further, only one of the most relevant studies included in the table below was conducted in the United States, which may limit the generalizability of findings from this body of evidence to the United States. |
| **Asthma-Related ED Visits** | 7 | 2b (study type: cohort studies)<br>2b (study type: cohort studies)<br>2c (study type: outcomes research)<br>4 (study type: case series)<br>4 (study type: case series)<br>4 (study type: case series)<br>4 (study type: case series) | **Grade C:** Studies found since 2008 provide some evidence that addresses the focus of the evidence review. With regard to primary care (models such as medical home), studies showed fewer ED visits in children with asthma. In regard to studies examining the association between followup after ED visits and ED readmission in children with asthma, no significant effects were noted in the evidence found. |
| **Followup Visits for Children Prescribed ADHD Medication** | 5 | 1b (study type: randomized control trials)<br>2b (study type: cohort studies)<br>2b (study type: cohort studies)<br>3a (study type: case control<br>4 (study type: case series) | **Grade C:** Studies found since 2008 do not provide sufficient evidence that directly addresses the review's focus; however, prior to 2008, the *Multimodal Treatment Study of ADHD* provided evidence that high-quality medication treatment (including monthly followup visits) was associated with improved outcomes. This study informed the AAP 2000/2001 guidelines (updated in 2011) recommending followup for chronic care management at least two times per year. |
| **HbA1c Testing for Children With T1DM and T2DM** | 4 | 4 (study type: case series)<br>4 (study type: case series)<br>2b (study type: cohort studies)<br>2b (study type: cohort studies) | **Grade D for T1DM:** Little evidence on the timing of HbA1c testing was found.<br><br>**Grade D for T2DM:** Grade is consistent with the 2013 AAP evidence rating for clinical HbA1c monitoring. Additionally, little evidence regarding timing was found.<br><br>No studies found since 2008 provide evidence that addresses the focus of the evidence review (i.e., annual clinical HbA1c testing for T1DM and T2DM). For T2DM, no studies have examined the timing of clinical HbA1c monitoring for children or adolescents. For T1DM, evidence and expert consensus prior to 2008 suggest that testing is associated with better health outcomes, but it is not clear what interval versus annual testing is effective. Evidence cited in the 2013 AAP clinical practice guideline, *Management of Newly Diagnosed Type 2 Diabetes Mellitus in Children and Adolescents,* indicates a low evidence quality (grade D) for T2DM; however, AAP recommends clinical monitoring of HbA1c every 3 months. This guideline does not directly address the frequency of HbA1c testing for T1DM. |

**Abbreviations:** AAP = American Academy of Pediatrics; ADHD: attention deficit hyperactivity disorder; ED = emergency department; HbA1c = glycated hemoglobin; T1DM = Type 1 diabetes mellitus; T2DM = Type 2 diabetes mellitus.

### 3. Review by 2013 SNAC and Revisions

Detailed results are provided in the primary manuscript to this supplement: Dougherty D, Mistry K, Lindly O, Desoto M, LLanos K, Chesley F. Systematic evidence-based quality measurement life-cycle approach to measure retirement in CHIPRA. Acad Pediatr 2014;14:S97–S104.

## References

1. Mangione-Smith R, Schiff J, Dougherty D. Identifying children's health care quality measures for Medicaid and CHIP: an evidence-informed, publicly transparent expert process. Acad Pediatr 2011;11(3S):S11-S21.
2. U.S. Preventive Services Task Force. Screening for Obesity in Children and Adolescents, Topic Page. January 2010. U.S. Preventive Services Task Force. Available at: http://www.uspreventiveservicestaskforce.org/uspstf/uspschobes.htm. Accessed August 2013.
3. OCEBM Levels of Evidence Working Group. The Oxford 2009 Levels of Evidence. Oxford, England: Oxford Centre for Evidence-Based Medicine; 2009.