

# Identifying Children's Health Care Quality Measures for Medicaid and CHIP: An Evidence-Informed, Publicly Transparent Expert Process

Rita Mangione-Smith, MD, MPH; Jeffrey Schiff, MD, MBA; Denise Dougherty, PhD

From the Department of Pediatrics, University of Washington, and the Center for Child Health, Behavior, and Development, Seattle Children's Hospital Research Institute (Dr Mangione-Smith); Health Services Medical Management, Minnesota Department of Human Services, St Paul, Minn (Dr Schiff); and the Office of Extramural Research, Education and Priority Populations, Agency for Healthcare Research and Quality, Rockville, Md (Dr Dougherty)

The views expressed in this article are those of the authors and do not necessarily reflect the views of the US Department of Health and Human Services, the Agency for Healthcare Research and Quality, or the Centers for Medicare & Medicaid Services.

The authors have no conflicts of interest to disclose.

Publication of this article was supported by the US Department of Health and Human Services or the Agency for Healthcare Research and Quality.

Address correspondence to Rita Mangione-Smith, MD, MPH, Department of Pediatrics, University of Washington, Center for Child Health, Behavior and Development, Seattle Children's Hospital Research Institute, 1100 Olive Way, Suite 500, Seattle, Washington 98101 (e-mail: [Rita.Mangione-Smith@seattlechildrens.org](mailto:Rita.Mangione-Smith@seattlechildrens.org)).

Received for publication February 27, 2010; accepted November 6, 2010.

## ABSTRACT

**OBJECTIVE:** To describe the process used to identify the recommended core set of quality measures as mandated by the Children's Health Insurance Program Reauthorization Act of 2009 (CHIPRA) and provide an overview of the measures selected.

**METHODS:** In May 2009, the multidisciplinary Agency for Healthcare Research and Quality (AHRQ) National Advisory Council for Healthcare Research and Quality Subcommittee on Children's Healthcare Quality Measures for Medicaid and CHIP Programs (SNAC) was formed. The SNAC established criteria to evaluate quality measures on the basis of their validity, feasibility, and importance. Subsequently, AHRQ implemented a measure nomination process. Nominators supplied key information related to measure validity, feasibility, and importance. Oxford Centre for Evidence-Based Medicine (CEBM) criteria were used to assess evidence supporting the validity/scientific soundness of nominated measures. SNAC members applied an adaptation of the RAND-UCLA modified Delphi process to all nominated measures. Measures passing the Delphi process were further assessed on the basis of criteria pertaining to legislative priorities.

**RESULTS:** Seventy of 119 nominated measures met criteria for validity, feasibility, and importance according to Delphi scoring. After further prioritization, 25 measures were recommended for the initial core set. Twelve of the recommended measures focus on preventive care and health promotion including prenatal/perinatal care (4), well-child care (1),

immunizations (2), screening for: developmental delays (1), obesity (1), and sexually transmitted infections (1), and receipt of preventive dental services (2). Five acute care measures were recommended which focus on management of upper respiratory illnesses (2), receipt of acute care dental services (1), emergency department utilization (1), and inpatient rates of central line associated bloodstream infections (1). Five of the recommended measures focus on chronic care, specifically asthma (1), attention-deficit/hyperactivity disorder (1), diabetes (1), and care for children with mental health conditions (2). Two of the measures focus on family experiences with care, and one of the measures assesses utilization of outpatient primary care services. Thirteen (52%) of the measures were derived from the National Committee for Quality Assurance (NCQA) Healthcare Effectiveness Data and Information Set (HEDIS). Eighteen of the measures were supported by relatively high levels of evidence (Oxford CEBM grade A or B).

**CONCLUSIONS:** An open national public process combined with an evidence-informed evaluation methodology resulted in identification of a balanced, grounded, and parsimonious core set of measures that should become feasible to implement on a widespread scale over time.

**KEYWORDS:** children's health care quality measurement; CHIP; CHIPRA; Medicaid

**ACADEMIC PEDIATRICS** 2011;11:S11-S21

WITH THE PASSAGE of the Children's Health Insurance Program (CHIP) legislation in 1997 (Public Law 105-33), Congress made their first commitment to developing quality-of-care performance measures to assess the care provided to children under this program.<sup>1</sup> However, as illustrated by deLone and Hess in this supplement,<sup>2</sup> interpretation of the requirements in that legislation has resulted in wide variation in state approaches to quality

measurement and a limited number of measures that can be compared across states. Further, a recently released annual report from the Secretary of Health and Human Services (HHS) on the Quality of Health Care for Children in Medicaid and CHIP noted that the Center for Medicare and Medicaid Services (CMS) has no uniform system for assessing quality of care for children across states.<sup>3</sup> Concerns at the federal level related to this lack of

a uniform approach to quality measurement led to the inclusion of several specific provisions in the Children's Health Insurance Program Reauthorization Act (CHIPRA) legislation of 2009 aimed at addressing these deficits. These provisions included that the secretary identify a core set of children's healthcare quality measures for voluntary use by states starting in December of 2011.<sup>4</sup>

## THE LEGISLATION

Title IV of CHIPRA (Public Law 111-3) required the Secretary of HHS to identify and post for public comment by January 1, 2010, an initial recommended core set of children's health care quality measures for voluntary use by CHIP, health insurance issuers and managed care entities that enter into contracts with such programs, and providers of items and services under such programs. The legislation called for identification of "existing quality of care measures for children that are in use under public and privately sponsored health care coverage arrangements or are part of reporting systems that measure both the presence and duration of health insurance coverage over time."<sup>4</sup>

The legislation specifically called for identification of measures on the following topics, although others could be included: duration of enrollment and coverage; preventive and health promotion services; treatment and management for acute and chronic conditions in children; family experiences of care, most integrated health care settings; and availability of services. CHIPRA also called for evidence-based measures and measures that could identify disparities in health care quality by race and ethnicity, socioeconomic status, and special health care need.

## AHRQ/CMS PARTNERSHIP

In response to this legislative directive, the Agency for Healthcare Research and Quality (AHRQ) and the Centers for Medicare & Medicaid Services (CMS) signed a Memorandum of Understanding giving AHRQ leadership responsibilities for identifying the initial core set, working in close partnership with CMS. CMS has the authority for implementation of all CHIPRA provisions.

We present a brief summary of the processes used to identify an initial recommended core set of children's health care quality measures and the key steps that will need to be taken to successfully implement them on a wide-spread scale.

## METHODS

The initial core set of children's health care quality measures for voluntary use by Medicaid and CHIP was developed by means of a transparent and evidence-informed process that included broad input from multiple stakeholders (Figure 1). Key components included multiple opportunities for public comment; an AHRQ National Advisory Council on Healthcare Research and Quality (NAC) Subcommittee (SNAC) that contributed expertise on validity, feasibility, and importance of measures in use; and supportive background work by

AHRQ, CMS, members of the CHIPRA Federal Quality Workgroup (FQW),<sup>5</sup> as well as authors of other papers included in this supplement.

## ESTABLISHMENT OF THE SNAC

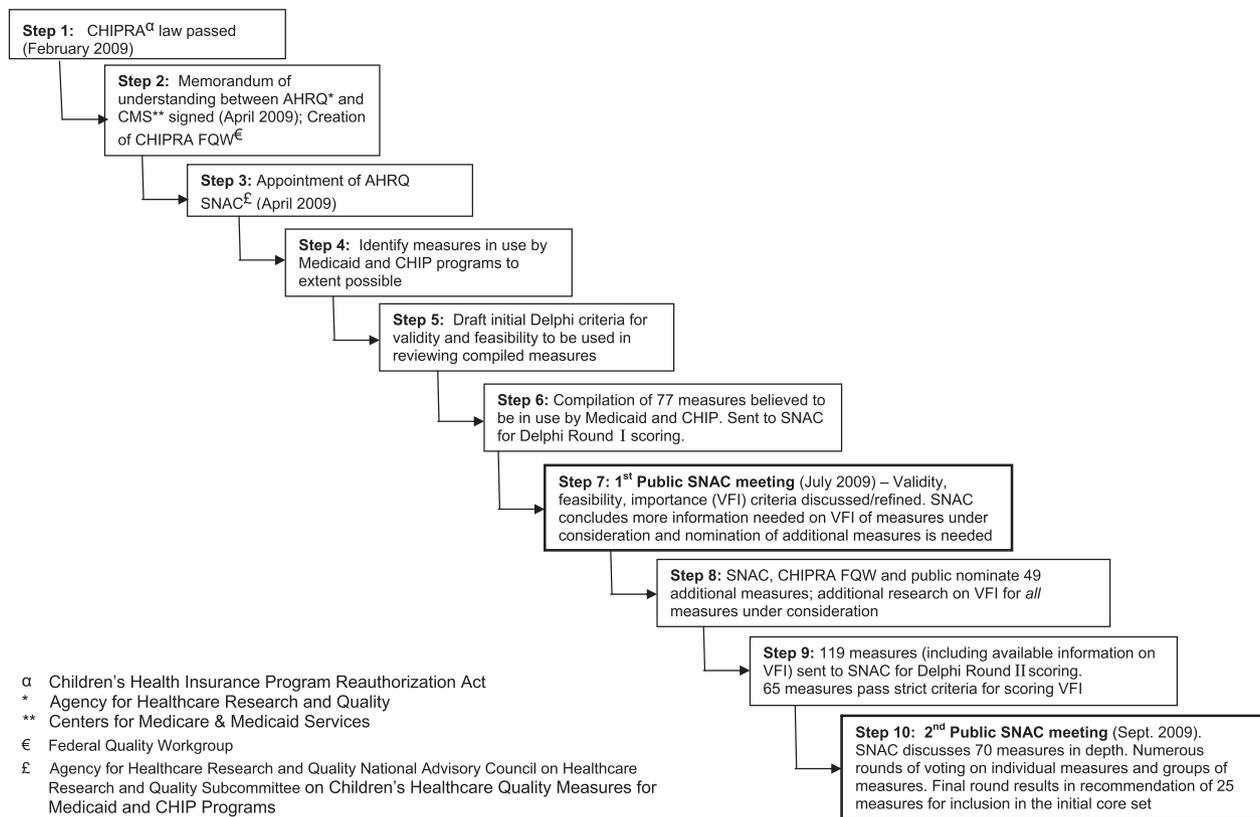
As one of the first steps in the process of identifying the recommended core set of measures, the AHRQ director approved a charter creating the SNAC. The AHRQ NAC had agreed to provide advice to AHRQ and CMS to facilitate their work to recommend an initial core set of measures of children's health care quality for Medicaid and CHIP programs. To provide the requisite expertise and input from the range of stakeholders identified in the CHIPRA legislation, the NAC established the SNAC.

The SNAC was charged with providing guidance on measure evaluation criteria to be used in identifying an initial core measurement set, providing guidance on a strategy for gathering additional measures and measure information from state programs and others, and reviewing and applying criteria to a compilation of measures currently in use by Medicaid and CHIP programs to begin selection of the initial core measurement set. SNAC recommendations were to be provided to CMS and the NAC, which in turn would advise the Director of AHRQ. The Directors of AHRQ and CMS would then review and decide on the final recommended core set to be presented to the Secretary of HHS for consideration.

Nominations for SNAC members to represent the range of stakeholders were sought from CMS and the CHIPRA FQW.<sup>6</sup> Emphasis was placed on including Medicaid and CHIP officials because of their unique role as potential implementers of the initial core set. Four state Medicaid program officials (from Alabama, Minnesota, Missouri, and the District of Columbia) and one state CHIP official (from Alabama) were able to participate. Others represented Medicaid, CHIP, and other state programs more generally (ie, representatives of the National Academy on State Health Policy, National Association of State Medicaid Directors, and the Association of Maternal and Child Health Programs).

Representatives of health care provider groups came from the American Academy of Family Physicians, American Academy of Pediatrics, American Board of Pediatrics, the National Association of Children's Hospitals and Related Institutions, the National Association of Pediatric Nurse Practitioners, and a Medicaid health plan representative. The interests of families and children were represented by the March of Dimes. Individual SNAC members provided expertise in children's health care quality measurement, children's health care disparities, tribal health care, pediatric dental care, substance abuse and mental health care, adolescent health, and children's health care delivery systems in general. Two members of the NAC also participated in the SNAC.

The SNAC held 2 public meetings and accomplished a substantial amount of work outside the meetings in order to provide recommendations to the NAC, AHRQ, CMS, and the HHS Secretary to meet the CHIPRA legislative deadline of January 1, 2010.<sup>7</sup>



**Figure 1.** Process of identifying an initial, recommended core set of children's health care quality measures for use by Medicaid and CHIP programs.

## PUBLIC INPUT

Multiple ongoing opportunities for public input were provided as part of this process. Early on, AHRQ established a Web site to provide information on its role in CHIPRA implementation, in close collaboration with CMS, and an e-mail address through which the public could comment on the process. In addition, both SNAC meetings were open to the public and provided opportunities for formal public comments. A conference call for members of the Medicaid Medical Directors Learning Network was also conducted to seek input on the measure identification and nomination process. Several members of the network responded by nominating children's health care quality measures in use by their states for consideration for the initial core measure set.

Those making public comments through these mechanisms included individual health care practitioners, additional Medicaid and CHIP program officials, representatives of professional societies and industry groups, child and family advocates, and members of the CHIPRA FQW.

## FIRST SNAC MEETING

This section describes preparation for the first SNAC meeting, the focus of SNAC discussions, presentations to the SNAC, refinements to methodology made during the meeting, and the identification of a preliminary group of measures to further consider for inclusion in the final core set, as well as needs for additional information and work.

## PREPARATION

Seventy-seven measures in use by Medicaid and CHIP programs were identified by AHRQ staff, with the assistance of CMS, and a process for initial evaluation of those measures was agreed upon by the SNAC co-chairs, AHRQ, CMS, and the FQW.

The SNAC co-chairs, working through AHRQ, provided subcommittee members with standard definitions and criteria recommended for use in evaluating the validity and feasibility (including reliability) of quality measures (Table 1). SNAC members were asked to apply these evaluation criteria to the 77 measures using the RAND-UCLA modified Delphi method.<sup>8</sup> Previous work has shown this method of evaluating quality measures to be reliable and to have content, construct, and predictive validity in other applications.<sup>9–11</sup>

The modified Delphi process involved individual SNAC members scoring the initial identified set of Medicaid and CHIP quality measures separately for validity and feasibility on a 1- to 9-point scale (1 = measure not valid or feasible; 9 = measure definitely valid or feasible). Objective information (eg, on underlying scientific soundness of the measures) related to both measure validity and feasibility was provided to the extent it was available. However some measures were scored in this first round without adequate identification of numerators, denominators, or measure specifications, all of which are essential for evaluating feasibility of measure implementation.

**Table 1.** Evaluation Criteria Used in Delphi Round I to Assess Validity and Feasibility of Candidate Measures for the Recommended Core Set**A Quality Measure Should be Considered Valid If:**

1. There is adequate scientific evidence or, where evidence is insufficient, expert professional consensus to support the stated relationship between:
  - **Structure and process** (eg, that there is a demonstrated likelihood that a clinical decision support system [a structural or capacity measure] in a hospital or ambulatory office leads to increased rates of appropriate flu vaccination in the hospital or practice).
  - **Structure and outcome** (eg, higher continuity of care in the outpatient setting [influenced by how appointments are organized] is associated with fewer ambulatory care sensitive hospitalizations [eg, hospitalizations for dehydration]).
  - **Process or structure and outcome** (eg, that there is a demonstrated likelihood that prescribing inhaled corticosteroids [a clinical process] to specified patients with asthma will improve the patients' outcomes and vice versa (eg, if we measure health outcomes there is a sufficient demonstrated likelihood that the outcomes can be attributed to either clinical processes of care or health care delivery structures or a combination of both).
2. The health care system can be said to be responsible for performance and/or the related health outcome. The majority of factors that determine adherence to a measure are under the control of the clinician, clinic, hospital, health plan, or the Medicaid or CHIP program subject to measurement.

**A Quality Measure Should be Considered Feasible If:**

1. The information necessary to determine adherence to the measure is likely to be found in available data sources (eg, administrative billing data, medical records, or routinely collected survey data).
2. Estimates of adherence to the measure based on available data sources are likely to be reliable and unbiased. Reliability is the degree to which the measure is free from random error.

CHIP = Children's Health Insurance Program.

Instructions to the SNAC for Delphi Round I noted that scores for validity could be guided by professional consensus when published evidence to support the measure's validity was insufficient.

The RAND-UCLA modified Delphi method outlines cut points for passing scores on validity and feasibility. For validity, the median passing score used is more stringent, ie, 7–9 on the 9-point scale, than the passing score for feasibility, which requires a median score of 4–9. The rationale for this difference is that for validity, the evidence or expert consensus that exists to support the assessment of the measure is relatively objective. Feasibility is a more subjective and variable assessment than validity. Some Medicaid or CHIP programs may find a measure quite feasible to implement (due to their infrastructure, amount of available funding, etc), while others will not. Feasibility of implementing a quality measure can also be field tested. If it is determined that a measure is less feasible to implement than initially assumed, needed technical assistance could be provided by CMS, or the measure might be dropped from the core set.

Nineteen (70%) of the 27 SNAC members participated in the first round of Delphi scoring. Median scores and a display of the distribution of scores across voting members were calculated and prepared for SNAC review by AHRQ staff before the first meeting. The median scores summarized the individual scores of SNAC members on these 2 domains (ie, validity and feasibility). The median scores were used to determine whether candidate measures would be discussed further at the meeting. Twenty-five measures passed with sufficient scores to be noncontroversial. Measures with a median validity score of 6 or 7, a median feasibility score  $\geq 4$ , and a relatively wide distribution of scores across members, suggesting little consensus among the group, were discussed by the SNAC. These measures were selected for discussion because they were deemed controversial and in need of further consideration by the group. Forty-five of the originally identified 77 measures in use by Medicaid or CHIP programs met these scoring criteria.

**FIRST SNAC MEETING**

Presentations by the AHRQ Director, CMS's Director of the Center for Medicaid, CHIP, and Survey & Certification (CMCS), and Director of the Division of Evaluation, Quality, and Health Outcomes in CMCS set the stage for the meeting.<sup>12,13</sup> The AHRQ director provided the charge to the SNAC, and the CMCS director expressed a strong desire for the SNAC to recommend a grounded and parsimonious core set of measures that could be implemented voluntarily by state programs, health plans, and provider groups.

The SNAC spent a large part of this first meeting reviewing the criteria for validity and feasibility used in Delphi Round I, identifying criteria for importance, and discussing the measures that were deemed controversial after Delphi Round I.

**REFINEMENTS TO THE MEASURE EVALUATION METHODOLOGY**

The refinement process involved reviewing, discussing, and reaching consensus on criteria the SNAC would use to evaluate the validity and feasibility (including reliability) of candidate measures in future rounds of the Delphi process.<sup>14</sup>

In addition, importance was added as a third domain, along with validity and feasibility, to consider when evaluating potential measures. The SNAC worked to establish consensus on the criteria to be used to rank the importance of measures under consideration. To be considered important, the measure had to meet at least some of the following criteria. The criteria are listed in order of decreasing weight as determined through a voting process by SNAC members:

- The measure should be *actionable*. State Medicaid and CHIP programs, managed care plans, and relevant health care organizations should have the ability to improve their performance on the measure with implementation of quality improvement efforts.
- The *cost* to the nation for the area of care addressed by the measure should be substantial.

- Health care systems should clearly be *accountable* for the quality problem assessed by the measure.
- The *extent of the quality problem* addressed by the measure should be substantial, ie, a significant proportion of the US child population should be affected by poor performance on the measure.
- There should be documented *variation* in performance on the measure.
- The measure should be *representative* of a class of quality problems, ie, it should be a “sentinel measure” of quality of care provided for preventive care, mental health care, dental care, etc.
- The measure should assess an aspect of health care where there are known *disparities*.
- The measure should contribute to a final core set that represents a *balanced portfolio* of measures that is consistent with the intent of the legislation.
- Improving performance on measures included in the core set should have the potential to *transform care* for our nation's children.

Similar to feasibility, the threshold for a passing score on importance was also set at  $\geq 4$  on the 9-point Delphi scale. The SNAC chose this cut point because the necessary information to objectively assess these criteria may not be available in all cases, and some of the criteria were deemed somewhat subjective in nature (eg, whether improvement on the measure would “transform care”).

The SNAC members were asked to score each of the 25 measures that had unequivocally passed the first round of Delphi scoring for validity and feasibility on the new criterion of importance. AHRQ staff then summarized these scores using the median value.

#### OTHER STEPS AND DECISIONS

The SNAC's discussion of the 45 controversial measures resulted in the recommendation that further information related to measure validity, feasibility, and importance would be needed before further consideration of the measures. The SNAC also determined that a call for nominations of additional pediatric quality measures in use (either within or outside of Medicaid and CHIP programs) should be used to identify a larger set of measures to consider for the final core set. AHRQ staff were asked to identify relevant information related to validity, feasibility, and importance on the controversial measures. SNAC members agreed to nominate additional measures in use.

SNAC members expressed a strong desire to recommend a grounded and parsimonious core set of measures that could be implemented voluntarily by state programs, health plans, and provider groups, and they agreed on a target number of no more than 25 measures. The SNAC acknowledged that such a core set would be incomplete, but efforts would be made to balance the set to accomplish the legislative goals and the goals articulated in the SNAC discussion of measure importance. The SNAC agreed to bring forth to the NAC's attention measures not accepted into the core set. Furthermore, SNAC members agreed to

identify important aspects of child health for which valid measures may not currently exist.

#### ADDITIONAL INPUT AND DISCUSSION

Representatives from the National Quality Forum, the National Committee on Quality Assurance, and the Center for Health Care Strategies also spoke at the meeting on the challenges of implementing health care quality measures for children. In addition, several experts who had been asked to write federally supported papers on specific aspects of measurement in the legislation presented their early thoughts about their work. These experts addressed the charges to them of conceptualizing and assessing the validity, feasibility, and importance of measures of mental and behavioral health care, family experiences of care, duration of enrollment and coverage, availability of services, and the “most integrated health care setting.” AHRQ and CMS also asked that papers be prepared analyzing data sets of the National Academy for State Health Policy, Health Management Associates, and the Child and Adolescent Health Measurement Initiative database from the 2007 National Survey on Children's Health.<sup>15</sup> Some of the final papers and their key findings are included in this supplement.

#### PREPARATION AND THE SECOND SNAC MEETING

In addition to being open to public participation on site, the second meeting was Webcast allowing for greater participation and public comment.<sup>16</sup>

#### ADDITIONAL MEASURE NOMINATIONS

AHRQ staff, in collaboration with the SNAC co-chairs, developed a measure nomination template.<sup>17</sup> This template was created in order to collect a standardized set of information on all measures nominated for potential inclusion in the core set. The nomination template was made available for approximately 1 month before the second meeting. In addition to measure nominations by SNAC members, public nominators included members of the Medicaid Medical Directors Learning Network, the American Medical Association Physician Consortium for Performance Improvement, the National Partnership for Women and Families, and the Child and Adolescent Health Measurement Initiative on behalf of The Commonwealth Fund. Additional nominations were obtained through e-mail to the AHRQ public comment e-mail address. CHIPRA FQW nominations also came from CMS and the Health Resources and Services Administration.

In addition to all newly nominated measures, each measure that either passed Delphi Round I (25 measures) or was considered controversial by the SNAC during their first meeting in July (45 measures) was entered into the measure template, with required information, by AHRQ staff. Authors of the CHIPRA-commissioned papers (presented in this supplement) also recommended 4 measures for consideration and additional sources of data for quality measurement on the basis of their works in progress.

At a minimum, nominators were asked to identify the measure numerator and denominator, measure

specifications, and current use of the measure. Each template was then supplemented with additional information where necessary by AHRQ staff and the SNAC co-chairs. Through this work, a standardized set of information was made available for almost all measures for consideration by the SNAC members during their second round of Delphi scoring. A 1-page summary sheet was developed for each measure under consideration that included abstracted information from the measure nomination template and a summary of the evidence base for the measure's underlying scientific soundness using the Oxford Centre for Evidence-Based Medicine (CEBM) criteria (Table 2).<sup>18</sup>

In total, the SNAC had 119 measures to consider during a second modified Delphi process.

### *DELPHI ROUND II SCORING BY THE SNAC*

Members were sent the 1-page summary sheets for the 119 measures under consideration and a scoring instruction guide.<sup>14</sup> By means of a second modified Delphi process, 19 (70%) of 27 SNAC members participated in scoring. The scoring process resulted in selection of 65 of the 119 measures as meeting criteria for validity, feasibility, and importance when we strictly adhered to the preset cutoff points. As in Delphi Round I, SNAC members were told they could use professional consensus on the underlying scientific soundness of the measures in cases of insufficient published evidence.

### *SNAC MEETING DELIBERATIONS*

As a result of time constraints and the need to identify for AHRQ and CMS consideration a reasonable core set of measures near the SNAC's target number of 25, the initial plan was to discuss and consider only the 65 measures that passed the second modified Delphi scoring process as candidates for the core set. However, initial discussions at the meeting resulted in adding back 5 measures that did not strictly pass the second Delphi round (ie, those with high median feasibility and importance scores [ $>7$ ] and median validity scores of 6 or 6.5 rather than the cutoff of 7) to the list of measures to be discussed and voted on during the meeting. Thus, 70 of the 119 measures scored in Delphi Round II were discussed and considered for the core set at the meeting.<sup>19,20</sup>

### *ELECTRONIC VOTING PROCESS*

Throughout the meeting, a method of confidential electronic voting was used extensively by SNAC members. This method was chosen because in small groups some members may dominate a discussion, leading to group decisions that do not reflect the true sense of the group membership.<sup>21</sup> Through private electronic voting, the SNAC process was most likely to obtain the candid individual preferences of members, accumulating to a consensus of the SNAC.

### *BALANCING MEASURES ACROSS MULTIPLE DOMAINS*

The SNAC reviewed and prioritized measures on the basis of several characteristics pertaining to legislative

and feasibility criteria, including: data source (administrative, medical record, electronic health records, other health information technology, or survey); site of care (primary care, specialty care, inpatient, emergency, mental health, substance abuse, dental); measure type (outcome, process, structural); care continuum (screening, prevention, diagnosis, treatment, care coordination); accountable entity (state, program, health plan, provider); child ages to which the measure applied; and availability of data to report disparities.

### *ELIMINATION OF OVERLAPPING MEASURES, MERGING OF SOME MEASURES, AND VOTING*

SNAC members engaged in detailed discussions of measures felt to have substantial overlap. For example, multiple measures pertaining to healthy birth (including the prevention of premature birth) passed the criteria for validity, feasibility, and importance, as did multiple dental measures. After discussions were completed, a series of votes was conducted which resulted in elimination of multiple measures and merging of some measures within a given category. For example, 3 separate well-child care visit measures that apply to different age groups were combined into one measure for voting purposes. Similarly, a number of measures of the quality of services to promote healthy birth were eliminated, narrowing measures in this area to 4, including one outcome measure (low birth weight). Measures in each legislative category or subcategory (eg, prevention/health promotion, care of children with chronic disease) were rank ordered within the category or subcategory. Lowest scoring measures were eliminated from further consideration. This process resulted in 31 measures for final consideration.

### *GETTING TO 25 MEASURES TO RECOMMEND TO THE AHRQ NAC*

Three rounds of voting were conducted in succession on the 31 remaining measures. SNAC members could vote for their top 20 measures out of the 31 that remained. In round 1, members individually voted for their top 10 measures; in round 2 their next 5 measures; and in round 3 their final 5 measure choices. Measures voted for in the first round received 3 points per vote, measures voted for in the second round received 2 points per vote, and measures voted for in the third round received 1 point per vote. A priority score was then calculated for each measure that represented the total points assigned to that measure by SNAC members after the 3 rounds of voting. The final rank order of the measures based on priority scores was examined by the SNAC to assess how the acceptance of various cut-points (ie, 10, 15, 20, 25 total measures) would fulfill the goal of arriving at a grounded, parsimonious, balanced core set of measures.<sup>22</sup>

## **RESULTS**

Among the 119 measures nominated for consideration by the SNAC, information needed to assess the validity,

**Table 2.** Oxford Centre for Evidence-Based Medicine (CEBM) Evidence Grades

| Evidence Grade | Definition of Grade   | Definition of Study Type   |
|----------------|---|--|
| A              | Consistent level 1 studies  | Level 1: Randomized controlled trials                                      |
| B              | Consistent level 2 or 3 studies or extrapolations* from level 1 studies           | Level 2: Cohort studies; outcome research<br>Level 3: Case-control studies |
| C              | Level 4 studies or extrapolations from level 2 or 3 studies                       | Level 4: Case series   |
| D              | Level 5 evidence or troublingly inconsistent or inconclusive studies of any level | Level 5: Expert consensus opinion  |

\*Extrapolations are where data are used in a situation that has potentially clinically important differences compared to the original study situation.

feasibility, and importance of the measures was incomplete in some cases (Table 3).

The SNAC ultimately voted to recommend 25 measures (Table 4), including 12 for prevention/health promotion, 5 for acute care, 5 for chronic condition management, 2 to assess family experiences of care, and 1 focused on availability of services. These represent a set of measures that address care across multiple settings (ambulatory [primary care, specialty care], emergency department, and inpatient); multiple conditions (pregnancy, vaccine-preventable conditions, sexually transmitted infections, overweight, social and behavioral developmental delays, dental care, appropriate antibiotic use for respiratory conditions, asthma, diabetes, and mental health); and multiple ages (in utero, newborn, early childhood, school age, and adolescence). The well-child care visit, general screening, and family experiences of care measures are relevant to all 37.3 million children enrolled in Medicaid or CHIP.<sup>11</sup> Although some measure topics address conditions with relatively low prevalence (eg, children with diabetes, children with central lines hospitalized in intensive care units), the costs of inadequate or unsafe treatment for these children are high.

In terms of levels of evidence for underlying scientific soundness of the measures (Table 2), 2 measures are Oxford CEBM grade A (8%), 14 are grade B (56%), 2 are both B and D (8%), 1 is grade C (4%), 3 are grade D alone (12%), and 3 could not be graded (12%) as a result of insufficient evidence available to make this assessment.

Thirteen (52%) of the measures are currently National Committee for Quality Assurance (NCQA) Healthcare Effectiveness Data and Information Set (HEDIS) measures

**Table 3.** Characteristics of the Nominated Measures Related to Validity, Feasibility, and Importance Criteria (N = 119)

| Criteria                    | n (%)   |
|-----------------------------|---------|
| Missing information         |         |
| No specifications           | 26 (22) |
| No reliability data         | 59 (50) |
| No validation data          | 42 (35) |
| No disparities data         | 76 (64) |
| Not in use                  | 29 (24) |
| Evidence grade              |         |
| A                           | 18 (15) |
| B                           | 58 (49) |
| C                           | 3 (3)   |
| D                           | 26 (22) |
| No evidence/unable to grade | 14 (12) |

reported by Medicaid managed care plans; 2 are CMS–Early Periodic Screening, Diagnosis, and Treatment (EPSDT) program measures (both dental); 3 are measures submitted by individual state Medicaid or CHIP programs; 2 are measures developed by the American Medical Association Physician Consortium for Performance Improvement; 1 is a measure used by state Medicaid programs under a grant program; 1 is a measure derived from national vital statistics using state birth certificate data; 1 as nominated is stewarded by the California Maternal Quality Collaborative; and 1 is a measure used by the National Health Safety Network at the Centers for Disease Control and Prevention (Table 4).

Ten of the measures as nominated currently use a hybrid of administrative and medical records data, 8 require administrative data only, one uses a combination of Medicaid enrollment data and state birth certificate data, one uses survey data, and others use other combinations. Information on the sources of data for some measures was not provided.

All but 3 of the SNAC-recommended measures were accepted by HHS for posting for public comment.<sup>23</sup> The deleted measures were “child and adolescent suicide risk assessment” (not in current use, except as a Physician Quality Reporting Initiative pilot measure for adults)<sup>24</sup>; the Consumer Assessment of Healthcare Providers and Systems (CAHPS) Clinician and Group Survey (insufficient documented experience in pediatric care for Medicaid and CHIP enrollees); and the NCQA HEDIS annual dental visit measure (more useful information provided by the EPSDT measures of preventive and treatment services; EPSDT measures already reported by Medicaid programs and future requirement for CHIP programs [CHIPRA requirement]). In addition, HHS decided to list the 3 well-child care visit measures individually. These changes resulted in a total of 24 measures being posted for public comment.<sup>23</sup>

## DISCUSSION

Taken together as a set, the measures are relevant to evaluate the quality of many aspects of care for all 37.3 million children enrolled in Medicaid or CHIP.<sup>11</sup> The work of the SNAC represents a survey of currently available measures, careful evaluation of available evidence, and expert judgment to create a grounded and parsimonious core measurement set, all done in a publicly transparent way.

**Table 4.** SNAC-Recommended Initial Core Set of Children's Health Care Quality Measures, Grouped by Legislative Topic\*

| Legislative Topic Area/Subtopic/Brief Measure Label  | Current Measure Steward |
|--|-------------------------|
| <b>Prevention and Health Promotion</b>   |                         |
| Prenatal/perinatal   |                         |
| Frequency of ongoing prenatal care   | NCQA                    |
| Timeliness of prenatal care  | NCQA                    |
| Percentage of live births weighing less than 2500 g  | NVSS                    |
| Cesarean rate for low-risk first-birth women   | CMQC                    |
| Immunizations  |                         |
| Immunizations for 2-year-olds  | NCQA                    |
| Adolescent immunization  | NCQA                    |
| Screening  |                         |
| Body mass index documentation for 2–18-year-olds   | NCQA                    |
| Rates of screening using standardized screening tools for potential delays in social and emotional development   | None†                   |
| Chlamydia screening for 16–20-year-old women   | NCQA                    |
| Well-child care  |                         |
| Well-child visits include 3 NCQA measures combined into one measure: 1) well-child care visits in the first 15 months of life; 2) well-child care visits in years 3, 4, 5, and 6 of life; 3) adolescent well-child care visits | NCQA                    |
| Dental   |                         |
| Total eligible to receive preventive dental services (EPSDT measure Line 12B)  | States/CMS              |
| Annual dental visit  | NCQA                    |
| <b>Management of Acute Conditions</b>  |                         |
| Upper respiratory—appropriate use of antibiotics   |                         |
| Pharyngitis—appropriate testing  | NCQA                    |
| Otitis media with effusion—avoidance of inappropriate use of systemic antimicrobial drugs  | AMA/PCPI                |
| Dental   |                         |
| Total EPSDT eligibles who received dental treatment services (EPSDT CMS Form 416 Line 12C)   | States/CMS              |
| Emergency department   |                         |
| Emergency department utilization—average number of emergency room visits per member per reporting period   | S/ME                    |
| Inpatient  |                         |
| Pediatric catheter-associated bloodstream infection rates (PICU and NICU)  | Hospitals/CDC           |
| <b>Management of Chronic Conditions</b>  |                         |
| Asthma   |                         |
| Annual number of asthma patients (>1 year old) with >1 asthma-related emergency room visit   | S/AL                    |
| ADHD   |                         |
| Follow-up care for children prescribed ADHD (medication continuation and maintenance phase)  | NCQA                    |
| Mental health  |                         |
| Child and adolescent major depressive disorder—suicide risk assessment   | AMA/PCPI                |
| Follow-up after hospitalization for mental illness   | NCQA                    |
| Diabetes   |                         |
| Annual hemoglobin A1C testing (all children and adolescents diagnosed with diabetes)   | S/AL                    |
| <b>Family Experiences With Care</b>  |                         |
| HEDIS CAHPS Medicaid 4.0 including supplemental items for children with chronic conditions   | NCQA                    |
| CAHPS Clinician and Group Survey for primary care practitioners participating in Medicaid and CHIP   | CAHPS                   |
| <b>Availability of Services</b>  |                         |
| Access to primary care practitioners, by age and total   | NCQA                    |

\*NCQA = National Committee for Quality Assurance; NVSS = National Vital Statistics System; CMQC = California Maternal Quality Collaborative; EPSDT = Early Periodic Screening, Diagnosis, and Treatment; CMS = Centers for Medicare & Medicaid Services; AMA/PCPI = American Medical Association Physician Consortium for Performance Improvement; S/ME = State of Maine; PICU = pediatric intensive care unit; NICU = neonatal intensive care unit; ADHD, attention-deficit/hyperactivity disorder; CDC = US Centers for Disease Control and Prevention; S/AL = State of Alabama; CAHPS = Consumer Assessment of Healthcare Providers and Systems.

†Developed with support from The Commonwealth Fund, which is also supporting implementation and testing in state Medicaid programs and elsewhere. States have different data collection strategies.

When considering the measures individually and as a group, it is critical to note a number of themes identified by the SNAC, Medicaid and CHIP officials, and the public.

#### WHAT THIS PROCESS ACCOMPLISHED

The initial recommended core set and its voluntary implementation provide a solid first step that will allow for progression toward the stated CHIPRA goal of establishing a measurement system that allows for feasible and valid comparisons of child health care quality across all populations, states, programs, and providers. The

recommended list of measures (and the compilations of measures from which it was drawn) represents a significant achievement of a number of persistent and talented communities: those intensely engaged in measuring and improving health care quality; child health researchers; child health advocates; state public program officials; representatives of Federal programs that support the development and dissemination of children's health care quality measures; and many others. The process of identifying the core measurement set brought together the expertise of these talented but formerly fragmented groups

to take this first important step as laid out by the CHIPRA legislation.

#### **WHAT THE PROCESS DID NOT DIRECTLY ADDRESS: CHALLENGES OF IMPLEMENTATION**

In relation to the multiple health care needs of children and the CHIPRA legislation measurement domains, the initial recommended core set of 25 measures may seem relatively small. However, to many who are being asked to implement the measures, the number may seem quite large. It is clear that implementation will not be automatic, and the number of measures in the recommended initial core set should be considered in the overall context of the CHIPRA legislation and other legislative and Federal Executive Branch initiatives, as well as in the context of the economic crisis faced by most states.<sup>25</sup>

By law, measures in the core set are intended for voluntary use; states, health plans, and providers are not required to use them. CMS will assist states by identifying standard measure specifications and providing focused, tailored technical assistance on information systems and measurement. States will receive a matching Federal Medical Assistance Percentage for the "effective collection and reporting of measures." With the implementation of the American Recovery and Reinvestment Act of 2009 for health IT adoption, state concerns about the burden of using medical records for quality measurement should lessen over time. Eligible providers will receive incentive payments for "meaningful use" of measures and information technology. Additional incentives for use of a core set of measures with standardized specifications will include the ability of states to benchmark their own performance against aggregated data. Some states are already using many of the measures identified for the initial core set. Inevitably, implementation of the core measure set will take place over time, using a carefully staged process.

#### **REFINEMENTS AND ENHANCEMENTS TO THE CORE SET ARE NEEDED**

Implementation of the recommended core set of measures as currently specified will not be sufficient to achieve the vision of a comparable, evidence-based, understandable set of measures that can identify racial and ethnic, socioeconomic, and health condition disparities in health care for children. In making its recommendations for the initial core measure set, the SNAC emphasized that the measures would need to be reconfigured to be able to reflect children's health care quality across all Medicaid and CHIP programs, providers, consumers, and intermediaries (eg, health plans contracting with state Medicaid programs). Modifying the measures for a more comprehensive set of programs and beneficiaries will take time and resources.

One example of the need for modification is the NCQA HEDIS measures, which make up a large proportion of the initial, recommended core measure set. NCQA HEDIS measures are currently used for reporting by Medicaid managed care plans and, as currently specified, cannot be

used for children enrolled in CHIP or children in states with limited or no Medicaid managed care. In 2008, 51% of children enrolled in Medicaid and CHIP (17 million for Medicaid and 5.3 million for CHIP) were enrolled in managed care organizations.<sup>26</sup> Many states have no managed care presence.

Other measures, specifically those that use data from state birth certificates reported to the National Vital Statistics System (NVSS), those used by nongovernment entities (eg, California Maternal Quality Collaborative), and those currently in use by selected states vary across states in measurement methodology or do not appear to be in use by states as currently specified.

#### **WORK STILL TO BE DONE**

Few of the proposed measures are currently used to report data that distinguish care quality by race, ethnicity, tribe, socioeconomic status, or special health care need status among children, all of which are required by the CHIPRA legislation and are critical in light of the demographics of the US child population.<sup>27,28</sup> Attention to improving the capacity of measures and data sets to assess disparities is needed. For all measures, a common duration of enrollment calculation is essential to make valid and reliable assessments of health care quality in programs and populations and comparisons across and within institutions, programs, and states. Additionally, the SNAC agreed that it is critical to identify the appropriate entities that should be held accountable for health care quality in the multi-layered approach to care delivery.<sup>29</sup>

In addition to the work needed on the initial core set of measures, new measure development is needed in several areas. The SNAC noted in particular the need for additional measures of specialty care, inpatient care, substance abuse care, and mental health treatment, as well as measures that link mainstream clinical care with other services that children receive (eg, coordination of care), health outcome measures, and measures of the medical home.

Fortunately, CHIPRA provided support for advancing and improving pediatric quality measures and called for priorities to be set to guide a new pediatric quality measures program. States are also encouraged to experiment with and evaluate the core measure set using funds from the CMS quality demonstration projects.<sup>30</sup>

#### **LIMITATIONS**

The efforts of the SNAC were limited to some degree by the time available to arrive at the recommended core set of quality measures. Some of the information the SNAC needed to objectively assess the merits of nominated measures was not available during their deliberations (Table 3); as noted below, some information would not be available for years, even assuming that action was taken to develop the needed data. Similar efforts in the future could be conducted over a longer timeframe, allowing for establishment of consensus around measure evaluation criteria before initiating any measure assessment activities. Additionally, the call for measures, the nomination

process, and the process of summarizing submitted information could be completed before engaging in assessments of the measures under consideration to increase efficiency and decrease rework. That said, the abbreviated timeline had its advantages, in that it helped focus the process and sharpen deliberations. Past efforts to identify core sets of measures, absent the CHIPRA requirements (eg, evidence based, requirement for public posting in less than 10 months from passage of the legislation) while able to make considerable progress, have not resulted in the measurement consistency desired by many, including some state programs.<sup>2,31–36</sup>

Other factors contributing to a less than ideal process could not have been remedied with more time (unless time was measured in years). These include the lack of a timely and detailed compilation of health care quality measures in use by state Medicaid and CHIP programs (to identify measures in use) and the paucity of a research base to establish the underlying scientific soundness for many measures of the quality of services used or potentially used by many Medicaid and CHIP enrollees (eg, perinatal services, mental health services, periodicity and content of well-child care visits, coordination efforts, and dental treatment).

### HOW THE MEASURES COULD BE USED

If most states can overcome the aforementioned challenges to broad implementation of the core set of quality measures, this measurement set will provide a method to routinely assess our nation's child health care quality by means of standardized methods. The availability of these national and local benchmarking data could facilitate identification of quality deficits at multiple different levels where the core set of measures could be used (eg, health care systems, health plans, health centers, and individual providers). This routine measurement could inform and hopefully drive quality improvement efforts at all of these levels. That said, this list of 25 quality measures is far from comprehensive. This lack of comprehensiveness may have the unintended consequence of states, programs, health plans, and providers focusing their improvement efforts on this limited set of health care issues which may ultimately take resources away from other areas of care that need attention. However, even improvements for this limited set of health care conditions would be a positive first step toward improving the quality of care children receive in the US. Implementation of this initial set of measures should also allow organizations to establish the necessary infrastructure to later implement a more comprehensive, expanded core set of measures, including those developed under funding from the Pediatric Quality Measures Program mandated by the CHIPRA and administered through AHRQ.

### CONCLUSION

An open, national, public process combined with an evidence-informed evaluation methodology resulted in identification of a balanced, grounded, and parsimonious

core set of recommended measures that should become feasible to implement on a widespread scale over time.

As stimulated by CHIPRA and building on AHRQ's, CMS's and the states' long-standing interest in health care quality improvement, HHS and the states are rejoining their efforts to use these recommended measures to identify areas in need of improvement and monitor progress toward the goal of a high quality health care system for all children.

### ACKNOWLEDGMENTS

The activities described herein were supported by the US Department of Health and Human Services' Agency for Healthcare Research and Quality and the Centers for Medicare & Medicaid Services.

### REFERENCES

1. Ryan J. The Children's Health Insurance Program (CHIP): the fundamentals. National Health Policy Forum. Background paper no. 68. April 2009.
2. deLone SE, Hess CA. Medicaid and CHIP Children's Health Care Quality Measures: What States Use and What They Want. *Acad Pediatr*. 2011;11:S68–S76.
3. First annual report on the quality of care for children in Medicaid and CHIP. Available at: [http://www.cms.gov/MedicaidCHIPQualPrac/02\\_Spotlight.asp](http://www.cms.gov/MedicaidCHIPQualPrac/02_Spotlight.asp). Accessed October 25, 2010.
4. Public Law 111-3. See Title IV for quality provisions. Available at: [http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=111\\_cong\\_public\\_laws&docid=f:publ003.111](http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=111_cong_public_laws&docid=f:publ003.111). Accessed October 15, 2010.
5. CHIPRA HHS Federal Quality Workgroup by operating division/staff division. Available at: <http://www.ahrq.gov/chipra/corebackground/corebackapa4.htm>. Accessed October 15, 2010.
6. National Advisory Council Subcommittee on Quality Measures for Children's Healthcare in Medicaid and the Children's Health Insurance Program (SNAC) member list. Available at: <http://www.ahrq.gov/chipra/chipralist.htm>. Accessed October 15, 2010.
7. Background report for the request for public comment on the initial, recommended core set of children's healthcare quality measures for voluntary use by Medicaid and CHIP. Available at: <http://www.ahrq.gov/chipra/corebackgrnd.htm>. Accessed October 15, 2010.
8. Brook RH. The RAND/UCLA appropriateness method. In: McCormick KA, Moore SR, Siegel RA, eds. *Clinical Practice Guidelines Development: Methodology Perspectives*. Rockville, Md: Agency for Health Care Policy and Research; 1994.
9. Hemingway H, Crook A, Feder G, et al. Underuse of coronary revascularization procedures in patients considered appropriate candidates for revascularization. *N Engl J Med*. 2001;344:645–654.
10. Kravitz RL, Park RE, Kahan JP. Measuring the clinical consistency of panelists' appropriateness ratings: the case of coronary artery bypass surgery. *Health Policy*. 1997;42:135–143.
11. Shekelle PG, Chassin MR, Park RE. Assessing the predictive ability of the RAND/UCLA appropriateness method criteria for performing carotid endarterectomy. *Int J Technol Assess Health Care*. 1998;14:707–727.
12. Clancy C. Introductory remarks and charge to the Subcommittee. Available at: <http://www.ahrq.gov/chipra/slides/chipra072209/clancyslides-contents.htm>. Accessed October 15, 2010.
13. Mann C, Dailey B. Overview of the Centers for Medicare & Medicaid Services (CMS) and Agency for Healthcare Research and Quality (AHRQ) partnership for improving the quality of health care for children in Medicaid and Children's Health Insurance Program (CHIP). Available at: <http://www.ahrq.gov/chipra/slides/chipra072209/mannslides-contents.htm>. Accessed October 15, 2010.
14. Evaluation criteria and instructions for Delphi, round II. Available at: <http://www.ahrq.gov/chipra/corebackground/corebackapa8.htm>. Accessed October 15, 2010.

15. National Survey on Children's Health 2007. Available at: <http://cahmi.org/pages/Home.aspx>. Accessed October 15, 2010.
16. Webcast. Available at: <http://www.connectlive.com/events/ahrq2009/>. Accessed October 15, 2010.
17. Measure nomination template. Available at: <http://www.ahrq.gov/chipra/corebackground/corebackapa7.htm>. Accessed October 15, 2010.
18. One-page summary sheet. Available at: <http://www.ahrq.gov/chipra/corebackground/corebackapa9.htm>. Accessed October 15, 2010.
19. Measures that met the Delphi II criteria for discussion at the September 17–18 SNAC meeting but were not part of the SNAC's top 25 initial, core recommended set. Available at: <http://www.ahrq.gov/chipra/corebackground/corebackapb.htm>. Accessed October 15, 2010.
20. Measures that did not meet the Delphi II criteria for discussion and voting at the September 17–18 SNAC meeting. Available at: <http://www.ahrq.gov/chipra/corebackground/corebackapb2.htm>. Accessed October 15, 2010.
21. McGlynn EA, Koseoff J, Brook RH. Format and conduct of consensus development conferences: a multi-nation comparison. In: Goodman C, Baratz S, eds. *Improving Consensus Development for Health Technology Assessment*. Washington, DC: National Academy Press; 1990.
22. SNAC-recommended initial core set of children's healthcare quality measures, in priority order by number of SNAC votes, September 17–18, 2009. Available at: <http://www.ahrq.gov/chipra/listtable.htm>. Accessed October 15, 2010.
23. Agency for Healthcare Research and Quality. Activities related to the Children's Health Insurance Program Reauthorization Act (CHIPRA) Web site. Available at: <http://www.ahrq.gov/chipra>. Accessed October 27, 2010.
24. Carlucci C. AMA PCPI measures to be considered for SNAC deliberations. In: American Medical Association, ed. *The Physicians Consortium for Practice Improvement*. Chicago, Ill: AMA PCPI; 2009:351.
25. Kaiser Commission on Medicaid and the Uninsured. State fiscal conditions and Medicaid. Medicaid facts. Menlo Park, Calif: The Henry J. Kaiser Family Foundation. September 2009. Available at: <http://www.kff.org/medicaid/7580.cfm>. Accessed October 25, 2010.
26. Centers for Medicare & Medicaid Services. Enrollment by delivery type, Medicaid and CHIP children, 2008. Excel spreadsheet provided in personal communication via e-mail from CMS staff Lekisha Daniel-Robinson to AHRQ staff Denise Dougherty, January 2009.
27. United States: People in Poverty—Poverty Rate by Age, States (2007–2008), US (2008). Menlo Park, Calif: The Henry J. Kaiser Family Foundation. Available at: <http://www.statehealthfacts.org/profileind.jsp?cat=1&sub=2&rgn=1>. Accessed October 25, 2010.
28. United States: Population Distribution of Children by Race/Ethnicity, States (2007–2008), US (2008). Menlo Park, Calif: The Henry J. Kaiser Family Foundation. Available at: <http://www.statehealthfacts.org/profileind.jsp?ind=7&rgn=1&cat=1>. Accessed October 25, 2010.
29. Berwick D. A user's manual for the IOM's "Quality Chasm" Report. *Health Affairs*. May/June 2002;21:80–90.
30. Centers for Medicare & Medicaid Services. Invitation to apply for FY2010 CHIPRA Quality Demonstration Grants (CFDA 93.767). September 30, 2009. Available at: [http://www.cms.hhs.gov/CHIPRA/Downloads/CHIPRA\\_Quality\\_Demo.pdf](http://www.cms.hhs.gov/CHIPRA/Downloads/CHIPRA_Quality_Demo.pdf). Accessed October 25, 2010.
31. Fairbrother G, Simpson L. Measuring and Reporting Quality of Health Care for Children: CHIPRA Beyond. *Acad Pediatrics*. 2011; 11:S77–S84.
32. Smith VK, Edwards JN, Reagan E, Roberts D. *Medicaid and CHIP Strategies for Improving Child Health*. New York, NY: The Commonwealth Fund; July 2009.
33. Rosenbach M, Irvin C, Merrill A, et al. *National Evaluation of the State Children's Health Insurance Program: A Decade of Expanding Coverage and Improving Access*. Washington, DC: Mathematica Policy Research; September 2007.
34. Partridge L. *Review of Access and Quality of Care in SCHIP Using Standardized National Performance Measures*. Washington, DC: George Washington University; April 4, 2007.
35. National Quality Forum. Child Healthcare Quality Measurement and Reporting: Workshop Proceedings. 2004. Available at: [http://www.qualityforum.org/Publications/2004/07/Child\\_Healthcare\\_Quality\\_Measurement\\_and\\_Reporting.aspx](http://www.qualityforum.org/Publications/2004/07/Child_Healthcare_Quality_Measurement_and_Reporting.aspx). Accessed October 27, 2010.
36. Simpson L, Dougherty D, Krause D, et al. Measuring children's health care quality. *Am J Med Qual*. 2007;22:80–84.