National Quality Measures for Child Mental Health Care: Background, Progress and Next Steps

Bonnie T. Zima, MD, MPH (RWJCS, UCLA 1991)
Professor-in-Residence, Child and Adolescent Psychiatry
Associate Director, UCLA Center for Health Services & Society
UCLA Semel Institute for Neuroscience and Human Behavior

J. Michael Murphy, EdD
Staff psychologist, Massachusetts General Hospital
Associate Professor of Psychology, Harvard Medical School

Sarah Hudson Scholle, MPH, DrPH
Vice President, Research and Analysis
National Committee for Quality Assurance

Kimberly Eaton Hoagwood, PhD
Professor of Clinical Psychology and
Vice Chair for Research
Department of Child Psychiatry
New York University Medical Center

Ramesh C. Sachdeva, MD, PhD
Professor of Pediatrics
Medical College of Wisconsin
Corporate VP and Chief Quality Officer
Children’s Hospital and Health System

Rita Mangione-Smith, MD, MPH (RWJCS, UCLA 1997, RWJ Generalist Physician
Faculty Scholar 2000-04)
Professor of Pediatrics
University of Washington
Seattle Children's Research Institute
Center for Child Health, Behavior, and Development

Donna Woods, EdM, PhD
Research Associate Professor
Institute for Healthcare Studies
Feinberg School of Medicine
Northwestern University

Hayley S. Kamin, BA
PhD Student, Department of Psychology, University of Florida

Michael Jellinek, MD
Chief of Clinical Affairs, Partners Healthcare
Professor of Psychiatry and Pediatrics, Harvard Medical School

Abbreviations: ADHD=Attention-deficit/Hyperactivity Disorder

Key words: quality measures, child mental health, ADHD, depression, clinical validity, quality improvement research

Financial disclosure: This study was supported by the National Institute of Mental Health (P30 MH082760), the Agency for Healthcare Research and Quality (1U18HS020506, U18 HS020503, 1U18HS020498).

Conflicts of interest: Drs. Zima, Murphy, Scholle, Hoagwood, Sachdeva, Woods, Mangione-Smith and Jellinek, and Ms. Kamin report no biomedical financial interests or potential conflicts of interest.

Correspondence: Dr. Bonnie T. Zima, UCLA Center for Health Services and Society, 10920 Wilshire Blvd. #300, Los Angeles, CA 90024; T: 310 794 3714; F: 310 794 3724; bzima@mednet.ucla.edu

The authors would like to thank Evan M. Williamson, MPH, MS for his technical consultation on the evidence bases for the NQF Behavioral Health Measures, and thank and acknowledge the work of all who participated in the ADHD Measurement Leadership Team for their contribution to evidence-based pediatric ADHD diagnosis, follow-up and treatment quality measure development. Mark Antman and Molly Siegel from the American Medical Association, Physician Consortium for Performance Improvement; Jonathan Klein, Fan Tait and Keri Theissen from the American Academy of Pediatrics and Nicole Muller and Caroline Mazurek from the Institute for Healthcare Studies, in the Feinberg School of Medicine at Northwestern University and Mark Wolraich and Karen Pierce, the Co-Chairs of the ADHD Measures Expert Workgroup.
Contributor’s Statement

Bonnie Zima, MD, MPH: Drafted and submitted an abstract for consideration for manuscript submission, developed the conceptual framework, provided oversight to tabulations, coordinated coauthor contributions, drafted earlier versions of the manuscript and made final edits.

J. Michael Murphy, EdD: Provided consultation to the paper’s conceptual framework, oversight on the literature review, and participated in writing early and final manuscript drafts.

Sarah Hudson Scholle, MPH, DrPH: Provided consultation on the development of new depression screening measures, and participated in writing early and final manuscript drafts.

Kimberly Eaton Hoagwood, PhD: Provided consultation on the conceptual framework and development of new depression screening measures, and participated in writing early and final manuscript drafts.

Ramesh C. Sachdeva, MD, PhD: Provided consultation on the refinement of ADHD quality measures, and participated in writing early and final manuscript drafts.
Rita Mangione-Smith, MD, MPH: Provided consultation on the refinement of the algorithm to identify children with complex health care needs, and participated in writing early and final manuscript drafts.

Donna Woods, EdM, PhD: Provided consultation on the refinement of the ADHD quality measures, and participated in writing early and final manuscript drafts.

Hayley S. Kamin, BA: Conducted literature reviews, and participated in writing early and final manuscript drafts.

Michael Jellinek, MD: Participated in development of the conceptual framework, and the writing of early and final manuscript drafts.
What’s Known, What’s New

What’s Known on This Subject

Between 2008 and 2011, a number of child mental health care quality measures were recommended for national use by multi-stakeholder groups convened through The Children’s Health Insurance Project Reauthorization Act (CHIPRA) and/or the National Quality Forum.

What’s New

The development of new child mental health quality measures poses methodological challenges that will require a paradigm shift to align research with its accelerated pace.
Abstract

**Objective:** To review recent health policies related to measuring child health care quality, the selection processes of national child health quality measures, the nationally recommended quality measures for child mental health care and their evidence strength, the progress made toward developing new measures, and early lessons learned from these national efforts.

**Methods:** Description of the selection process of child health care quality measures from two independent national initiatives, the recommended quality measures for child mental health care, and the strength of scientific evidence supporting them.

**Results:** Of the child health quality measures recommended or endorsed during these national initiatives, only nine unique measures were related to child mental health.

**Conclusions:** The development of new child mental health quality measures poses methodological challenges that will require a paradigm shift to align research with its accelerated pace.
Recent health policies have accelerated the development and use of quality measures for children receiving publicly-funded care.\textsuperscript{1, 2} In response, a legislatively mandated national committee and a non-profit organization systematically rated large pools of quality measures and recommended a limited number to monitor the quality of care received by U.S. children. Although these initiatives were independent and used different approaches to select and rate child health care quality measures, each recommended few measures related to child mental health care.\textsuperscript{3, 4} This gap is of public health significance because improving the quality of child mental health care is a long-standing national priority\textsuperscript{5-9} and there is substantial room for improvement in mental health care for both private and publicly-insured populations.\textsuperscript{10-18} This paper reviews:

- recent relevant health policy initiatives;
- the selection of national child health quality measures;
- existing national standards for child mental health care including the strength of the evidence supporting them;
- an update on development of new quality measures related to child mental health care; and,
- early lessons learned from these national efforts.

**Background**

The Children’s Health Insurance Program Reauthorization Act of 2009 (CHIPRA) called for identification, refinement and development of child health care quality measures for voluntary use in Medicaid and Children’s Health Insurance Programs.\textsuperscript{19} Developed under the auspices of the Agency for Healthcare Research and
Quality (AHRQ), an initial core set of twenty-four quality measures was submitted to the Secretary of the U.S. Department of Health and Human Services on January 1, 2010. For the subsequent Pediatric Quality Measures Program, 55 million dollars were made available to support seven Centers of Excellence in 2010 to develop new measures and refine existing ones for potential core set enhancements in January 2013, 2014, and 2015.


Accessed 9/5/12

Under the leadership of the Centers for Medicare & Medicaid Services (CMS), CHIPRA also funded 10 five-year demonstration projects to states at an estimated total cost of $100 million on February 2010, of which seven propose to develop, test, evaluate and/or report adherence to quality measures. Outreach and technical assistance efforts to the states to report on adherence to 12 of the 24 measures in the initial core set began in 2011. The use of the measures is likely to be sustained through financial incentives to collect and report on adherence rates on quality indicators through a matching Federal Medical Assistance Percentage that is part of the American Recovery and Reinvestment Act of 2009 (ARRA). Eligible providers will receive these payments for demonstrating “meaningful use” of quality measures under the Electronic Health Records Incentive Program and are anticipated to be given the capacity to benchmark their own performance against aggregated data. Together, these activities are envisioned to be “the first steps taken” to reach the goal of a quality-driven, evidence-based national system of child health care.
Consistent with this vision, the National Quality Strategy (NQS) was established “to improve the delivery of health care, services, patient health outcomes, and population health” for all Americans, as part of the 2010 Patient Protection and Affordable Care Act (ACA).\textsuperscript{2,23,24} This is the first legislation to set national goals to improve the quality of health care in public and private health care programs. It will guide all US Department of Health and Human Service quality improvement programs and regulations, and set criteria to measure the quality of health care to align with national efforts for quality improvement.\textsuperscript{24} The three aims of the NQS are to improve the overall quality of care, improve the health of the U.S. population, and reduce the cost of quality health care.\textsuperscript{\#23} To adapt the NQS for behavioral health care, the Substance Abuse and Mental Health Services Administration (SAMHSA) developed the Behavioral Health Quality Framework that tailors the six national priority areas to behavioral health care, reinforcing how the three aims of the NQS could be equally applied to the care of mental health problems.\textsuperscript{3}

Contemporaneously, the National Quality Forum (NQF) is a private, non-profit organization that was given federal funding to conduct a parallel effort to identify and endorse measures that could be used to assess the quality of children’s healthcare. The NQF is dedicated to improving the quality of American health care by: 1) building consensus on national priorities and goals for performance improvement and working in partnership to achieve them; 2) endorsing national consensus standards for measuring and publicly reporting on performance; and 3) promoting the attainment of national goals through education and outreach programs.\textsuperscript{25} As part of their mission, the NQF organized a standardized process to evaluate and endorse voluntary consensus standards for patient
outcomes for child health and mental health, and child health candidate standards. The projects were undertaken between 2009-2011 and known as the Patient Outcomes (Phase III): Child Health and Child Health Measures Projects. While specific approaches across these different national initiatives varied, they raised similar questions about how to address barriers that limit the feasibility of these quality measures, the acceptable threshold for sufficient scientific evidence for clinical validity, and how to address methodological limitations that could influence the interpretation of findings.

**Quality Measure Selection Process**

**CHIPRA: Development of Initial Core Set of Measures**

In partnership with AHRQ and CMS, the initial core measure set was identified using an evidence-informed process that integrated input from a broad array of stakeholders and public comments.\(^2^6\) A multidisciplinary AHRQ National Advisory Council Subcommittee on Children’s Healthcare Quality Measures for Medicaid and Children’s Health Insurance Programs (SNAC) was formed in May 2009. The SNAC was charged with establishing quality measure evaluation criteria, identifying a strategy for gathering measures, and applying the evaluation criteria to the measures.\(^2^6\) It was comprised of multiple stakeholders, including officials from publicly-insured programs, national professional organizations, and child and family advocacy organizations as well as national experts in health care quality measurement.\(^2^2\)

Over a four-month period, the SNAC held two public meetings and undertook substantial work outside of these meetings. This work included assessing an initial set of
quality measures in use by Medicaid and CHIP programs using an adapted version of the RAND/UCLA modified Delphi method, identifying a process to supplement these measures through a public call for nominations, and subsequently assessing the nominated measures using the same modified Delphi method. The RAND/UCLA appropriateness method is a well-established approach that integrates scientific evidence with expert clinical judgment\textsuperscript{27} and that has been successfully used to assess the quality of outpatient general health care among children nationally.\textsuperscript{28} It has also been used to assess the quality of mental health care statewide among children receiving publicly-funded outpatient specialty mental health care.\textsuperscript{18} The process integrates a review of the evidence-base for a proposed measure, and two rounds of structured expert ratings. During this process, the SNAC assessed the validity, feasibility, and importance of 119 measures, of which 12 were specific to child mental health. For each measure, the SNAC rated the level of scientific evidence supporting the measure, feasibility of implementing the measure, and the measure’s importance. When considering importance, highest priority was given to measures that were deemed actionable (by which the SNAC meant the extent to which a publicly insured program would likely be able to improve their performance) and likely to substantially reduce health care costs.

The initial modified Delphi process reduced the pool of candidate measures under consideration to 70. During the second public meeting, a series of private electronic votes were conducted to eliminate overlapping measures, merge conceptually similar measures, and prioritize the remaining pool to select the final measures. The SNAC recommended 25 measures that were then reviewed by the CHIPRA Federal Quality Workgroup, Medicaid and CHIP officials, and other key stakeholders. From this process,
two measures were dropped due to lack of field testing, including one pertaining to suicide risk assessment for children with major depression. Details of the methods and administrative review pathways prior to final submission of the initial core set of measures are described elsewhere.\textsuperscript{22, 26, 29} 

In addition to selecting measures, the SNAC provided guidance to the Pediatric Quality Measures Program. It found that measures lacked the capacity to stratify adherence by race/ethnicity, tribe, socioeconomic state, or special health care need status, characteristics called for in the CHIPRA legislation.\textsuperscript{30, 31} Content gaps led to recommendations for new measures for substance abuse care, and mental health treatment as well as several areas not specifically related to mental health: specialty care, inpatient care, availability of services, coordination of care, medical home, family experiences of care, and outcomes.\textsuperscript{26, 29, 32, 33} Further, public agency partners recommended that quality measures be aligned with the priorities of state Medicaid and CHIP agencies \textsuperscript{34, 35} and parent and provider representatives advocated for engaging key stakeholders during the quality measure development process.\textsuperscript{36, 37}

**National Quality Forum: Endorsement of Child Health Quality Measures**

The NQF consensus development process involves nine main steps that typically occur over a 12- to 18-month time period. The steps are: 1) call for intent; 2) call for nominations; 3) call for candidate standards; 4) candidate consensus standards review; 5) public and member comment; 6) member voting; 7) Consensus Standards Approval Committee (CSAC) Decision; 8) board ratification; and 9) 30-day appeals.\textsuperscript{38} The review of the candidate standards for the aforementioned child health-related projects was conducted by steering committees comprised of child health and family advocates, health
care system and provider professional organizations, clinicians, and health care quality measurement experts. Following a set of standardized training sessions, the committee conducted a detailed review of the candidate standards during an in-person meeting with follow-up as required by conference call. Similar to the development of the CHIPRA initial core set, transparency was of high priority. The steering committee meeting was open to the public, member voting was done openly, information about the meeting was posted on the NQF website, and time for public comment was allocated on the agenda.

The measures were rated on four main criteria: 1) importance to measure and report the nominal topic; 2) scientific acceptability; 3) usability; and 4) feasibility. Within these four domains, the reviewer also rated subdomains to standardize the rationale for the main criterion rating. If the measure was deemed not to be important, the rating stopped. The extent a measure met the remaining criteria was rated on a four-point scale (i.e., completely, partially, minimally, or not at all). During the vote for recommendation for endorsement, each reviewer personally weighed his or her item ratings. Recommendations were then classified as with or without consensus by NQF staff. Details of the rating criteria used for both initiatives are summarized in Table 1. The NQF criteria are regularly updated, and more rigorous criteria for scientific acceptability are being applied for the 2012 Behavioral Health Measures Evaluation.39

Recommended or Endorsed Quality Measures for Child Mental Health Care

INSERT TABLE 1 HERE
Although the approaches varied, both processes yielded relatively few child mental health quality measures (Table 2). Of the 70 measures considered for the CHIPRA Initial Core Set, 12 pertained to child mental health care and of these, 3 were recommended. Of the 101 candidate measures reviewed during the NQF projects, 15 pertained to child mental health care. Five of these overlapped with the three CHIPRA measures, two were the same measure for two different age groups of teens, and one measured maternal mental health. Thus there were nine unique measures of the quality of child mental health care in CHIPRA and NQF combined.

For both initiatives, priority was placed on the development of a balanced set of measures to build capacity to track a wide breadth of quality care. For these measures, the age ranges varied in the specifications, such that one was restricted to children aged 0-5, two to ages 13-18 and six included all or most child age groups. The focus of concern also ranged from specific to general problem areas. Two measures focused on depression, two on attention-deficit/hyperactivity disorder, one on risky behaviors, one on suicidality, and three on general problem areas. Two of the measures involved monitoring, three called for screening, and four required clinicians to make assessments.

Evidence Strength for Child Mental Health Care Quality Measures

One potential next step for the creation of quality standards is to rate the empirical evidence that supports each measure. The Oxford Centre for Evidence-Based Medicine (CEBM)\(^40\) has put together detailed methods for conducting these kinds of ratings and all of the CHIPRA measures were reviewed according to the CEBM standards.\(^26\) The CEBM protocol involves assigning a letter grade of A (the best evidence) to D (the worst) for the
quality of the evidence for a given measure based on the types of studies which have been
done to validate its use as a standard. A letter grade of ‘A’ corresponds to ‘consistent
Level 1 studies (randomized controlled trials (RCTs)). A grade of ‘B’ corresponds to
consistent Level 2 or 3 studies or extrapolations from Level 1 studies, with Level 2
studies defined as those that include: a) systematic reviews of cohort studies, or b)
individual cohort studies (including low-quality RCTs and "outcomes" research). Level 3
studies are systematic reviews with homogeneity of case-control studies or an individual
case-control study. A grade of ‘C’ is given if there are only Level 4 studies or
extrapolations from level 2 or 3 studies, with Level 4 defined as case series and poor
quality cohort and case control studies. A grade of ‘D’ is given if the evidence is only of
Level 5 (expert opinion) or if the evidence is inconsistent or inconclusive.

As noted above, the quality of the evidence for the three CHIPRA measures had
been graded according to CEBM standards. Although one of the CHIPRA measures
received the low grade of ‘D’, two measures were graded as ‘B’, but even these measures
were noted to have limitations in the quality of their evidence. One measure had been
assessed in studies that did not specify age (‘CHIPRA #21: “Follow up after
hospitalization for mental illness”) and the other revealed ‘no data on whether screening
using standardized tools ultimately leads to better outcomes for these children’ (CHIPRA
#8; ‘Screening using standardized screening tools for potential delays in social and
emotional development”).

Although NQF did not utilize CEBM standards, there was a section on evidence
and all relevant studies on the NQF website for each measure. For the purposes of this
paper we reviewed the studies cited there and supplemented this with a review of studies
on the website of the steward listed for each measure. We also conducted a search using Ovid and PubMed studies published from 2001 to 2011 using the six measure names as specific and general search terms.

For only two of the measures did we find studies suggesting higher than a ‘D’ level of evidence: the NQF summary for “Depression Screening by 13/18 years of age” (NQF # 1394 & 1515) noted that this measure had been rated by the United States Preventive Services Task Force as having a B level of evidence, citing studies\textsuperscript{41, 42} that reported that screening instruments both performed well and increased the use of effective treatments; and that use of the ‘Pediatric Symptom Checklist was associated with increased rates of referral and improved functioning for children after intervention.\textsuperscript{43-48}

Overall, the evidence strength supporting the child mental health quality measures was variable. None of the measures were supported by research using randomized clinical trials to examine the relationship between adherence and outcomes that were meaningful to “decision makers” (i.e., parents, providers, payers)\textsuperscript{49} or impact on health.\textsuperscript{50} Such a research gap is consistent with adult mental health and substance abuse care quality measures.\textsuperscript{51}

\textbf{INSERT TABLE 2 HERE}

\textbf{Improving Quality Measures for Child Mental Health Care: Next Steps}

As part of the Pediatric Quality Measures Program, three of the Pediatric Quality Measures – Centers of Excellence received first round assignments that included the development and refinement of quality measures related to child mental health. The topic

16
areas were ADHD, depression, and identifying eligible populations for mental health quality measurement. The lead centers for these activities were respectively the AHRQ-CMS CHIPRA Pediatric Measurement Center of Excellence (PMCoE) based at the Medical College of Wisconsin (PI: Sachdeva), the AHRQ-CMS CHIPRA National Collaborative for Innovation in Quality Measurement (NCINQ) based at the National Committee on Quality Assurance (NCQA) (PI: Scholle), and the AHRQ-CMS CHIPRA Center of Excellence on Quality of Care Measures for Children with Complex Needs (COE4CCN) based at Seattle Children’s Research Institute (PI: Mangione-Smith).

Second round assignments also included topics related to child mental health care, and the AHRQ-CMS CHIPRA Mount Sinai Collaboration for Advancing Pediatric Quality Measures (CAPQuaM) (PI: Kleinman) will also develop behavioral health measures. The new areas for measure development are: 1) psychotropic (mental health) medication reconciliation; 2) follow-up after psychiatric hospitalization; 3) alcohol and substance abuse screening, use, diagnosis, treatment and follow-up; 4) developmental screening and follow-up diagnosis, treatment and management of follow-up diagnosis; 5) emergency department and hospital use and avoidable use for mental health problems; 6) adherence to recommended care processes for common mental health problems in emergency department and hospital settings; 7) antipsychotic medication management; and 8) quality for children served in child welfare. Described below are brief updates of the centers’ early activities.

Attention-Deficit/Hyperactivity Disorder

The PMCoE is working collaboratively with the American Medical Association (AMA) Physician Consortium for Performance Improvement, the American Academy of
Pediatrics and the American Board of Pediatrics and research academic centers Northwestern University and the Medical College of Wisconsin on the development and refinement of quality measures related to the care of Attention Deficit Hyperactivity Disorder (ADHD). This disorder was selected because it is prevalent, affecting an estimated 3-9% of U.S. children, one of the most common reasons children are referred for mental health services, and represents 15 -45% of the mental health conditions diagnosed in children and youth. Considerable variations and gaps in care have been documented in the literature. Priority was therefore placed on establishing metrics for effective ADHD diagnosis, follow-up, and treatment, first, as a part of the development of an initial set of 25 pediatric measures and then as an assigned topic for pediatric quality measure development and testing through the PMCoE.

Several recent studies have provided important guidance about effective ADHD diagnosis, follow-up and treatment. To incorporate the current best evidence about these topics, the American Academy of Pediatrics (AAP) conducted a two year process to revise and update the 2003 AAP ADHD guideline. The most recent ADHD guideline was published in November of 2011 making several changes to the prior guideline recommendations to direct the field towards care based on the best existing evidence through six primary recommendations. Based on these new AAP ADHD guideline recommendations investigators for Northwestern University, along with investigators and staff from the American Medical Association, the American Academy of Pediatrics and the American Board of Pediatrics established and engaged an Expert Workgroup comprised of experts across the broad spectrum of stakeholders related to the diagnosis, follow-up, and treatment of ADHD, including pediatricians, child and adolescent
psychologists, child and adolescent psychiatrists, neurologists, parents, teachers, school nurses, family physicians, and an occupational therapist. Critical changes to the AAP’s ADHD guideline recommendations included that ADHD diagnosis should be determined based on DSM-IV criteria or through the use of a validated tool based on these criteria, lowering the potential age of ADHD diagnosis to include children ages 4 and 5, and making specific recommendations for behavior therapy and medication treatment.

The draft measures address known quality gaps and variations in ADHD care in accordance with the recommendations in the new 2011 AAP ADHD guideline for effective diagnosis, follow-up, and treatment for pediatric patients, ages 4 to 18 years, after a diagnosis of ADHD has been made. Following development and specification of pediatric quality measures for ADHD, these measures will be tested for: 1) performance of the measure (using) manual chart review; 2) feasibility and validity of using the Electronic Health Record to calculate the measure; and 3) the feasibility of specifying the measures for construction using administrative data sources and the reliability of the resulting measure output.

Major Depression

The NCINQ is taking the lead on the development and refinement of quality measures related to adolescent depression. Major depressive disorder (MDD) is a disabling condition that is associated with long-term complications and may lead to suicide. MDD affects over seven percent of adolescents in the U.S. In 2006, around 2.3 million adolescents 12 to 17 years of age reported experiencing a major depressive episode at some point in their lives. Depression can have a major impact on children’s functioning, disrupting daily life at home, school or in the community, and resulting in
serious long-term morbidities such as generalized anxiety disorder and panic disorder. Depression may also lead to engagement in risky behaviors such as substance use (e.g. alcohol, illicit drugs, tobacco), and it may also lead to suicide. Suicide, the third leading cause of death among 15- to 24-year olds, is often preceded by depression or long-term MDD. Adolescent-onset depression increases the risk of attempted suicide by five-fold and is strongly correlated with chronic and recurring depression in adulthood. Further, depressive symptoms can be both prolonged and episodic, recurring over weeks and months. The Centers for Disease Control and Prevention noted that individuals who experience just one episode of depression are at a 50 percent higher risk of experiencing further episodes.

Based on a review of all major guidelines, evidence reviews, and advice from family partners, clinicians and researchers NCQA has developed a logic model for adolescent depression management and follow-up. This model addresses several key aspects of management including: 1) screening and assessment; 2) treatment options and initiation of treatment; and 3) symptom monitoring, treatment course and remission. The logic model uses a “measurement-based care” approach to conceptualize the steps involved in optimizing care. For depression management, measurement-based care starts with use of standardized tools to screen for depression in primary care, followed by confirmatory assessment and monitoring of symptom and functioning throughout the episode of depression to guide treatment decisions and to assess response and remission. The model also acknowledges that successful implementation depends on adequate readiness of primary and specialty providers. NCINQ stakeholder panels provided
feedback both on the overall approach and to identify the most salient opportunities where quality measures are likely to improve quality and outcomes.

Identifying Eligible Populations for Mental Health Quality Measurement

The COE4CCN is working to develop several measures intended to advance quality measurement in the area of general child mental health care. One of the Center’s early efforts has focused on ways of coding the presence of mental health conditions based on diagnostic codes available in administrative data. Use of these codes to identify children with mental health problems will go through a process of validation using abstracted medical record data as the gold standard. If the methodology developed is found to be valid, it will then be further tested and refined using existing, large datasets like Medicaid claims from entire states. These analyses are being carried out using data from one state Medicaid agency as well as a large urban tertiary care children’s hospital.

Through this approach, the COE4CCN is working to build capacity to use existing data infrastructure to identify children with mental health conditions, describe the services delivered and explore new approaches to link measure adherence with clinical outcomes. The long-accepted observation that mental health problems are under-recognized in pediatrics suggests that the prevalence of child mental health problems may be underestimated. Delivery of mental health care may also be under-reported because procedure codes for evidence based mental health care are often missing in Medicaid claims data. Nevertheless, this new direction has the potential to bring a kind of ‘parity’ with physically based medical diagnoses in the identification of mental health problems. Secondly, an algorithm to identify children with “social complexity” using Medicaid claims and enrollment data is under development. For the purposes of
this project, social complexity is defined as the presence of one or more social risk factors hypothesized to be strong correlates of mental health. Valid identification of social complexity may enhance the identification of mental health problems that may be under-reported as diagnoses in Medicaid service encounter data, stand in as a proxy, or serve as a marker for children at risk for mental health problems who might benefit from early preventive interventions. Data sources will include Medicaid claims and encounter data from one state and surveys from parents and health care providers.

**Early Lessons Learned**

The inclusion of quality measures related to child mental health care and recent priority placed on developing new ones are major advances that are consistent with the recommended trajectory of integrating mental health care into the patient-centered medical home. The early work within the Pediatric Quality Measures Program is stimulating the refinement of existing child mental health measures and generating new proposed measures. The NQF is also embarking on re-evaluating existing and proposed behavioral health measures. At every phase, these processes are being conducted in collaboration with multi-stakeholder groups including parent and family representatives, providers, state agency representatives, and health services researchers who bring a breadth of perspectives on what makes adherence to a quality measure “meaningful.” Across the Centers of Excellence, the work of thousands of participant stakeholders has yielded a number of measures with strong expert support that will be systematically considered for a pediatric quality measure set to be submitted to the Secretary of the U.S. Department of Health and Human Services by January 2013.
The development of new child mental health quality measures poses methodological challenges. The constraints of existing data infrastructure, at the state and provider levels, must be addressed to enhance the capacity to capture data that links measure adherence to improved care and meaningful outcomes. Generating these desired data however demands time; therefore, priority must also be placed on reducing provider and parent burden. Further, new research models that promote engagement of community clinicians may require adaptation to test the clinical validity of child mental health quality measures.64

To continue to strive to find a “common ground”, a new partnered research paradigm for quality measurement for children is needed to capitalize on the rich network of collaboration from CHIPRA, NQF and other related projects. Early dialogue and sustained communication channels for information exchange, funding that cuts across these facets, and continuing to search for common goals aimed at improving the outcomes of children can serve as a starting point to address this important challenge. The adoption of electronic health care records may also serve as an additional mechanism to further strengthen these collaborations through active engagement in their development and implementation. Together, these activities share the original vision of a quality-driven health care system for children which can only truly improve through a continuous process of quality improvement conducted in full partnership.
References

1. PL 111-113 Children’s Health Insurance Program Reauthorization Act.


63. Agency for Healthcare Research and Quality (AHRQ). CHIPRA Core Set Candidate Measures: Follow-up after Hospitalization for Mental Illness.

65. Agency for Healthcare Research and Quality. CAHPS Health Plan Survey 4.0: Medicaid Version.

Table 1: Assessment Criteria for Proposed Quality Measures by National Initiative

<table>
<thead>
<tr>
<th>CHIPRA Initial Core Set&lt;sup&gt;26, 65&lt;/sup&gt;</th>
<th>NQF Patient Outcomes (Phase III) and Child Health Outcome Measures&lt;sup&gt;38&lt;/sup&gt;</th>
</tr>
</thead>
</table>

Table 1: Assessment Criteria for Proposed Quality Measures by National Initiative
Importance (descending order)

- The measure should be actionable. State Medicaid and CHIP programs, managed care plan, 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 (i.e., 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 (i.e., a “sentinel measure” of quality of care provided for preventive care, mental health care, or 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 balance portfolio of measures and is consistent with the intent of the legislation.

Importance to measure and report: the extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance.

- High impact

- Opportunities for improvement

- Outcome or evidence to support measure focus
- Improving performance on measures included in the core set should have the potential to transform care for our nation’s children.

**Scientific Acceptability**

Validity: the degree to which a quality measure is associated with what it purports to measure.

- It meets criteria for scientific soundness, defined as adequate scientific evidence or, where evidence is insufficient, expert professional consensus to support the relationship between structure and process, structure and outcome, or process and outcome.

- The measure itself is valid—that is, it should truly assess what it purports to measure.

**Scientific acceptability of measure properties**

- Precisely specified
- Reliability testing
- Validity testing
- Exclusions justified
- Risk adjustment for outcomes/resource use measures
- Identification of meaningful difference in performance
- Comparability of multiple data sources/methods
- Disparities in care

**Feasibility**

Feasibility: the degree to which the measure is free from random error

- The data necessary to score the measure are available to state Medicaid and CHIP programs.

- Detailed specifications are available for the measure

- Estimates of adherence to the measure based on available data sources are likely to be reliable and unbiased. This allows for meaningful comparisons across states, programs, individual

Feasibility: the extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement.
providers or institutional providers.

Usability$^d$

Usability: the extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.

- Meaningful, understandable and useful information
- Relation to other NQF-endorsed measures (harmonization, distinctive or additive value)
- Data generated as a byproduct of care processes
- Electronic sources
- Exclusions
- Susceptibility to inaccuracies, errors, or unintended consequences
- Data collection strategy/implementation

$^a$ CHIPRA Rating: 7-9= definitely important and meets several of the criteria, 4-2= uncertain level of importance and meets some of the criteria but fails to meet some of the criteria given higher weight (1-4), 1-3=fails to meet most of the criteria; CHIPRA Median Pass Score: ≥4; NQF Rating: Yes/No (must pass).

$^b$ CHIPRA Rating: 7-9=scientifically sound and the measure itself is definitely valid (i.e., sufficient evidence), 4-2=uncertain scientific soundness (i.e., insufficient evidence) and the measure itself has uncertain validity, 1-3= not scientifically sound and the measure itself is not valid; CHIPRA Median Pass Score: ≥7; NQF Rating: Completely, partially, minimally, not at all.

$^c$ CHIPRA Rating: 7-9=definitely feasible, 4-2=uncertain feasibility, 1-3=not feasible; CHIPRA Median Pass Score: ≥4; NQF Rating: Completely, partially, minimally, not at all.

$^d$ NQF Rating: Completely, partially, minimally, not at all.
<table>
<thead>
<tr>
<th>Measure</th>
<th>Evidence Grade</th>
<th>Steward</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Follow-up care for children prescribed ADHD medication</td>
<td>D</td>
<td>National Committee for Quality Assurance</td>
<td>Percentage of children 6 – 12 years of age as of the Index Prescription Episode Start Date with an ambulatory prescription dispensed for and ADHD medication and who had one follow-up visit with a practitioner with prescribing authority during the 30-Day Initiation Phase.</td>
</tr>
<tr>
<td>Management of attention deficit hyperactivity disorder in primary care for school age children and adolescents</td>
<td>NR</td>
<td>Institute for Clinical Systems Improvement</td>
<td>Percentage of patients treated with psycho-stimulant medication for the diagnosis of attention deficit hyperactivity disorder (ADHD) whose medical record contains documentation of a follow-up visit at least twice a year.</td>
</tr>
<tr>
<td>Follow-up after hospitalization for mental illness</td>
<td>B</td>
<td>National Committee for Quality Assurance</td>
<td>Percentage of discharges for members 6 years of age and older who were hospitalized for treatment of selected mental health disorders and who had an outpatient visit, an intensive outpatient encounter or partial hospitalization with a mental health practitioner. Two rates are reported. Rate 1. The percentage of members who received follow-up within 30 days of discharge. Rate 2. The percentage of members who received follow-up within 7 days of discharge.</td>
</tr>
<tr>
<td>Measure</td>
<td>Agency</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>---------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Developmental screening in the first three years of life</td>
<td>National Committee for Quality Assurance, The Children and Adolescent Health Measurement Initiative</td>
<td>Percentage of children screened for risk of developmental, behavioral and social delays using a standardized screening tool in the first three years of life. This is a measure of screening in the first three years of life that includes three, age-specific indicators assessing whether children are screened by 12 months of age, by 24 months of age and by 36 months of age.</td>
<td></td>
</tr>
<tr>
<td>Developmental screening by 2 years of age</td>
<td>National Committee for Quality Assurance</td>
<td>Percentage of children who turned 2 years old during the measurement year who had a developmental screening</td>
<td></td>
</tr>
<tr>
<td>Pediatric Symptom Checklist</td>
<td>Massachusetts General Hospital</td>
<td>The Pediatric Symptom Checklist is a brief parent report questionnaire that is used to measure overall psychosocial functioning in children from 4 to 16 years of age.</td>
<td></td>
</tr>
<tr>
<td>Depression screening by 13 years of age</td>
<td>National Committee for Quality Assurance</td>
<td>Percentage of adolescents who turn 13 years of age in the measurement year who had a screening for depression using a standardized tool.</td>
<td></td>
</tr>
<tr>
<td>Depression screening by 18 years of age</td>
<td>National Committee for Quality Assurance</td>
<td>Percentage of adolescents who turn 18 years of age in the measurement year who had a screening for depression using a standardized tool.</td>
<td></td>
</tr>
<tr>
<td>Risky behavior assessment by age 13 years</td>
<td>National Committee for Quality Assurance</td>
<td>Percentage of children with documentation of a risk assessment or counseling for risky behaviors by the age of 13 Years. Four rates are reported: Risk Assessment or Counseling for Alcohol Use, Risk Assessment or Counseling for Drug Use, Risk Assessment or Counseling for Tobacco Use, Risk Assessment or Counseling for Other Use.</td>
<td></td>
</tr>
<tr>
<td>Measure</td>
<td>Agency</td>
<td>Percentage Description</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------------------------------</td>
<td>---------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Risky behavior assessment by age 18 years (NQF-1515)</td>
<td>NR National Committee for Quality Assurance</td>
<td>Percentage of children with documentation of assessment or counseling for risky behavior. Four rates are reported: assessment or counseling for alcohol use, tobacco use, other substance use, and sexual activity.</td>
<td></td>
</tr>
<tr>
<td>Suicide risk assessment (NQF-1365)</td>
<td>NR American Medical Association</td>
<td>Percentage of patient visits for those patients aged 6 through 17 years with a diagnosis of major depressive disorder with an assessment for suicide risk.</td>
<td></td>
</tr>
<tr>
<td>Documentation of DSM-IV diagnostic evaluation for depression (NQF-1364)</td>
<td>NR American Medical Association</td>
<td>Percentage of patients aged 6 through 17 years with a diagnosis of major depressive disorder with documented evidence that they met the DSM-IV criteria [at least 5 elements with symptom duration of two weeks or longer, including 1) depressed mood (can be irritable mood in children and adolescents) or 2) loss of interest or pleasure] during the visit in which the new diagnosis or recurrent episode was identified.</td>
<td></td>
</tr>
<tr>
<td>Diagnosis of attention deficit hyperactivity disorder in primary care for school age children and adolescents</td>
<td>NR Institute for Clinical Systems Improvement</td>
<td>Percentage of patients newly diagnosed with attention deficit hyperactivity disorder whose medical record contains documentation of</td>
<td></td>
</tr>
</tbody>
</table>
(NQF-106) Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) or Diagnostic and Statistical Manual for Primary Care (DSM-PC) criteria being addressed.


b Evidence grades reported in this column are quoted from the official measure materials which can be found on the CHIPRA and NQF websites. More specifically, evidence grades for the CHIPRA measures are shown for each of the three measures in the summary table for all of the measures (http://www.ahrq.gov/chipra/corebackground/corebacktab.htm). For the measures endorsed by NQF, the evidence grade or lack thereof can be found on the Measure Submission and Evaluation Worksheet 5.0 that is posted for each measure on the NQF website (http://www.qualityforum.org/Home.aspx).

c From CHIPRA Website: The types and rigor of studies at various levels of evidence depend on the study purposes (e.g., therapy/prevention, prognosis, diagnosis, differential diagnosis/symptom prevalence; economic and decision analyses). Most of the studies submitted or identified as documentation of underlying scientific soundness for the measures were therapy or prevention studies. For those studies, Level 1 studies are systematic reviews of randomized controlled trials (RCTs). Level 2 studies include systematic reviews of cohort studies, individual cohort studies, including low-quality RCTs), and "outcomes" research. Level 3 studies are systematic reviews with homogeneity of case-control studies or an individual case-control study. Level 4 studies are case-series and poor quality cohort and case-control studies. Level 5 evidence is defined as expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles."

NR = Not rated as to grade of evidence