

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

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: Methods used included description of the selection process of child health care quality measures from 2 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 9 unique measures were related to child mental health.

CONCLUSIONS: The development of new child mental health quality measures poses methodologic challenges that will require a paradigm shift to align research with its accelerated pace. *Pediatrics* 2013;131: S38–S49

AUTHORS: Bonnie T. Zima, MD, MPH,^a J. Michael Murphy, EdD,^b Sarah Hudson Scholle, MPH, DrPH,^c Kimberly Eaton Hoagwood, PhD,^d Ramesh C. Sachdeva, MD, PhD,^e Rita Mangione-Smith, MD, MPH,^f Donna Woods, EdM, PhD,^g Hayley S. Kamin, BA,^h and Michael Jellinek, MDⁱ

^aUCLA Center for Health Services & Society, UCLA Semel Institute for Neuroscience and Human Behavior, University of California at Los Angeles.; ^bMassachusetts General Hospital, Harvard Medical School, Boston, Massachusetts; ^cNational Committee for Quality Assurance, Washington, DC; ^dDepartment of Child Psychiatry, New York University Medical Center, New York, New York; ^eMedical College of Wisconsin, Children's Hospital and Health System, Milwaukee, Wisconsin; ^fUniversity of Washington, Seattle Children's Research Institute, Center for Child Health, Behavior, and Development, Seattle, Washington; ^gInstitute for Healthcare Studies, Feinberg School of Medicine, Northwestern University, Chicago, Illinois; ^hDepartment of Psychology, University of Florida, Gainesville, Florida; and ⁱPartners Healthcare, Harvard Medical School, Boston, Massachusetts

KEY WORDS

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

ABBREVIATIONS

AAP—American Academy of Pediatrics
ADHD—attention-deficit/hyperactivity disorder
AHRQ—Agency for Healthcare Research and Quality
CEBM—Centre for Evidence-Based Medicine
CHIP—Children's Health Insurance Program
CHIPRA—Children's Health Insurance Program Reauthorization Act of 2009
CMS—Centers for Medicare & Medicaid Services
COE4CCN—Center of Excellence on Quality of Care Measures for Children with Complex Needs
DSM-IV—*Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*
MDD—major depressive disorder
NCINQ—National Collaborative for Innovation in Quality Measurement
NQF—National Quality Forum
NQS—National Quality Strategy
PI—Principal Investigator
PMCoE—Pediatric Measurement Center of Excellence
RCT—randomized controlled trial
SNAC—Subcommittee on Children's Healthcare Quality Measures for Medicaid and Children's Health Insurance Programs

(Continued on last page)

Recent health policies have accelerated the development and use of quality measures for children receiving publicly funded care.^{1,2} In response, a legislatively mandated national committee and a nonprofit organization systematically rated large pools of quality measures and recommended a limited number to monitor the quality of care received by US 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.^{3,4} This gap is of public health significance because improving the quality of child mental health care is a longstanding national priority,^{5–9} and there is substantial room for improvement in mental health care for both private and publicly insured populations.^{10–18}

This article reviews the following: 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 (CHIP).¹⁹ Developed under the auspices of the Agency for Healthcare Research and Quality (AHRQ), an initial core set of 24 quality measures was submitted to the Secretary of the US Department of Health and Human Services on January 1, 2010. For the subsequent Pediatric Quality Measures Program, \$55 million

was made available to support 7 Centers of Excellence in 2010 to develop new measures and refine existing ones for potential core set enhancements in January 2013, 2014, and 2015.²⁰

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 in February 2010; 7 of them propose to develop, test, evaluate, and/or report adherence to quality measures.²¹ 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 2011.²² The use of the measures is likely to be sustained through financial incentives to collect and report on adherence rates regarding quality indicators through a matching Federal Medical Assistance Percentage that is part of the American Recovery and Reinvestment Act of 2009.²³ 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.²³ 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.²²

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.^{2,24,25} 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 Services quality improvement programs and regulations, and set criteria to measure the

quality of health care to align with national efforts for quality improvement.²⁵ The 3 aims of the NQS are to improve the overall quality of care, improve the health of the US population, and reduce the cost of quality health care.²⁴ To adapt the NQS for behavioral health care, the Substance Abuse and Mental Health Services Administration developed the Behavioral Health Quality Framework that tailors the 6 national priority areas to behavioral health care, reinforcing how the 3 aims of the NQS could be equally applied to the care of mental health problems.⁵

Contemporaneously, the National Quality Forum (NQF) is a private, nonprofit 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 health care. The NQF is dedicated to improving the quality of US 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.²⁶ 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, undertaken between 2009 and 2011, are known as the Patient Outcomes (Phase III): Child Health and Child Health Measures Projects. Although 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 methodologic 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 by using an evidence-informed process that integrated input from a broad array of stakeholders and public comments.²⁷ 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. It comprised 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.²⁸

Over a 4-month period, the SNAC held 2 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 by 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 by using the same modified Delphi method. The Rand/UCLA appropriateness method is a well-established approach that integrates scientific evidence with expert clinical judgment²⁹; it has been successfully used to assess the quality of outpatient general health care among children nationally.³⁰ It has also been used to assess the

quality of mental health care statewide among children receiving publicly funded outpatient specialty mental health care.¹⁸ The process integrates a review of the evidence base for a proposed measure and 2 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, 2 measures were dropped due to lack of field testing, including 1 pertaining to suicide risk assessment for children with major depression. Details of the methods and administrative review pathways before final submission of the initial core set of measures are described elsewhere.^{27,28,31}

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 according to race/ethnicity, tribe, socioeconomic state, or special

health care need status, characteristics called for in the CHIPRA legislation.^{32,33} Content gaps led to recommendations for new measures for substance abuse care and mental health treatment as well as in several areas relevant to child mental health: specialty care, inpatient care, availability of services, coordination of care, medical home, family experiences of care, and outcomes.^{27,31,34,35} Furthermore, the SNAC strongly encouraged new quality measures to be aligned with the priorities of state Medicaid and CHIP agencies,^{36,37} providers, and parents.^{38,39}

NQF: Endorsement of Child Health Quality Measures

The NQF consensus development process involves 9 main steps that typically occur over a 12- to 18-month period. The steps are as follows: (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 Decision; (8) board ratification; and (9) 30-day appeals.⁴⁰ The review of the candidate standards for the aforementioned child health-related projects was conducted by steering committees composed of child health and family advocates, health care system and provider professional organizations, clinicians, and health care quality measurement experts. After 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 Web site, and time for public comment was allocated on the agenda.

The measures were rated on 4 main criteria: (1) importance to measure and report the nominal topic; (2) scientific acceptability; (3) usability; and (4) feasibility. Within these 4 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 4-point scale (ie, completely, partially, minimally, 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.⁴¹

RECOMMENDED OR ENDORSED QUALITY MEASURES FOR CHILD MENTAL HEALTH CARE

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; 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 3 CHIPRA measures, 2 were the same measure for 2 different age groups of teenagers, and 1 measured maternal mental health. Thus, there were 9 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 1 was restricted to children aged 0 to 5 years, 2 to ages 13 to 18 years, and 6 included all or most child age groups. The focus of concern also ranged from specific to general problem areas. Two measures focused on depression, 2 on attention-deficit/hyperactivity disorder (ADHD), 1 on risky behaviors, 1 on suicidality, and 3 on general problem areas. Two of the measures involved monitoring, 3 called for screening, and 4 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)⁴² has put together detailed methods for conducting these kinds of ratings, and all of the CHIPRA measures were reviewed according to the CEBM standards.²⁷ 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 that have been conducted 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 either systematic reviews of cohort studies or 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 earlier, the quality of the evidence for the 3 CHIPRA measures had been graded according to CEBM standards. Although 1 of the CHIPRA measures received the low grade of D, 2 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 use CEBM standards, there was a section on evidence and all relevant studies on the NQF Web site for each measure.²⁶ For the purposes of the current article, we reviewed the studies cited there and supplemented this with a review of studies on the Web site of the steward listed for each measure. We also conducted a search by using Ovid and PubMed of studies published from 2001 to 2011 with the 6 measure names as specific and general search terms.

For only 2 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 and 1515) noted that this measure had been rated by the US Preventive Services Task Force as having a B level of evidence, citing studies^{43,44} which 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.^{45–48}

TABLE 1 Assessment Criteria for Proposed Quality Measures by National Initiative

CHIPRA Initial Core Set ^{27,69}	NQF Patient Outcomes (Phase III) and Child Health Outcome Measures ⁴⁰
<p>Importance^a Importance (descending order)</p> <ul style="list-style-type: none"> • 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, 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, a “sentinel measure” of quality of care provided for preventive care, mental health care, or dental care) • The measure should assess an aspect of health care in which 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 • Improving performance on measures included in the core set should have the potential to transform care for our nation’s children 	<p>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 health care where there is variation in or overall poor performance.</p> <ul style="list-style-type: none"> • High impact • Opportunities for improvement • Outcome or evidence to support measure focus
<p>Scientific acceptability^b Validity: the degree to which a quality measure is associated with what it purports to measure</p> <ul style="list-style-type: none"> • It meets criteria for scientific soundness, defined as adequate scientific evidence or, where evidence is insufficient, expert professional consensus to support the relation 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 	<p>Scientific acceptability of measure properties</p> <ul style="list-style-type: none"> • 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
<p>Feasibility^c Feasibility: the degree to which the measure is free from random error</p> <ul style="list-style-type: none"> • 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 providers or institutional providers 	<p>Feasibility: the extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement</p>
<p>Usability^d</p>	<p>Usability: the extent to which intended audiences (eg, consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision-making</p> <ul style="list-style-type: none"> • 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 (ie, sufficient evidence); 4–2 = uncertain scientific soundness (ie, 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 2 Recommended and/or Endorsed Child Mental Health Quality Measures

Measure (Source, Number)	Evidence Grade ^{a,b,c}	Steward	Description
1a. Follow-up care for children prescribed ADHD medication (CHIPRA-20; NQF-108)	D	National Committee for Quality Assurance	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 1 follow-up visit with a practitioner with prescribing authority during the 30-day initiation phase
1b. Management of ADHD in primary care for school-aged children and adolescents (NQF-107)	NR	Institute for Clinical Systems Improvement	Percentage of patients treated with psychostimulant medication for the diagnosis of ADHD whose medical record contains documentation of a follow-up visit at least twice a year
2. Follow-up after hospitalization for mental illness (CHIPRA-21; NQF-576)	B	National Committee for Quality Assurance	Percentage of discharges for members aged ≥ 6 years 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 was the percentage of members who received follow-up within 30 days of discharge. Rate 2 was the percentage of members who received follow-up within 7 days of discharge
3a. Developmental screening in the first 3 years of life (CHIPRA-8; NQF-1448)	B	National Committee for Quality Assurance, The Children and Adolescent Health Measurement Initiative	Percentage of children screened for risk of developmental, behavioral, and social delays by using a standardized screening tool in the first 3 years of life. This is a measure of screening in the first 3 years of life that includes 3 age-specific indicators assessing whether children are screened by 12 months, 24 months, or 36 months of age
3b. Developmental screening by 2 years of age (NQF-1399)	NR	National Committee for Quality Assurance	Percentage of children who turned 2 years old during the measurement year who had a developmental screening performed between 12 and 24 months of age
4. Pediatric Symptom Checklist (NQF-722)	NR	Massachusetts General Hospital	The Pediatric Symptom Checklist is a brief parent-report questionnaire that is used to measure overall psychosocial functioning in children aged 4 to 16 years
5a. Depression screening by 13 years of age (NQF-1394)	NR	National Committee for Quality Assurance	Percentage of adolescents who turn 13 years of age in the measurement year who had a screening for depression by using a standardized tool
5b. Depression screening by 18 years of age (NQF-1515)	NR	National Committee for Quality Assurance	Percentage of adolescents who turn 18 years of age in the measurement year who had a screening for depression by using a standardized tool.
6a. Risky behavior assessment by age 13 years (NQF-1406)	NR	National Committee for Quality Assurance	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 tobacco use, risk assessment or counseling for other substance abuse, and risk assessment or counseling for sexual activity
6b. Risky behavior assessment by age 18 years (NQF-1515)	NR	National Committee for Quality Assurance	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
7. Suicide risk assessment (NQF-1365)	NR	American Medical Association	Percentage of patient visits for those patients aged 6 through 17 years with a diagnosis of MDD with an assessment for suicide risk

TABLE 2 Continued

Measure (Source, Number)	Evidence Grade ^{a,b,c}	Steward	Description
8. Documentation of DSM-IV diagnostic evaluation for depression (NQF-1364)	NR	American Medical Association	Percentage of patients aged 6 through 17 years with a diagnosis of MDD with documented evidence that they met the DSM-IV criteria (at least 5 elements with symptom duration of ≥2 weeks, 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
9. Diagnosis of ADHD in primary care for school-aged children and adolescents (NQF-106)	NR	Institute for Clinical Systems Improvement	Percentage of patients newly diagnosed with ADHD whose medical record contains documentation of DSM-IV or <i>Diagnostic and Statistical Manual for Primary Care</i> criteria being addressed

NR, not rated as to grade of evidence.

^a Oxford Centre for Evidence-based Medicine Levels of Evidence (March 2009). Produced by Bob Phillips, Chris Ball, Dave Sackett, Doug Badenoch, Sharon Straus, Brian Haynes, Martin Dawes since November 1998. Updated by Jeremy Howick March 2009. Available at www.cebm.net. Accessed January 24, 2013.

^b Evidence grades reported in this column are quoted from the official measure materials, which can be found on the CHIPRA and NQF Web sites. More specifically, evidence grades for the CHIPRA measures are shown for each of the 3 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 Web site (<http://www.qualityforum.org/Home.aspx>).

^c From the CHIPRA Web site: the types and rigor of studies at various levels of evidence depend on the study purposes (eg, 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 RCTs. Level 2 studies include systematic reviews of cohort studies or 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.”

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

IMPROVING QUALITY MEASURES FOR CHILD MENTAL HEALTH CARE: NEXT STEPS

As part of the Pediatric Quality Measures Program, 3 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 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 (Principal Investigator [PI]: Dr Sachdeva), the AHRQ-CMS CHIPRA National Collaborative for Innovation in Quality Measurement (NCINQ) based at the National Committee on Quality Assurance (PI: Dr 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: Dr 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 (PI: Dr 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. The following discussion offers brief updates of the centers’ early activities.

Attention-Deficit/Hyperactivity Disorder

The PMCoE is working collaboratively with the American Medical Association Physician Consortium for Performance Improvement, the American Academy of Pediatrics (AAP), the American Board of Pediatrics, and the research academic centers Northwestern University and the Medical College of Wisconsin on the development and refinement of quality measures related to the care of ADHD. This disorder was selected because it is

prevalent, affecting an estimated 3% to 9% of US children.⁵² It is 1 of the most common reasons children are referred for mental health services and represents 15% to 45% of the mental health conditions diagnosed in children and youth.^{53,54,55} Considerable variations and gaps in care regarding ADHD have been documented in the literature.^{11,12,17} 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 regarding effective ADHD diagnosis, follow-up, and treatment. To incorporate the current best evidence about these topics, the AAP conducted a 2-year process to revise and update the 2003 AAP ADHD guideline. The most recent ADHD guideline was published in November 2011, making several changes to the previous guideline recommendations to direct the field toward care based on the best existing evidence through 6 primary recommendations.⁵⁶ Based on these new AAP ADHD guideline recommendations, investigators for Northwestern University, along with investigators and staff from the American Medical Association, the AAP, and the American Board of Pediatrics, established and engaged an expert workgroup comprising experts across the broad spectrum of stakeholders related to the diagnosis, follow-up, and treatment of ADHD. This workgroup included 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

Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (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 years, 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 of pediatric patients, ages 4 to 18 years, after a diagnosis of ADHD has been made. After 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 by 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.⁴⁴ MDD affects >7% of adolescents in the United States. In 2006, ~2.3 million adolescents 12 to 17 years of age reported experiencing a major depressive episode at some point in their lives.⁴⁴ 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.^{57–62} Depression may also lead to engagement in risky behaviors such as substance

use (eg, alcohol, illicit drugs, tobacco), and it can also lead to suicide.^{58–61} Suicide, the third leading cause of death among 15- to 24-year olds, is often preceded by depression or long-term MDD.^{44,60} Adolescent-onset depression increases the risk of attempted suicide by fivefold⁴⁴ and is strongly correlated with chronic and recurring depression in adulthood.⁶³ Furthermore, depressive symptoms can be both prolonged and episodic, recurring over weeks and months.⁵⁷ The Centers for Disease Control and Prevention noted that individuals who experience just 1 episode of depression are at a 50% higher risk of experiencing additional episodes.⁶⁴

Based on a review of all major guidelines, evidence reviews, and advice from family partners, clinicians, and researchers, the National Committee on Quality Assurance 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.⁶⁵ 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 by 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 by using existing, large data sets such as Medicaid claims from entire states. These analyses are being conducted by using data from 1 state Medicaid agency as well as a large urban tertiary care children's hospital. Through this approach, the COE4CCN is working to build the capacity of using 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 pediatrics⁶⁶ suggests that the prevalence of child mental health problems may be underestimated. Delivery of mental health care may also be underreported because procedure codes for evidence-based mental health care are often missing in Medicaid claims data.^{17,18} 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" by using Medicaid claims and enrollment data is under development. For the purposes of this project, social complexity is defined as the presence of ≥ 1 social risk factor

hypothesized to be a strong correlate of mental health. Valid identification of social complexity may enhance the identification of mental health problems that might be underreported 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 1 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.⁶⁷ 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 multiple stakeholder groups, including parent and family representatives, providers, state agency representatives, and health services researchers. They all bring a breadth of perspectives on what makes adherence to a quality measure "meaningful."

The development of new child mental health quality measures poses methodologic challenges. The constraints of existing data infrastructure, at the state and provider levels, must be addressed to enhance the capacity to capture data that link measure adherence to improved care and meaningful outcomes. Generating these desired data demands time; therefore, priority must also be placed on reducing provider and parent

burden. Furthermore, new research models that promote engagement of community clinicians may require adaptation to test the clinical validity of child mental health quality measures.⁶⁸

A paradigm shift for quality measurement for children is needed to align research with its accelerated pace and 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 sharing the common goal of improving outcomes for children can serve as a starting point. The adoption of electronic health care records may also serve as a 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 that can be attained through a continuous process of quality improvement conducted in full partnership.

ACKNOWLEDGMENTS

The authors thank Evan M. Williamson, MPH, MS, for his technical consultation on the evidence bases for the NQF behavioral health measures. They also 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 AAP 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.

- CHIPRA and beyond. *Acad Pediatr*. 2011;11 (suppl 3):S77–S84
34. Co JP, Sternberg SB, Homer CJ. Measuring patient and family experiences of health care for children. *Acad Pediatr*. 2011;11 (suppl 3):S59–S67
 35. Sternberg SB, Co JP, Homer CJ. Review of quality measures of the most integrated health care settings for children and the need for improved measures: recommendations for initial core measurement set for CHIPRA. *Acad Pediatr*. 2011;11(suppl 3): S49–S58.e3
 36. deLeon SE, Hess CA. Medicaid and CHIP children's healthcare quality measures: what states use and what they want. *Acad Pediatr*. 2011;11(suppl 3):S68–S76
 37. Greene-McIntyre M, Caldwell C. State Medicaid and Children's Health Insurance Program's perspective on DHIPRA Core Measures. *Acad Pediatr*. 2011;11(suppl 3): S89–S90
 38. Palfrey JS, Brei TJ. Children's health care providers and health care quality measurement. *Acad Pediatr*. 2011;11(suppl 3): S87–S88
 39. Wells N, Partridge L. Families are key in improving quality. *Acad Pediatr*. 2011;11 (suppl 3):S85–S86
 40. National Quality Forum. Consensus Development Process. Available at: www.qualityforum.org/Measuring_Performance/Consensus_Development_Process.aspx. Accessed March 7, 2012
 41. National Quality Forum. Behavioral health phase 1 and 2. Available at: www.qualityforum.org/Projects/Behavioral_Health_Phase_1_and_2.aspx. Accessed March 7, 2012
 42. Oxford Centre for Evidence-based Medicine. Levels of evidence (March 2009). Available at: www.cebm.net/index.aspx?o=1025. Accessed February 24, 2012
 43. Weissman MM, Wolk S, Goldstein RB, et al. Depressed adolescents grown up. *JAMA*. 1999;281(18):1707–1713
 44. Williams SB, O'Connor E, Eder M, Whitlock E. Screening for child and adolescent depression in primary care settings: A systematic evidence review for the US Preventive Services Task Force. *Pediatrics*. 2009;123(4). Available at: www.pediatrics.org/cgi/content/full/123/4/e716
 45. Kuhlthau K, Jellinek M, White G, Vancleave J, Simons J, Murphy M. Increases in behavioral health screening in pediatric care for Massachusetts Medicaid patients. *Arch Pediatr Adolesc Med*. 2011;165(7):660–664
 46. Gall G, Pagano ME, Desmond MS, Perrin JM, Murphy JM. Utility of psychosocial screening at a school-based health center. *J Sch Health*. 2000;70(7):292–298
 47. Hacker KA, Williams S, Myagmarjv E, Cabral H, Murphy M. Persistence and change in pediatric symptom checklist scores over 10-18 months. *Acad Pediatr*. 2009;9(1):270–277
 48. Murphy JM, Arnett HL, Bishop SJ, Jellinek MS, Reede JY. Screening for psychosocial dysfunction in pediatric practice. A naturalistic study of the Pediatric Symptom Checklist. *Clin Pediatr (Phila)*. 1992;31(11): 660–667
 49. Houry MJ, Gwinn M, Yoon PW, Dowling N, Moore CA, Bradley L. The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? *Genet Med*. 2007;9(10):665–674
 50. Trochim W, Kane C, Graham MJ, Pincus HA. Evaluating translational research: a process marker model. *Clin Transl Sci*. 2011;4 (3):153–162
 51. Pincus HA, Spaeth-Rublee B, Watkins KE. Analysis & commentary: The case for measuring quality in mental health and substance abuse care. *Health Aff (Millwood)*. 2011;30(4):730–736
 52. American Academy of Pediatrics. ADHD: Clinical practice guideline for the diagnosis, evaluation, and treatment of Attention-Deficit/Hyperactivity Disorder in children and adolescents. *Pediatrics*. 2011;128(5): 2011–2654
 53. American Academy of Pediatrics. ADHD: Clinical practice guideline for the diagnosis, evaluation, and treatment of Attention-Deficit/Hyperactivity Disorder in children and adolescents. *Pediatrics*. 2011;128(5): 2011–2654
 54. National Center for Health Statistics. Summary Health Statistics for U.S. Children: National Health Interview Survey 2011 *Vital Health Stat* 10(254). 2011. Medicaid and CHIP Payment and Access commission (MACPAC). Report to the Congress on Medicaid and CHIP, March 2012
 55. Visser MS, Bitsko RH, Danielson ML, et al. Increasing Prevalence of Parent-Reported Attention Deficit/Hyperactivity Disorder Among Children; United States, 2003 and 2007 November 12, 2010 / 59(44);1439–1443 *SN*
 56. American Academy of Pediatrics. ADHD: Clinical practice guideline for the diagnosis, evaluation, and treatment of attention-deficit/hyperactivity disorder in children and adolescents. *Pediatrics*. 2010;128(5):1007–1022
 57. Child Mind Institute. Major depressive disorder. Available at: www.childmind.org/en/health/disorder-guide/major-depressive-disorder. Accessed September 20, 2011
 58. Foley HA, Carlton CO, Howell RJ. The relationship of attention deficit hyperactivity disorder and conduct disorders to juvenile delinquency: legal implications. *Bull Am Psychiatry Law*. 1996;24(3):333–345
 59. Friedman RM, Katz-Levey JW, Manderschied RW, Sondheimer DL. Prevalence of serious emotional disturbance in children and adolescents. In: Manderscheid RW, Sonnenschein MA, eds. *Mental Health*. Rockville, MD: Center for Mental Health Services; 1996: 71–78
 60. National Research Council and Institute of Medicine, Committee on Adolescent Health Care Services and Models of Care for Treatment Prevention, and Healthy Development, Board on Children, Youth, and Families, Division of Behavioral and Social Sciences and Education. *Adolescent Health Services: Missing Opportunities*. Washington, DC: The National Academies Press; 2009
 61. Taylor E, Chadwick O, Heptinstall E, Danckaerts M. Hyperactivity and conduct problems as risk factors for adolescent development. *J Am Acad Child Adolesc Psychiatry*. 1996;35(9):1213–1226
 62. US Preventive Services Task Force. Screening and treatment for major depressive disorder in children and adolescents: US Preventive Services Task Force recommendation statement. *Pediatrics*. 2009;123(6): 1611
 63. Garber J, Clarke GN, Weersing VR, et al. Prevention of depression in at-risk adolescents: a randomized controlled trial. *JAMA*. 2009;301(21):2215–2224
 64. Centers for Disease Control and Prevention. Depression. Available at: www.cdc.gov/mentalhealth/basics/mental-illness/depression.htm. Accessed February 24, 2012
 65. Harding KJ, Rush AJ, Arbuckle M, Trivedi MH, Pincus HA. Measurement-based care in psychiatric practice: a policy framework for implementation. *J Clin Psychiatry*. 2011; 72(8):1136–1143
 66. Costello EJ, Edellbrock C, Costello AJ, Dulcan MK, Burns BJ, Brent D. Psychopathology in pediatric primary care: the new hidden

- morbidity. *Pediatrics*. 1988;82(3 pt 2):415–424
67. Agency for Healthcare Research and Quality (AHRQ). CHIPRA core set candidate measures: follow up after hospitalization for mental illness. Available at: www.ahrq.gov/chipra/corebackground/measures/CC-5.pdf. Accessed March 5, 2012
68. Ryan G, Berrebi C, Beckett M, et al. Reengineering the clinical research enterprise to involve more community clinicians. *Implementation Sci*. 2011;6(1):36
69. Agency for Healthcare Research and Quality. CAHPS health plan survey 4.0: Medicaid version. Available at: https://www.cahps.ahrq.gov/content/products/HP3/PROD_HP3_ExecSummary.asp?p=1021&s=211. Accessed June 1, 2011

(Continued from first page)

Dr Zima was a Robert Wood Johnson Foundation (RWJF) Clinical Scholar, University of California, Los Angeles, 1991; Dr Mangione-Smith was a RWJF Clinical Scholar, University of California, Los Angeles, 1997, and RWJF Generalist Physician Faculty Scholar, 2000–2004.

Dr Zima 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; Dr Murphy provided consultation regarding the article's conceptual framework, offered oversight on the literature review, and participated in writing early and final manuscript drafts; Dr Scholle provided consultation on the development of new depression screening measures and participated in writing early and final manuscript drafts; Dr Hoagwood provided consultation on the conceptual framework and development of new depression screening measures, and participated in writing early and final manuscript drafts; Dr Sachdeva and Dr Woods provided consultation on the refinement of attention-deficit/hyperactivity disorder quality measures, and participated in writing early and final manuscript drafts; Dr Mangione-Smith 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; Ms Kamin conducted literature reviews, and participated in writing early and final manuscript drafts; and Dr Jellinek participated in development of the conceptual framework, and the writing of early and final manuscript drafts.

www.pediatrics.org/cgi/doi/10.1542/peds.2012-1427e

doi:10.1542/peds.2012-1427e

Accepted for publication Dec 20, 2012

Address correspondence to Bonnie T. Zima, MD, MPH, UCLA Center for Health Services and Society, 10920 Wilshire Blvd, # 300, Los Angeles, CA 90024. E-mail: bzima@mednet.ucla.edu

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

Copyright © 2013 by the American Academy of Pediatrics

FINANCIAL DISCLOSURE: *The authors have indicated they have no financial relationships relevant to this article to disclose.*

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

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

Bonnie T. Zima, J. Michael Murphy, Sarah Hudson Scholle, Kimberly Eaton Hoagwood, Ramesh C. Sachdeva, Rita Mangione-Smith, Donna Woods, Hayley S. Kamin and Michael Jellinek

Pediatrics 2013;131;S38

DOI: 10.1542/peds.2012-1427e

Updated Information & Services	including high resolution figures, can be found at: /content/131/Supplement_1/S38.full.html
References	This article cites 39 articles, 5 of which can be accessed free at: /content/131/Supplement_1/S38.full.html#ref-list-1
Citations	This article has been cited by 4 HighWire-hosted articles: /content/131/Supplement_1/S38.full.html#related-urls
Subspecialty Collections	This article, along with others on similar topics, appears in the following collection(s): Administration/Practice Management /cgi/collection/administration:practice_management_sub Quality Improvement /cgi/collection/quality_improvement_sub
Permissions & Licensing	Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at: /site/misc/Permissions.xhtml
Reprints	Information about ordering reprints can be found online: /site/misc/reprints.xhtml

PEDIATRICS is the official journal of the American Academy of Pediatrics. A monthly publication, it has been published continuously since 1948. PEDIATRICS is owned, published, and trademarked by the American Academy of Pediatrics, 141 Northwest Point Boulevard, Elk Grove Village, Illinois, 60007. Copyright © 2013 by the American Academy of Pediatrics. All rights reserved. Print ISSN: 0031-4005. Online ISSN: 1098-4275.

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™



PEDIATRICS®

OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

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

Bonnie T. Zima, J. Michael Murphy, Sarah Hudson Scholle, Kimberly Eaton Hoagwood, Ramesh C. Sachdeva, Rita Mangione-Smith, Donna Woods, Hayley S. Kamin and Michael Jellinek
Pediatrics 2013;131;S38
DOI: 10.1542/peds.2012-1427e

The online version of this article, along with updated information and services, is located on the World Wide Web at:
/content/131/Supplement_1/S38.full.html

PEDIATRICS is the official journal of the American Academy of Pediatrics. A monthly publication, it has been published continuously since 1948. PEDIATRICS is owned, published, and trademarked by the American Academy of Pediatrics, 141 Northwest Point Boulevard, Elk Grove Village, Illinois, 60007. Copyright © 2013 by the American Academy of Pediatrics. All rights reserved. Print ISSN: 0031-4005. Online ISSN: 1098-4275.

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™

