

- Am J Psychiatry*. 1998;155:36–42.
6. Pollack MH, Otto MW, Worthington JJ, et al. Sertraline in the treatment of panic disorder: a flexible-dose multicenter trial. *Arch Gen Psychiatry*. 1998;55:1010–1016.
  7. Tesar GE, Rosenbaum JF, Pollack MH, et al. Double-blind, placebo-controlled comparison of clonazepam and alprazolam for panic disorder. *J Clin Psychiatry*. 1991;52:69–76.
  8. DeMartini NA, Schweizer F. An open-label trial of nefazodone in high comorbidity panic disorder. *J Clin Psychiatry*. 1996;57:245–248.
  9. Regier DA, Narrow WE, Rae DS, et al. The de facto US mental and addictive disorders service system. Epidemiologic catchment area prospective 1-year prevalence rates of disorders and services. *Arch Gen Psychiatry*. 1993;50:85–94.
  10. Blanco C, Laje G, Olfson M, et al. Trends in the treatment of bipolar disorder by outpatient psychiatrists. *Am J Psychiatry*. 2002;159:1005–1010.
  11. Goodwin R, Gould MS, Blanco C. Prescription of psychotropic medications to youths in office-based practice. *Psychiatric Ser*. 2001;52:1081–1087.
  12. Olfson M, Marcus SC, Pincus HA, et al. Antidepressant prescribing practices of outpatient psychiatrists. *Arch General Psychiatry*. 1998;55:310–316.
  13. Pincus HA, Tanielian TL, Marcus SC, et al. Prescribing trends in psychotropic medications: primary care, psychiatry, and other medical specialties. *JAMA*. 1998;279:526–531.
  14. Potthoff RF, Woodbury MA, Manton KG. “Equivalence sample size” and “equivalent degrees of freedom” refinements for inference using survey weights under superpopulation models. *J Am Stat Assoc*. 1992;87:383–396.
  15. Robins LN, Regier DS, eds. *Psychiatric Disorders in America*. New York: Free Press; 1991:27–28.
  16. Rogers EM. *Diffusion of Innovations*. New York, Free Press, 1995.
  17. Bruce SE, Vasile RG, Goisman RM, et al. Are benzodiazepines still the medication of choice for patients with panic disorder with or without agoraphobia? *Am J Psychiatry*. 2003;160:1432–1438.
  18. Ballenger JC, Lydiard RB. Panic disorder: Results of a patient survey. *Human Psychopharmacol*. 1997;12:S27–S23.
  19. Marchesi C, Ampollin P, Signifredi R, et al. The treatment of panic disorder in a clinical setting: a 12-month naturalistic study. *Neuropsychobiology*. 1997;36:25–31.
  20. Roy-Byrne P, Russo J, Dugdale DC, et al. Undertreatment of panic disorder in primary care: role of patient and physician characteristics. *J Am Board Fam Pract*. 2002;15:443–450.
  21. Hazlett-Stevens H, Craske MG, Roy-Byrne PP, et al. Predictors of willingness to consider medication and psychosocial treatment for panic disorder in primary care patients. *Gen Hosp Psychiatry*. 2002;24:316–321.

## Achieving Consensus Across Diverse Stakeholders on Quality Measures for Mental Healthcare

Richard C. Hermann, MD, MS,\*†‡ Heather Palmer, MB, BCh, SHM,‡ Stephen Leff, PhD,§ Michael Schwartz, PhD,¶ Scott Provost, MM, MSW,\*† Jeffrey Chan, BS,\*† Wai T. Chiu, AM,† and Greta Lagodmos, BA†

**Objective:** Quality-improvement efforts are hindered by a lack of consensus on meaningful and feasible measures of care. The objective of this study was to develop a core set of quality measures for mental health and substance-related care that are meaningful to stakeholders, feasible to implement, and broadly representative of diverse dimensions of the mental health system.

**Method:** A 12-member panel of stakeholders from national organizations evaluated 116 process measures in a 2-stage modified Delphi

consensus development process. Drawing on a conceptual framework and literature review, panelists rated each measure on 7 domains using a 9-point scale (1 = best). Measures were then mapped to a framework of system dimensions to identify a core set with the highest ratings for system characteristics within each dimension.

**Results:** Twenty-eight measures were identified assessing treatment (12), access (2), assessment (2), continuity (4), coordination (2), prevention (1), and safety (5). Overall, mean ratings for meaningfulness were: clinical importance 2.29; perceived gap between actual and optimal care 2.59; association between improved performance and outcome 2.61. For feasibility, mean ratings were clarity of specifications 3.39; acceptability of data collection burden 4.77; and adequacy of case mix adjustment 4.20. The measures address a range of treatment modalities, clinical settings, diagnostic categories, vulnerable populations, and other dimensions of mental healthcare.

**Conclusions:** A structured consensus process identified a core set of quality measures that are meaningful and feasible to multiple stakeholders, as well as broadly representative of the mental healthcare system. By yielding quantitative assessments of meaningfulness, feasibility and degree of consensus among stakeholders, these results can inform ongoing national efforts to adopt common quality measures for mental healthcare.

**Key Words:** quality, measure, mental health, consensus development

From the \*Institute for Clinical Research and Health Policy Studies at Tufts-New England Medical Center; the †Center for Quality Assessment and Improvement in Mental Health; the ‡Harvard School of Public Health; §The Evaluation Center@Human Services Research Institute, and the ¶Boston University School of Management, Boston, Massachusetts.

Supported by grants from AHRQ (R01-HS10303), NIMH (K08-MH001477), and SAMHSA.

Reprints: Richard C. Hermann, MD, MS, Tufts-New England Medical Center, 750 Washington Street, NEMC #345, Boston, MA 02111. E-mail: rhermann@cqaimh.org.

Measurement-based quality assessment and quality improvement (QA/QI) activities are underway throughout the US healthcare system. However, there is little agreement among “stakeholders” in the mental healthcare system (eg, consumers, families, clinicians, payers, purchasers, and oversight organizations) on which measures should be used. Lack of agreement on common measures increases the burden of data collection on providers, decreases the comparability of measure results, and slows efforts to test and refine measures with most promise.

The implementation of standardized sets of “core measures” is widely advocated to ensure that systems and providers focus on clinically important processes with known variations in quality of care.<sup>1–4</sup> Measures typically are selected on the basis of criteria for meaningfulness and feasibility as well as to represent diverse dimensions of the system under evaluation.<sup>5</sup>

There are a number of initiatives aimed at developing consensus on quality measures for mental healthcare.<sup>6–8</sup> Progress has been made for a few settings and conditions.<sup>9–12</sup>

It has proven more difficult to achieve consensus among stakeholders on a broader set of measures that are representative of the diverse conditions, treatments, settings and populations that comprise the mental healthcare system. Among the barriers are, first, existing measures are numerous and vary widely in content and scientific properties. A national inventory identified more than 300 process measures proposed for QA/QI.<sup>13–16</sup> Less than 40% were based on research evidence. Few were assessed for validity or reliability. They draw on a wide range of data sources, each with implications for feasibility of use. Second, stakeholder groups participating in initiatives to select mental health measures have widely divergent priorities with regard to measure content, rigor and burden.<sup>17</sup>

This study addresses the question, how can diverse stakeholders with competing values select quality measures for common use? We developed a method that brought stakeholders from national organizations together to systematically review measures for mental healthcare and identify a core set of 25–30 that were meaningful, feasible, and representative of diverse dimensions of the mental health system. We focused on a specific, widely used type of measure—single-item, technical process measures—which examine the clinical content of care, usually in comparison with a standard or norm.

**Method**

Building on prior work in consensus development,<sup>18–20</sup> we developed a method to select core measures by identifying areas of common interest, balancing competing interests, and providing a floor based on measure quality. First, we developed a framework to identify measure attributes and system dimensions. Second, we screened existing measures to identify promising candidates. Third, we convened a panel of diverse stakeholders to rate candidate measures. Fourth, core measures were selected on the basis of panel ratings and system representativeness.

**Conceptual Framework**

Broad agreement exists on several principles for quality measurement.<sup>21</sup> Measures should be meaningful, ie, address known quality problems in clinically important areas, be evidence-based and valid, and yield actionable results. Measures should be feasible, ie, precisely specified, reliably collected, adequately case mix adjusted, and based on available and affordable data. Core measure sets should reflect the diversity of systems they are intended to assess. Important dimensions of mental healthcare include different psychiatric conditions, modalities, settings, clinical processes (ie, prevention, access, assessment, treatment, continuity, coordination, and safety), and needs that are specific to vulnerable groups. We developed a conceptual framework that incorporates these principles but also identifies tensions among them.<sup>17</sup>

For example, measures based on research evidence often require more costly sources of data and are not equally available across conditions, treatment modalities or groups.<sup>13,14</sup> In this way, the goal of meaningfulness conflicts with the goals of feasibility and representativeness. The framework identified areas where stakeholder priorities conflict and require tradeoffs to produce consensus.

**Measure Collection and Screening**

Drawing on a nationwide inventory of technical process measures developed for QA/QI in mental healthcare,<sup>13,16</sup> we screened available measures for a face relationship to quality, specifications for use and use of existing data sources (Fig. 1). This process resulted in 116 measures for review by the panel. Detailed information on the measures and their properties are available at [www.cqaimh.org](http://www.cqaimh.org).

**Stakeholder Panel Composition**

Measure ratings were obtained from a diverse panel of stakeholders using a 2-stage modified-Delphi method.<sup>22</sup> In consultation with national stakeholder organizations, we selected 12 panelists representing each stakeholder perspective (Table 1).

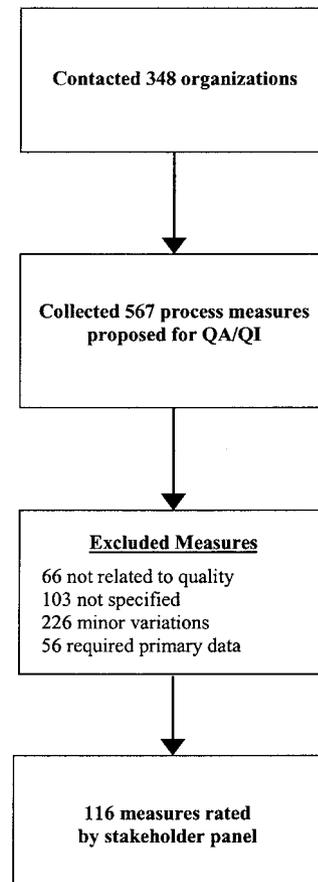


FIGURE 1. Collection and screening of quality measures.

**TABLE 1.** Characteristics of Members of Stakeholder Panel (n = 12)

	No. Panelists
Stakeholder perspective	
Accrediting organization	2
Federal payer/purchaser	1
State payer/purchaser	1
Private sector employer/purchaser	1
Clinician	1
Managed care organization	1
Delivery system manager	1
Researcher	1
Consumer/family advocacy	3
Clinical background (n = 9)	
Nurse	1
Psychiatrist	3
Psychologist	3
Social worker	1
Case manager	1
Geographic region	
Northeast	5
South	1
North central	4
West	2
Minority race/ethnicity (n = 2)	
Black	1
Hispanic	1
Gender	
Men	6
Women	6
Areas of specialized experience (n = 8)	
Children	3
Elderly	1
Substance abuse	3
Primary care	1

### Measure Rating Process

Panelists were provided with available information on 116 measures, including specifications, rationales, data requirements, research evidence, reliability and validity results and case mix adjustment. Profiles were developed by research staff based on training and guidelines and were verified by a study investigator. Attributes requiring judgment (eg, evidence) were subject to inter-rater reliability testing with kappa scores of 0.5 (moderate) to 1.0 (excellent).<sup>23</sup>

There were 2 rounds of ratings. First, panelists individually rated each measure on 3 scales assessing *meaningfulness* (importance of the process to the panelist's stakeholder group, perceived gap between actual and optimal care, and perceived relationship between improved performance on the

measure and patient outcomes) and 3 scales assessing *feasibility* (clarity of specifications, acceptability of data collection burden, and capacity to provide fair comparisons across healthcare organizations). A seventh scale assessed, *overall*, whether the measure should be included in the core set. Responses were obtained on a 9-point Likert scale: 1 to 3 indicating agreement, 4 to 6 neither agreement nor disagreement, and 7 to 9 disagreement. First-round ratings were provided to panelists as a basis for discussion, including the median, distribution and dispersion (mean absolute deviation from the median, indicating degree of agreement). Discussion focused on measures with overall scores of 6 or less and significant dispersion. If consensus for revision existed, the panel made changes for content or clarity. After discussion, panelists rerated each measure individually and confidentially.

### Analysis

Meaningfulness and feasibility ratings were calculated for each measure as the mean result from the 3 scales comprising each domain. Mean ratings across measures were calculated after the first and second rounds of rating and for the final core set. To explore differences among stakeholders, we grouped panelists by perspective: managers/oversight (5 panelists); clinicians/consumers (4); and researchers/evaluators (3). Sample sizes were insufficient to examine the statistical significance of these comparisons.

### Core Set Construction

A preliminary core set was selected on the basis of final measure scores. We set thresholds for meaningfulness of 3 or less with dispersion of 0.8 or less, and feasibility of 6 or less with dispersion 0.9 or less. The threshold for feasibility was lower because measures can often be improved by revising specifications.

To balance goals of measure quality and system representativeness, characteristics of the preliminary set were compared with mental health system dimensions described in the framework. To address areas that the panel identified as a priority but not represented in the preliminary set (eg, racial/ethnic disparities), the highest-rated measure in that category was added to the core set.

### Results

First-round ratings of 116 measures showed better scores and more agreement for meaningfulness than for feasibility. In the second round, 39 measures with potentially acceptable scores but poor agreement were discussed by the panel and rerated. Agreement on these measures improved after discussion, with the average dispersion in the overall score declining from 1.39 in the first round to 0.93 in the second.

Mean ratings of meaningfulness were similar across the 3 broader stakeholder groups (range: 2.89–3.07). Greater divergence was seen on feasibility, with clinicians/consumers (3.82) more likely to rate measures as feasible than managers/oversight (5.09) or researchers/evaluators (4.49).

Twenty-one measures met thresholds for inclusion in the preliminary set. Seven additional measures were added as the most highly rated in unrepresented priority areas. Table 2 presents the 28 measures comprising the final core set.<sup>24–45</sup> Mean panelist ratings (dispersion) of the meaningfulness of the set were clinical importance 2.29 (0.61); perceived gap between actual and optimal care 2.59 (0.76); and association between improved performance and outcome 2.61 (0.79). Feasibility ratings were clarity of specifications 3.39 (1.13); acceptability of data collection burden, 4.77 (1.21); and adequacy of case mix adjustment 4.20 (1.16).

With regard to the goal of representativeness, the core measures address a range of priorities across system dimensions. Of 28 measures, 12 assess treatment, 5 safety, 4 continuity, 2 access, 2 assessment, 2 coordination, and 1 prevention. Fifteen measures assess care for specific disorders: depression (4), schizophrenia (5), substance abuse/dependence (1), dementia (2), bipolar disorder (1), borderline personality disorder (1), and dual diagnosis (1). Four measures assess psychosocial interventions, 8 assess pharmacotherapy, and 2 address medical care for psychiatric patients. Settings assessed include outpatient, inpatient, residential, emergency, nursing home, and primary care. Measures addressed care for children (5), elderly patients (4), and racial/ethnic minorities (1).

Nearly all the core set measures (93%) require administrative data; 61% require data from additional sources. Sixteen measures (57%) evaluate clinical processes based on research evidence: 4 supported by randomized controlled trials and 12 by less-rigorous research.

The panel assessed readmission and ambulatory follow-up rates after hospital discharge based on commonly used intervals of 7, 14, 30, and 90 days. The panel rated the 7-day interval as most meaningful for each measure. Only the ambulatory follow-up measure met inclusion thresholds for the core set.

## Discussion

We applied a method of consensus development—incorporating stakeholder values, research data, and a means of identifying agreement while permitting trade-offs among competing priorities—to identify a core set of 28 quality measures for mental healthcare. To what extent did this core set reach the desired goals of representation, meaningfulness, and feasibility?

Breadth of representation was achieved in several dimensions. The core set includes measures addressing each process domain, 8 diagnostic categories, biologic and psy-

chosocial treatments, numerous levels of care, and clinical needs specific to several vulnerable groups.

With regard to meaningfulness, all 28 measures had scores in the “agreement” range. However, nearly half lacked supporting research evidence, and only 4 were supported by randomized controlled trials. This reflects the current state of measure development and an evidence base for mental healthcare that is narrow relative to the diverse circumstances encountered in practice.

On average, the feasibility of the core measures was rated worse than meaningfulness. In part this results from limited methods of case-mix adjustment for mental healthcare.<sup>46</sup> It also stems from concerns about the burden of data collection—the subject of greatest disagreement among panelists. Panelists from the manager/oversight groups (typically responsible for collecting data), rated measures as more burdensome than either consumers/clinicians or researchers/evaluators. Several national QA initiatives use only measures using administrative data because of the costs of medical record abstraction and administering surveys. Our results suggest that a broader array of data sources are needed to produce a set of measures that are meaningful, feasible, and diversely representative. This conclusion is consistent with the Institute of Medicine’s determination that a “mosaic of data sources” are needed to meet national quality-assessment needs.<sup>5</sup>

Achieving consensus on core quality measures presents a significant challenge, particularly in mental health care, where multiple clinical disciplines, an influential consumer movement, and a specialized managed behavioral healthcare industry bring diverse perspectives to the table. Both the method described in this report and the resulting measures may be useful to policy groups working toward standardized measures, including the Forum on Performance Measures in Behavioral Health and the National Quality Forum. The Delphi method permits identification of measures representing priorities of diverse stakeholders without requiring them to agree on each measure. Mapping ratings of measure attributes to priority system dimensions allows for balance between these considerations. The measures reported here, having been vetted by diverse stakeholders, may achieve greater acceptance and use in activities to improve care.

Our process identified gaps where further measure development is needed, including assessment of children, recovery-oriented treatment, racial and ethnic disparities, and care for posttraumatic stress disorder, developmental disorders, and mental retardation.

A limitation of Delphi processes is that few participants represent any one perspective. We sought to enhance reproducibility by selecting panelists belonging to national organizations that have well-established perspectives on quality measurement. A costly alternative would have been to conduct multiple panels and assess their reliability. Another

TABLE 2. Characteristics and Ratings of Core Measures

Measure	Meaningfulness* (Mean, Dispersion)	Feasibility† (Mean, Dispersion)	Representativeness			Data Source	Evidence Rating‡
			Population	Setting	Conditions		
<b>Treatment</b>							
≥ 1 visit with adult caregiver of child ≤ 13 treated for a psychiatric or substance-related disorder in 3-month period <sup>24</sup>	2.00 (0.67)	3.17 (0.97)	Children	Outpatient	Mental disorders, substance-use disorders	Administrative data, medical record	C
Clinician contact with family member of consenting individuals with schizophrenia at initial evaluation <sup>25</sup>	2.33 (0.67)	5.17 (0.64)	Adults	Inpatient, outpatient	Schizophrenia	Administrative data, medical record	B
Cumulative daily antipsychotic dosage between 300–1000 CPZ equivalents at hospital discharge for schizophrenia <sup>26</sup>	2.50 (0.75)	4.17 (0.89)	Adults	Inpatient	Schizophrenia	Administrative data, medical record	A
Prescription of atypical antipsychotic drug for individuals with ≥ 1 clinical service for schizophrenia in 6-month period <sup>27</sup>	2.50 (0.67)	4.67 (0.89)	Adults	All	Schizophrenia	Administrative data, pharmacy data	A
Length of treatment ≥ 90 days after initiation for substance-related disorder <sup>28§</sup>	2.50 (0.50)	3.17 (0.69)	All	All	Substance-use disorders	Administrative data	B
≥ 3 medication visits or ≥ 8 psychotherapy visits in 12-week period after new diagnosis of major depression <sup>29,30§</sup>	2.67 (0.64)	3.83 (0.89)	Adults	Outpatient	Depression	Administrative data	B
≥ 12-week continuation after initiation of antidepressant drug for major depression <sup>29</sup>	2.67 (0.50)	4.83 (0.81)	All	Outpatient	Depression	Administrative data, pharmacy data	A
Daily antipsychotic dosage between 0.5–9.0 CPZ equivalents per kg body weight at discharge for individual < 18 hospitalized for psychotic disorder <sup>31§</sup>	2.67 (0.75)	5.00 (0.64)	Children	Inpatient, residential	Psychotic disorders	Administrative data, medical record	B
Daily antipsychotic dosage ≥ 200 CPZ equivalents for nursing home resident with dementia without psychotic symptoms in 3-month period <sup>32§</sup>	3.00 (0.61)	4.83 (0.78)	Elderly	Nursing home	Dementia	Minimum data set	B
≥ 1 serum drug level taken for individuals with bipolar disorder treated with mood stabilizers in 12-month period <sup>33</sup>	3.00 (0.67)	5.50 (1.17)	All	Outpatient	Bipolar disorder	Administrative data, pharmacy data	B
Avoidance of an anticholinergic antidepressant drug for individuals ≥ 65 prescribed antidepressants <sup>34</sup>	3.33 (0.92)	5.00 (0.69)	Elderly	Inpatient, outpatient	Depression	Administrative data, pharmacy data	B
≥ 1 psychotherapy visit for individuals within 6 months of hospitalization or ER visit for borderline personality disorder <sup>35§</sup>	3.50 (1.14)	4.83 (1.00)	Adults	Outpatient	Borderline personality disorder	Administrative data	B
<b>Safety</b>							
Number of involuntary physical restraint events per patient day in 3-month period	2.17 (0.50)	2.83 (0.77)	Adults, children	Inpatient, residential	Mental disorders	Administrative data, medical record	C
Number of inpatient injuries per patient day in 3-month period <sup>36§</sup>	2.33 (0.50)	3.00 (0.63)	Adults, children, elderly	Inpatient, residential	Mental disorders, substance-use disorders	Administrative data, occurrence reports	C
Number of nursing home residents with dementia restrained physically in 3-month period <sup>32§</sup>	2.33 (0.43)	3.83 (1.17)	Elderly	Nursing home	Dementia	Minimum data set	B
Assessment of suicidal ideation at initial evaluation for patients with major depression <sup>37</sup>	2.33 (0.80)	5.33 (0.43)	All	Inpatient, outpatient, emergency service	Depression	Administrative data, medical record	C
Number of unplanned departures per patient discharge in 3-month period	2.50 (0.40)	3.17 (0.80)	Adults, children	Inpatient, residential	Mental disorders	Administrative data, medical record	B
<b>Access</b>							
Beneficiaries with ≥ 1 mental health or substance-related service in 12-month period <sup>29</sup>	2.50 (0.75)	2.83 (0.72)	All	All	Mental disorders, substance-use disorders	Administrative data	C
Denials for mental health or substance-related services per number of requests in 12-month period <sup>38§</sup>	4.00 (0.69)	3.67 (0.53)	All	Inpatient, outpatient	Mental disorders, substance-use disorders	Administrative data	C
<b>Assessment</b>							
Assessment of drug and alcohol use at initial evaluation for psychiatric disorder <sup>35</sup>	2.33 (0.73)	5.00 (1.17)	All	Inpatient, outpatient, emergency service	Mental disorders	Administrative data, medical record	C
Assessment of general medical status at initial evaluation for psychiatric disorder <sup>39</sup>	2.50 (0.80)	4.83 (0.77)	All	Inpatient, outpatient, emergency service	Mental disorders, substance-use disorders	Administrative data, medical record	C

(Continued)

TABLE 2. (Continued)

Measure	Meaningfulness* (Mean, Dispersion)	Feasibility† (Mean, Dispersion)	Representativeness			Data Source	Evidence Rating‡
			Population	Setting	Conditions		
<b>Continuity</b>							
Outpatient visit within 7 days of hospital discharge for psychiatric disorder <sup>40,41</sup>	2.00 (0.25)	3.83 (0.78)	All	Outpatient	Mental disorders, substance-related disorders	Administrative data	C
≥ 4 psychiatric and ≥ 4 substance abuse visits following discharge for dual diagnoses (psychiatric and substance abuse) in 12-month period <sup>42</sup>	2.33 (0.36)	3.50 (0.69)	Adults	Outpatient	Dual diagnosis	Administrative data	B
≥ 1 visit for individuals in 12-month period after initial visit, stratified by race/ethnicity <sup>43,8</sup>	2.33 (0.67)	4.00 (1.14)	Racial/ethnic minorities	Outpatient	Mental disorders, substance-use disorders	Administrative data	C
≥ 1 visit per month for 6 months after hospitalization for psychiatric or substance-related disorder <sup>41</sup>	2.67 (0.75)	4.33 (0.81)	Adults	Outpatient	Mental disorder, substance-related disorder	Administrative data	B
<b>Coordination</b>							
Contact with primary care clinician for consenting inpatients with primary psychiatric disorder <sup>44</sup>	2.17 (0.70)	4.67 (0.77)	All	Inpatient	Mental disorders, substance-use disorders	Administrative data, medical record	C
Enrollment in intensive case management (ICM) for patients with 4 ER visits or 2 hospitalizations for schizophrenia in 12-month period <sup>25</sup>	2.33 (0.56)	4.67 (0.83)	Adults	All	Mental disorders, substance-use disorders	Administrative data, ICM enrollment data	A
<b>Prevention</b>							
Depression screening for primary care patients during 12-month period <sup>30</sup>	2.17 (0.50)	4.67 (0.63)	All	Primary care	Mental disorders, substance-use disorders	Administrative data, medical record	C

\*Meaningfulness ratings were calculated as the average of measure scores on 3 scales: clinical importance, perceived gap between actual and optimal care, and association between improved performance and outcome.

†Feasibility ratings were calculated as the average of measure scores on 3 scales: clarity of specifications, acceptability of data collection, and adequacy of case-mix adjustment.

‡AHRQ rating categories were used to assess the research evidence supporting the association between the underlying clinical process and patient outcomes: A) good research evidence such as well-designed randomized controlled trials; B) fair research-based evidence such as retrospective analyses of observational data; C) no research evidence, principally clinical consensus or opinion.<sup>45</sup>

<sup>§</sup>Denotes that the measure was modified during the core set development process and differs from the specifications in the cited reference. Beneficiaries can refer to insurance program enrollees or to populations, as in the case of public sector programs.

approach, surveying stakeholders, would have allowed for a larger sample but without an exchange of views.

Our decision to restrict this study to a single, widely used type of measure should not be interpreted as advocating for process measures over other approaches to assessment. Multiple perspectives are essential to understanding quality—including measures of outcome, satisfaction, interpersonal process and fidelity to evidence-based treatment models. The strengths and weaknesses of these approaches complement one another and, collectively, provide a more robust armamentarium for improving quality of care.<sup>47</sup>

Resources for improving quality are limited, and measures used in national core sets will likely influence where funds are spent. These measures should thus be chosen wisely. Some of the information needed to inform their selection (eg, validity) may come only after implementation in large samples and analysis of results. This speaks to the need for an iterative process of selection and implementation that may initially rely on consensus but incorporates empirical information as it becomes available.

### Acknowledgments

We thank the contribution of the members of the stakeholder panel: John Oldham, MD (American Psychiatric Association); Mary Cesare-Murphy, PhD, and Sharon Sprenger, MPA (Joint Commission on Accreditation of Healthcare Organizations); Eric Goplerud, PhD (US Substance Abuse and Mental Health Services Administration); Mary E. Evans, PhD (University of South Florida College of Nursing); Edward R. Jones, PhD (PacifiCare Behavioral Health, Inc.); Dee Roth, MA (Ohio Department of Mental Health); Suzanne Paranjpe, PhD (Greater Detroit Health Council); Ramiro Guevara Sr. (Stanislaus Homeless Outreach Program); Wilma Townsend, MSW (Office Support Agency, Columbus, Ohio); Rex Cowdry, MD, MPH (National Alliance for the Mentally Ill); and Paul Barreira, MD (McLean Hospital/Partners Healthcare System). We also thank Kenneth Wells, MD, MPH; Deborah Garnick, ScD; Joshua C. Kendall; and Julie Regner, BS, for their helpful review of an earlier version of the manuscript.

## REFERENCES

- Institute of Medicine. *Leadership by Example: Coordinating Government Roles in Health Care Quality*. Washington, DC: The National Academies Press; 2003.
- US Public Health Service Office of the Surgeon General. *Report of the Surgeon General's Conference on Children's Mental Health: A National Action Agenda*. Washington, DC: US Public Health Service Office of the Surgeon General; 2001.
- World Health Organization. *The World Health Report 2001. Mental Health: New Understanding, New Hope*. Geneva, Switzerland: WHO; 2001. ISBN-92 4 1562013.
- President's Advisory Commission on Consumer Protection and Quality in the Health Care Industry. *Quality First: Better Health Care for All Americans Final report of the president's advisory commission on consumer protection and quality in the health care industry*. Washington, DC: U.S. Government Printing Office; 1998.
- Institute of Medicine. *Envisioning the National Health Care Quality Report*. Washington, DC: National Academy Press; 2001.
- Forum on Performance Measures in Behavioral Healthcare and Related Systems. Adult Mental Health Workgroup (AMHW). Available at: <http://www.mhindicators.org/amhw.asp>. Accessed September 13, 2004.
- MHSIP Task Force. *The Final Report of the Mental Health Statistics Improvement Program (MHSIP) Task Force on a Consumer-Oriented Mental Health Report Card*. Washington, DC: Center for Mental Health Services; April 1996.
- Mental Health Statistics Improvement Project (MHSIP). Decision Support 2000+. Available at: <http://www.mhsip.org/ds2000/newindex.htm>. Accessed September 13, 2004.
- Schacht LM, Hines H. Recent applications of risk adjustment for performance measures used by state inpatient psychiatric facilities. *Harvard Rev Psychiatry*. 2003;11:220–224.
- National Committee for Quality Assurance. *Health Plan Employer Data and Information Set (HEDIS) 2000*. Washington, DC: NCQA; 2000. Volume 1.
- Center for Mental Health Services (CMHS). CMHS Core Performance Indicator Chart. *Community Mental Health Services Block Grant Application Guidance*. Washington, DC: Substance Abuse and Mental Health Services Administration; 2004.
- McCorry F, Garnick DW, Bartlett J, et al. Developing performance measures for alcohol and other drug services in managed care plans. *Joint Commission J Quality Improvement*. 2000;26:633–643.
- Hermann RC, Leff HS, Palmer RH, et al. Quality measures for mental health care: results from a national inventory. *Medical Care Res Rev*. 2000;57(Suppl 2):135–154.
- Hermann R, Leff S, Provost S, et al. Process measures for quality assessment and improvement: are they based on research evidence? Paper presented at: NIMH Conference on Evidence in Mental Health Services Research: What Types, How Much and What Then? Available at <http://www.cqaimh.org/research.html>, 2002; Washington, DC.
- Hermann RC, Finnerty M, Provost S, et al. Process measures for the assessment and improvement of quality of care for schizophrenia. *Schizophrenia Bull*. 2002;28:95–104.
- Hermann RC. *Quality Measurement for Improving Mental Healthcare: A Guide With Results from a National Inventory of Measures*. Washington, DC: American Psychiatric Press, Inc; in press.
- Hermann RC, Palmer RH. Common ground: a framework for selecting core quality measures. *Psychiatric Services*. 2002;53:281–287.
- Schuster M, Asch S, McGlynn E, et al. Development of a quality of care measurement system for children and adolescents. *Arch Pediatr Adolesc Med*. 1997;151:1085–1092.
- Asch S, Kerr E, Lapuerta P, et al. A new approach for measuring quality of care for women with hypertension. *Arch Intern Med*. 2001;161:1329–1335.
- Normand S, McNeil B, Peterson L, et al. Eliciting expert opinion using the Delphi technique: identifying performance indicators for cardiovascular disease. *Int J Quality Health Care*. 1998;10:247–260.
- Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington DC: National Academy Press; 2001.
- Brook R. *The RAND/UCLA Appropriateness Method*. Santa Monica: RAND Corporation. RAND/RP-395.
- Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
- American Managed Behavioral Healthcare Association. *PERMS 2.0: Performance Measures for Managed Behavioral Healthcare Programs*. Washington, DC: AMBHA; July 1998.
- Young A, Sullivan G, Burnam A, Brook R. Measuring the quality of outpatient treatment for schizophrenia. *Arch Gen Psychiatry*. 1998;55:611–617.
- Lehman AF, Steinwachs DM. Patterns of usual care for schizophrenia: initial results from the schizophrenia patient outcomes research team (PORT) client survey. *Schizophrenia Bull*. 1998;24:11–20.
- National Association of State Mental Health Program Directors. *Performance measures for mental health systems: a standardized framework*. Alexandria, VA: National Association of State Mental Health Program Directors; December 4, 1998.
- New York State Office of Alcoholism and Substance Abuse Services. *IPMES/Workscope Objective Attainment System: FY2000 and 2000–2001*. Albany, NY: New York State Office of Alcoholism and Substance Abuse Services; July 2000.
- National Committee for Quality Assurance. *Health Plan Data and Information Set (HEDIS 2000)*. Washington, DC: National Committee for Quality Assurance; 1999.
- Veterans Health Administration/Department of Defense. *VHA/DOD Performance Measures for the Management of Major Depressive Disorder in Adults, Version 2.0*. Washington, DC: March 2000.
- APA Task Force on Quality Indicators for Children. *Workbook of Quality Indicators for Children*. Washington, DC: American Psychiatric Association; August 21, 2000.
- Zimmerman DR, Karon SL, Arling G, et al. Development and testing of nursing home quality indicators. *Health Care Financing Rev*. 1995;16:107–127.
- Marcus S, Olfson M, Pincus H, et al. Therapeutic drug monitoring of mood stabilizers in Medicaid patients with bipolar disorder. *Am J Psychiatry*. 1999;156:1014–1018.
- Wells K, Norquist G, Benjamin B, et al. Quality of antidepressant medications prescribed at discharge to depressed elderly patients in general medical hospitals before and after prospective payment system. *Gen Hospital Psychiatry*. 1994;16:4–15.
- American Psychiatric Association. *Report of the American Psychiatric Association Task Force on Quality Indicators*. Washington, DC: APA; March 1999.
- New York State Office of Mental Health. *Report of the Performance Measurement/Management Workgroup Project*. Albany, NY: New York State Office of Mental Health; November 18, 1998.
- Joint Commission on Accreditation of Healthcare Organizations. *National Library of Healthcare Indicators*. Oakbrook Terrace, IL; 1997.
- Washington Circle Group. *Improving Performance Measurement by Managed Care Plans for Substance Abuse: Year 1 Report of the Washington Circle Group*. Rockville, MD: Washington Circle Group; March 16, 1999.
- American Psychiatric Association Office of Quality Improvement. *APA Practice Guideline Measures*. Arlington, VA: American Psychiatric Association; 2001.
- McCorry F, Garnick DW, Bartlett J, Cotter F, Chalk M. *Improving Performance Measurement for Alcohol and Other Drug Services: Report of the Washington Circle Group*. Rockville, MD: Washington Circle Group and the Center for Substance Abuse Treatment; March 2000.
- Leslie D, Rosenheck R. Comparing quality of mental health care for public-sector and privately insured populations. *Psychiatric Services*. 2000;51:650–655.
- Rosenheck R, DiLella D. *Department of Veterans Affairs National Mental Health Program Performance Monitoring System: Fiscal Year 1999 Report*. West Haven, CT: Northeast Program Evaluation Center; March 14, 2000.
- Center for Mental Health Services. *The Final Report of the Mental Health Statistics Improvement Project (MHSIP) Task Force on a Consumer-Oriented Mental Health Report Card*. Rockville, MD: Center for Mental Health Services; 1996.

44. Massachusetts Division of Medical Assistance. *FY'99 Performance Standards*. Boston, MA: Massachusetts Behavioral Health Partnership; 1999.
45. Agency for Health Care Policy and Research Depression Guideline Panel. *Depression in Primary Care: Volume 1. Detection and Diagnosis*. Washington, DC: US Department of Health and Human Services; 1993.
46. Hermann RC. Risk Adjustment for Mental Health Care. In: Iezzoni LI, ed. *Risk Adjustment for Measuring Health Care Outcomes*. Third ed. Chicago: Health Administration Press; 2003.
47. Hermann RC. Linking outcome measurement with process measurement for quality improvement. In: IsHak W Burt T, Sederer L, eds. *Outcome Measurement in Psychiatry: A Critical Review*. Washington, DC: American Psychiatric Press, Inc.; 2002.