

# Quality Measures for Mental Health Care: Results from a National Inventory

*Richard C. Hermann*

Center for Quality Assessment and Improvement in  
Mental Health, Cambridge Hospital and Harvard Medical School

*H. Stephen Leff*

Evaluation Center, Human Services Research Institute

*R. Heather Palmer*

Harvard School of Public Health

*Dawei Yang*

*Terri Teller*

*Scott Provost*

*Chet Jakubiak*

*Jeff Chan*

Center for Quality Assessment and Improvement in Mental Health

*The National Inventory of Mental Health Quality Measures was funded by the Agency for Healthcare Research and Quality to (1) inventory process measures for assessing the quality of mental health care; (2) identify clinical, administrative, and quality domains where measures have been developed; and (3) identify areas where further research and development is needed. Among the 86 measures identified, most evaluated treatment of major mental disorders, for example, schizophrenia (24 percent) and major depression (21 percent). A small proportion focused on children (8 percent) or the elderly (9 percent). Domains of quality included treatment appropriateness (65 percent), continuity (26 percent), access (26 percent), coordination (13 percent), detection (12 percent), and prevention (6 percent). Few measures were evaluated for reliability (12 percent) or validity (3 percent). Measures imposing a lower burden were more likely to be in use ( $\chi^2 = 4.41$ ,  $p =$*

---

This work was supported by grants by the National Institute of Mental Health (K08-MH001477), the Agency for Healthcare Research and Quality (R01-HS10303), the Substance Abuse and Mental Health Services Administration, and the Evaluation Center@HSRI

Medical Care Research and Review, Vol. 57 Supplement 2, (2000) 136-154  
© 2000 Sage Publications, Inc.

*.036). Further measures are needed to assess care for several priority clinical and demographic groups. Research should focus on measure validity, reliability, and implementation costs. In order to foster quality improvement activities and use of common measures and specifications for mental health care, the inventory of quality measures will be made available at [www.challiance.org/cqaimh](http://www.challiance.org/cqaimh).*

A number of factors have converged to establish a national agenda to monitor and improve the quality of health care. Rapid changes in the organization and financing of care have put unprecedented pressure on health care delivery organizations to reduce utilization and costs, leading to a need to ensure that quality is not adversely affected. Influential panels such as the President's Commission on Consumer Protection and Quality in the Health Care Industry (President's Advisory Commission on Consumer Protection and Quality in the Health Care Industry 1998) and the Institute of Medicine's National Roundtable on Health Care Quality (Chassin and Galvin 1998) have spearheaded national initiatives to improve quality, basing their recommendations on two decades of research showing widespread practice variation, including underuse, overuse, and misuse of services. Consumer dissatisfaction with the health care system has risen to high levels over the past decade and has led to legislative proposals for system reform (Donelan et al. 1999).

Frameworks and tools for the systematic improvement of care are at an early stage of development (Berwick, Godfrey, and Roessner 1990; Blumenthal and Kilo 1998; Shortell, Bennett, and Byck 1998). A priority is the development of measures of quality, which are urgently needed to facilitate improvement activities, regulatory oversight, and selection of plans and providers by purchasers and consumers. A number of measure-development efforts are under way nationally, including those led by the Agency for Healthcare Research and Quality (AHRQ), the Foundation for Accountability (FACCT), the Joint Commission on Accreditation of Healthcare Organizations (JCAHO), the National Committee for Quality Assurance (NCQA), and the Health Care Financing Administration (HCFA).

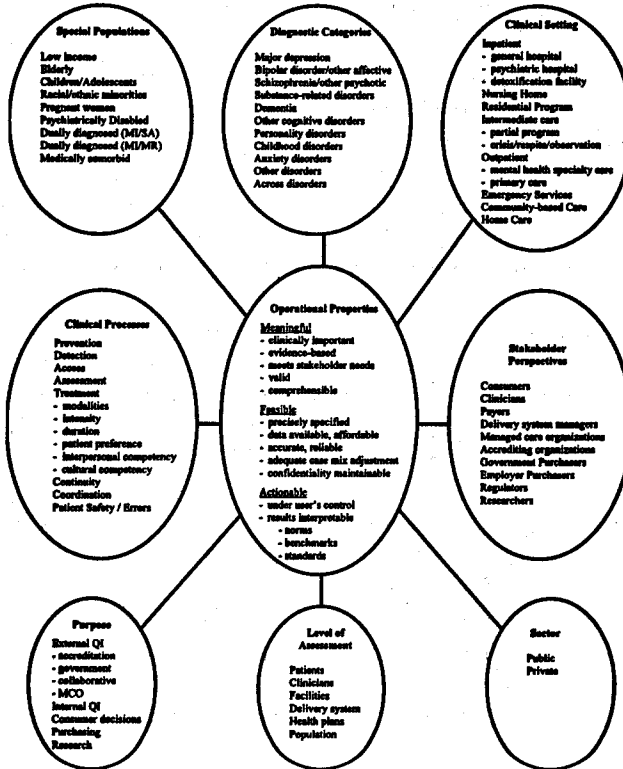
The President's Commission highlighted mental health as an area in which quality measures are underdeveloped (President's Advisory Commission on Consumer Protection and Quality in the Health Care Industry 1998). Severe and persistent psychiatric disorders affect more than 5 million Americans (Kessler et al. 1992; National Advisory Mental Health Council 1995) and \$79 billion are spent annually for mental health care in the United States (Mark et al. 1996). Medication and psychosocial treatments of proven efficacy exist for many mental disorders (Gabbard 1995; U.S. Department of Health and

Human Services 1999), yet variation in their use has been documented in mental health care as elsewhere in medicine (Hermann et al. 1995; Hermann 1996).

Recognition of the importance of quality assessment and improvement is emerging throughout the mental health community (Chisholm et al. 1997; McFarland et al. 1998; Rosenheck and Cicchetti 1998; Salzer et al. 1997; Young et al. 1998). For example, the National Advisory Mental Health Council called for advancing the development of quality measures in its 1998 report to Congress, "Bridging Science and Service" (National Advisory Mental Health Council's Clinical Treatment and Services Research Workgroup 1998). A number of measure-development efforts have been initiated by stakeholders, including accreditors, purchasers, managed behavioral healthcare organizations, government agencies, provider organizations, advocacy groups, and mental health services researchers. However, little is known about the status of these measures or the degree to which they meet the needs of stakeholders.

Figure 1 outlines an organizational framework for evaluating the status of quality measures in mental health. The framework takes a societal perspective that encompasses quality assessment of mental health care from the perspectives of all major stakeholders and across the U.S. healthcare system. The central category—operational properties—summarizes attributes for evaluating the meaningfulness, feasibility, and actionability of measures. The surrounding areas are dimensional components of the health care system that should be considered in measure development, selection, and implementation: clinical processes in mental healthcare, clinical settings, diagnostic categories, subpopulations with special clinical needs, stakeholder perspectives, measure purpose, level of assessment, and economic sector. Collectively, these considerations form a multidimensional matrix of considerations that can be used to evaluate the availability and adequacy of current quality measures and to identify key gaps to inform further measure development and research.

The National Inventory of Mental Health Quality Measures was funded by the AHRQ and the Substance Abuse and Mental Health Services Administration (SAMHSA) to evaluate process measures for the assessment of mental health care. In this report, we describe results from the inventory, focusing on (1) the characteristics of measures that have been developed; (2) diagnostic, treatment, and demographic areas where measures are available; and (3) priority areas for further development and research. The inventory is based on the quality measurement evaluation system, CONQUEST 2.0 (a Computerized Needs-Oriented Quality Measurement Evaluation System), developed by Palmer et al. to enable health plans, providers, and purchasers to identify, select, and use clinical quality measures (Lawthers and Palmer 1997). CONQUEST currently offers profiles of more than 1,100 measures. Few measures of mental health care were originally included in CONQUEST. Most of



**FIGURE 1 Organizational Framework for Assessing Mental Health Quality Measures**

Note: MI = mental illness, SA = substance abuse, MR = mental retardation, QI = quality improvement, MCO = managed care organization.

the mental health measures identified in the course of our research were developed after CONQUEST 2.0 was completed. To facilitate a focus on mental health measures, we added a number of categories and items to CONQUEST.

**METHODS**

For the purposes of this study, a quality measure is defined as “a quantitative measure that can be used as a guide to monitor and evaluate the quality of important patient care and support service activities” (Joint Commission of Accreditation of Healthcare Organizations 1989, 330). A number of other

terms are also used, such as *indicators* rather than *measures* and *performance* rather than *quality*. In the terms of Donabedian's (1980) framework, characteristics of structure, process, and outcome can all be measured, and measures from each domain can be used to evaluate quality. This inventory focuses on measures of clinical and administrative process. Process measurement is only one approach to assessing health care quality. Other approaches, including outcome measurement and evaluation of patient satisfaction, have important roles in quality assessment, and each presents different opportunities and challenges (Hermann in press). This study assesses those associated with mental health process measures.

Certain measures, such as hospital readmission rates, are sometimes referred to as process-based "proxies" for treatment outcomes. Because they are, in fact, treatment processes, they were included in this inventory. Clinical processes result in the provision of health care, while administrative processes do not directly result in the provision of health care but are closely related to care provision and can affect clinical outcome. An example of an administrative process measure, from the domain of continuity of care, is whether an outpatient visit was scheduled for a patient prior to discharge from an inpatient setting.

The vast majority of measures identified were rate-based, which is to say they were proportions constructed from a numerator and denominator. A small number of measures yielded means or a count of incidents over time (e.g., a hospital's number of physical restraints of inpatients over a 3-month period).

*Identification of measures.* Using chain-referral and snowball collection techniques, we systematically identified measures developed by U.S. individuals and organizations for the assessment of the quality of mental health care. Between November 1999 and April 2000, we contacted accreditors (e.g., NCQA and JCAHO); federal, state, and local agencies (e.g., all 50 state departments of mental health); the Department of Veterans Affairs; SAMSHA; the National Association of State Mental Health Program Directors; mental health provider organizations (e.g., the American Psychiatric Association and the National Association of Psychiatric Health Systems); delivery system managers (e.g., the American College of Mental Health Administration); managed care organizations (e.g., the American Managed Behavioral Healthcare Association and individual companies); payers; purchaser and consumer groups (e.g., FACCT); health systems; and quality of care researchers (e.g., investigators from the Schizophrenia PORT and the RAND study of the Medicare prospective payment system). Contact modalities included letters, phone calls, and web site reviews. We also conducted reviews of published quality of care re-

search literature and databases of active grants, using MEDLINE, PsycLit, and CRISP.

*Inclusion criteria.* To be included in the inventory, a process measure had to meet a minimum threshold of development, which was defined as having (1) a specified numerator and denominator, (2) a designated data source, and (3) a face relationship to quality. Measures that only described utilization without a directional association to quality (e.g., length of hospital stay) or prevalence of clinical conditions (e.g., the proportion of a nursing home population with depressive symptoms) were excluded. In focusing on discrete measures, we excluded multi-item instruments that assess quality of care. An exception to this rule occurred when an organization had proposed a stand-alone measure based on a subset of instrument items. An example is a measure from the Mental Health Statistics Improvement Project's (MHSIP's) Consumer-Oriented Mental Health Report Card. The measure is based on responses to a single question from the MHSIP consumer survey: the proportion of consumers surveyed who reported that they strongly agree or agree with the statement, "Services were available at times that were convenient."

*Inventory of measures.* We employed the CONQUEST inventory format to evaluate mental health measures, using a relational database—Microsoft Access 97—as the software platform. Prior to the inventory of each measure, the developer was contacted for supporting documentation. A literature review was conducted to obtain background information on the subject and clinical context and to evaluate the scientific evidence relating to the measure. Using this material, the measure was inventoried by a member of the research team, describing or evaluating the properties described below. Some of the properties were drawn from CONQUEST 2.0, with additional attributes developed by the research team.

## DEVELOPMENT

*Stakeholder involvement.* Stakeholders that may have participated in a measure's development include: accreditors, payers, employers, consumers, clinicians, managed care organizations, delivery system managers, and researchers.

*Level of development.* Measures meeting a minimum threshold of development were included in the inventory. Thus, we further classified them by whether they were fully operationalized (i.e., they could be implemented based on current documentation) or whether they needed further development.

*Development method.* Methods used to develop measures included consensus by participants, clinician panel using formal methods (e.g., Delphi panel), literature review, or the application of preexisting instruments or published guidelines.

*Basis in scientific evidence.* We used U.S. AHRQ categories for assessing the scientific evidence underlying the practice addressed by the following measure: (1) good research-based evidence, such as well-designed randomized controlled trials; (2) fair research-based evidence, such as retrospective analyses of observational data, with supporting clinical consensus or opinion; and (3) little research evidence, principally clinical consensus or opinion (Agency for Health Care Policy and Research Depression Guideline Panel 1993).

*References.* We listed primary references used in the coding, including documentation published by the developers and research articles used for the rationale to assess the scientific basis or other aspects of the measure.

## SPECIFICATIONS

*Numerator and denominator specifications.* These specifications included the measure's target population, sampling, time frame, and inclusion and exclusion criteria.

*Data source.* Data sources included administrative data (from enrollment and billing claims), medical records, pharmacy claims, patient surveys, and laboratory records.

## CONTENT

*Clinical rationale.* The rationale summarized the clinical context for the measure, including its association to quality of care.

*Domain of quality.* Domains of quality included prevention, detection, access, assessment, treatment, coordination, continuity, and patient safety.

*Applicable diagnoses.* These included major depression, schizophrenia, bipolar disorder, substance abuse and dependence, dementia, dysthymia, and personality disorders.

*Principal treatment modality.* Possible modalities included medication, psychotherapy, case management, assertive community treatment, other psychosocial treatments, and other somatic treatments.

*Special populations.* Several population subgroups have particular clinical or policy significance, including low income groups, racial/ethnic minorities, individuals with severe and persistent mental illness, dual diagnosis (i.e., patients having both psychiatric and substance-related disorder), children/adolescents, elderly individuals, and pregnant women.

*Location or level of care.* Location or level of care for which the measure was designed included hospitals (general, psychiatric, or substance abuse), ambulatory settings, rehabilitation, nursing homes, emergency service, and community-based or home care.

*Organizational settings.* The organizational settings for measure use include fee-for-service, managed care, or government health care facility or system. Measures were classified on the basis of the developers' intent or from settings of actual use.

## OPERATIONAL PROPERTIES

*Reliability.* Questions included the following: was reliability testing performed, the type of testing (i.e., test-retest, interrater, data accuracy, internal consistency), and the results (positive, negative, or mixed).

*Validity.* Questions included the following: was validity testing performed, the type of testing (i.e., comparison with the results from other measures, criterion, or gold standard), and the results (positive, negative, or mixed).

*Case mix adjustment.* In using measures to compare the performance of providers, facilities, or plans, case mix adjustment is sometimes needed to account for patient factors such as illness severity, comorbid conditions, and/or socioeconomic status. Typically, risk adjustment is unnecessary when the measure is only influenced by provider behavior, but becomes more important when the measure is influenced by patient behavior (e.g., showing up for visits). We classified each measure in terms of the need for risk adjustment: not needed, present and adequate, present but inadequate, and unknown. If present, we also classified the method of risk adjustment: paired data at patient level, analysis by subgroup, or other.

## USE

*Purpose.* We categorized measures by the purpose of use: internal quality improvement (measurement within a health care plan, delivery system, or facility to facilitate improvement activity within that entity), external quality



improvement (measurement conducted external to the organization providing care, such as an accreditor, managed care organization, government agency, or collaborative, for the purpose of facilitating improvement within the care-providing organization), health plan purchasing, consumer choice of plans or providers, or provider contracting.

*Implementation.* We documented whether the measure had been implemented, pilot-tested, or not used or tested. We also listed the extent of use: one site, at multiple sites within a single organization, or by multiple organizations.

*Conformance results.* If a measure had been used, we included conformance results in the inventory along with available norms or benchmarks.

*Potential comparisons.* We documented whether the measure could be used to compare the organization's results to their previous performance or to the performance of other organizations. We also documented any standards set by the developer or users of a measure.

*Cost.* We included any information available on the cost or cost-effectiveness of implementing the measure.

## REVIEW OF MEASURE INVENTORIES

Each inventory was drafted by a single member of the research team and presented at a weekly meeting where it was reviewed by at least three team members—including the principal investigator—for the accuracy and consistency of assessments. Divergent assessments, though infrequent, were discussed and a consensus was reached. Through this process, decision rules were formulated, documented in a coding manual, and disseminated to research staff weekly. Each inventory was then rereviewed by a single senior researcher to ensure conformance with the decision rules that had been established.

## ANALYSES

The inventory process is ongoing and this report presents an interim analysis of the first 86 measures. Characteristics of selected inventory results are presented in terms of the proportion of all measures. Associations among selected characteristics (data source, use, and scientific evidence) were analyzed using the Mantel-Haenszel chi-square test.

TABLE 1 Proportion of Measures by Development Characteristics

<i>Development Characteristic</i>	<i>Measures</i>	
	<i>n</i>	<i>%</i>
<b>Type of developer organization</b>		
Researchers	25	29.1
Clinician organizations	22	25.6
Government agencies	19	22.1
Managed behavioral health care organizations	13	15.1
Accreditors	6	7.0
Purchasers/consumers	1	1.2
<b>Stakeholder involvement in development</b>		
Researchers	77	89.5
Clinicians	74	86.0
Payers	35	40.7
Delivery system managers	34	39.5
Managed care organizations	26	30.2
Accrediting organizations	24	27.9
Consumers	18	20.9
Employers	8	9.3

Note: For the stakeholder involvement in development characteristics, numbers may add up to more than 100 percent because the categories were not exclusive.

## RESULTS

There were 149 process measures of the quality of mental health care that were identified; 22 were duplicates, 16 were insufficiently operationalized for inclusion, and 25 lacked a face association with quality, leaving 86 for further analysis.

As shown in Table 1, researchers developed the largest number of measures (29.1 percent), followed by clinician organizations (25.6 percent), government agencies (22.1 percent), and managed care organizations (15.1 percent). Only six measures (7.2 percent) were developed by accreditors and only one measure (1.2 percent) was developed by an organization made up of purchasers and consumers.

Most organizations had multiple stakeholders participate in the measure development process. Researchers and clinicians were most likely to participate, involved in 89.5 percent and 86 percent respectively, of the measures developed (see Table 1). Participation by other stakeholders included payers (40.7 percent), delivery system managers (39.5 percent), managed care organi-

zations (30.2 percent), accreditors (27.9 percent), consumers (20.9 percent), and employers (9.3 percent).

Most domains of quality were represented by a significant number of measures (see Table 2). The majority of measures (65.1 percent) assessed the appropriateness of treatment. The smallest proportion of measures (5.8 percent) assessed preventative aspects of care. Other domains included access (25.6 percent), continuity (25.6 percent), coordination (12.8 percent), detection (11.6 percent), and assessment (10.5 percent).

As shown in Table 2, more measures were developed to evaluate care across diagnoses (39.5 percent) than care for any single condition. Schizophrenia (24.4 percent) and major depression (20.9 percent) were the disorders most likely to be assessed. Few measures were identified for substance abuse and dependence (8.1 percent), bipolar disorder (3.5 percent), or dementia, personality disorders, or dysthymia (each 1.2 percent). Of all the measures, 15.1 percent focused explicitly on patients with severe and persistent mental illness. Fewer focused on other priority subgroups: elderly individuals (9.3 percent), children and adolescents (8.1 percent), or patients with dual diagnoses (4.7 percent), comorbid medical conditions (3.5 percent), or pregnancy (1.2 percent).

Of the measures, 37.2 percent focused on medication treatment, while psychosocial treatment was the focus of 38.4 percent (psychotherapy, 10.5 percent; case management and assertive community treatment, 10.5 percent; substance abuse counseling, 3.5 percent; and other psychosocial modalities, 16.3 percent). Of the measures, 3.5 percent assessed other modalities, while 20.9 percent did not focus on a specific treatment modality.

A substantial majority of measures were based at least in part on research evidence (see Table 2), with 39.5 percent evaluated as based on good research evidence (i.e., well-designed, randomized controlled trials), 40.7 percent based on fair research evidence, and 19.8 percent of measures were based largely on clinical consensus or opinion. Of measures that had been implemented, only 20 percent were based on good research evidence.

Table 3 shows the distribution of measures by selected technical characteristics. Of the measures, 37.2 percent required only administrative data for construction, while medical record data was required for 34.9 percent, patient survey data for 19.8 percent, pharmacy claims data for 22.1 percent, and laboratory data for 3.5 percent.

Only 11.6 percent of measures had been tested for reliability and 3.5 percent for validity. No cost assessments had been performed for any of the measures inventoried. Case mix adjustment was determined to be required and adequate in 14.3 percent of measures, required and inadequate in 38.4 percent of measures, and not needed in 23.3 percent of measures. In another 24.4 percent

TABLE 2 Proportion of Measures by Content Characteristics

Content Characteristic	Measures	
	n	%
Domain of quality		
Appropriateness	56	65.1
Access	22	25.6
Continuity	22	25.6
Coordination	11	12.8
Detection	10	11.6
Assessment	9	10.5
Prevention	5	5.8
Diagnosis		
Schizophrenia	21	24.4
Major depressive disorder	18	20.9
Substance abuse/dependence	7	8.1
Bipolar disorder	3	3.5
Dementia	1	1.2
Personality disorders	1	1.2
Dysthymia	1	1.2
Across diagnoses	34	39.5
Principal treatment modality		
Medication	32	37.2
Psychosocial modalities	33	38.4
Psychotherapy	9	10.5
Intensive case management/assertive community treatment	9	10.5
Substance abuse counseling	3	3.5
Other psychosocial	14	16.3
Other modalities	3	3.5
Not applicable	18	20.9
Use with specific populations		
Serious and persistent mental illness (SPMI)	13	15.1
Elderly	8	9.3
Children and adolescents	7	8.1
Dual diagnosis	4	4.7
Comorbid medical condition	3	3.5
Pregnant women	1	1.2
Strength of scientific evidence		
Good research-based evidence	34	39.5
Fair research-based evidence and supporting clinical opinion	35	40.7
Little research evidence, based principally on clinical opinion	17	19.8

Note: For the domain of quality and principal treatment modality characteristics, numbers may add up to more than 100 percent because the categories were not exclusive.

TABLE 3 Proportion of Measures by Selected Technical Characteristics

Technical Characteristic	Measures	
	n	%
Data source		
Administrative only	32	37.2
Medical record	30	34.9
Pharmacy claims data	19	22.1
Patient survey	17	19.8
Laboratory data	3	3.5
Development status		
Fully operationalized	43	50.0
Incomplete	43	50.0
Testing		
Measures with reliability testing	10	11.6
Measures with validity testing	3	3.5
Measures with established standards	14	16.3
Measures with cost assessments	0	—
Risk adjustment		
Adequate	12	14.3
Inadequate	33	38.4
Not needed	20	23.3
Need not known	21	24.4
Purpose		
External quality improvement	19	100
Internal quality improvement	12	63.2
Provider contracting	5	26.3
Consumer choice	0	
Purchaser decisions	0	
Measure use		
Measure defined, not tested or used	28	32.6
Measure pilot tested, not used	39	45.3
Measure implemented	19	22.1

Note: For the data source and purpose characteristics, numbers may add up to more than 100 percent because categories were not exclusive.

of measures, the need for adjustment based on patient characteristics was unclear.

Half of the measures included in the sample were fully operationalized and ready for use, while the other 43 needed further development. Of the measures, 22.1 percent had been implemented in routine quality assessment or

improvement activities and 45.3 percent had been pilot tested or used in a research study. Standards had been proposed for 16.3 percent of measures.

Of the 19 implemented, all had been used for external quality improvement, while 63.2 percent were additionally used for internal quality improvement and 26.3 percent for provider contracting. None of the measures had been used to facilitate consumer or purchaser selection of providers or plans.

Measures requiring only administrative data were more likely to have been used ( $\chi^2 = 4.41, p = .036$ ). Measures based on fair research evidence (AHRQ Level B) were more likely to have been implemented (37.1 percent) than either measures based on good evidence (Level A, 11.8 percent) or based principally on opinion (Level C, 11.8 percent).

## DISCUSSION

Over the past few years, the development of quality measures for mental health care has been in a stage of letting many flowers bloom. The strength of this process is that it has produced many measures covering a wide range of quality domains, clinical conditions, and types of treatment. However, current measures show wide variation in the extent of their development and testing. Comparing these measures with the framework described in Figure 1 identifies numerous gaps between existing measures, components of the health care system, and quality-related needs of primary stakeholders. These gaps, summarized below, provide an agenda for the improvement of the meaningfulness, feasibility, and actionability of existing measures and development of new measures in priority areas. In general, a wide range of stakeholders participated in the development of existing measures. Of greatest concern was the low rate of consumer participation and the near absence of participation by employer purchasers. Given the absence of these groups in the development process, it is perhaps not surprising that among the measures currently in use, none are being used for consumer or purchaser selection of providers or plans.

A majority of measures examined the appropriateness of treatment, including the type, intensity, and duration of care. Most other domains of quality—for example, access, continuity, and coordination—included a substantial number of measures. However, their numbers quickly decreased when secondarily sorted by setting, treatment modality, and other considerations, leaving a small number of measures—often none—in any one subdomain.

Most diagnosis-specific measures focused on more severe conditions, particularly schizophrenia and major depression. This initial focus may have been guided by the disability associated with these conditions and the high prevalence of the latter disorder. There were few measures for prevalent but less severe conditions such as dysthymia, anxiety disorders, or personality

disorders. There were few measures focusing on the mental health needs of elderly individuals—or for dementia, a high-prevalence condition in this cohort—despite the projected growth in this segment of the population and their importance in financial and policy issues regarding Medicare. Only a handful of measures were identified for other populations with specialized clinical needs, such as children and individuals with dual diagnoses.

Available measures were evenly distributed between biologic and psychosocial modalities. Although there were some measures related to psychotherapy, in general there was an absence of measures assessing the content of this modality, instead focusing on whether it was provided and the visit frequency. This was in contrast to measures of medication treatment, where several measures evaluated the content of care, including the type of medication selected, dose, duration of treatment, and monitoring of side effects. The absence of such measures for psychotherapy may reflect a lesser consensus among researchers and clinicians regarding parameters for psychotherapy use, such as indications, selection among types, and duration.

Only 22 percent of measures have been implemented for routine quality assessment or improvement. Half of all measures required further operationalization before they could be used, typically more precisely specified numerators, denominators, and data sources. To date, measures most commonly have been implemented for the purpose of externally driven quality improvement, followed by internal quality improvement and provider contracting. The small number of measures implemented has resulted in a paucity of aggregate data available for the calculation of norms and the comparative interpretation of results, although pilot testing of measures in research studies and in a National Association of State Mental Health Program Directors (NASMHPD) Research Institute multistate feasibility test has yielded some preliminary results. Few developers or users have set standards for these measures and little progress has been made toward developing benchmarks (i.e., quantitative results from best practices).

It is encouraging to observe that over 80 percent of measures had some foundation in research-based evidence; however, measures based on good research evidence (AHRQ Level A) were less frequently implemented than measures based on lesser evidence. Measures based on administrative data, such as billing claims, were more likely to be used than measures based on more costly and labor-intensive sources of data, such as medical record abstraction or patient surveys. Administrative claims lack much of the clinical detail found in medical records and surveys. Laboratory reports and pharmacy claims are computer-based sources of data with clinically relevant information, but their availability is currently limited by integration of information systems and, in the case of pharmacy claims, the benefit structure of plans.

Eddy (1998) has argued persuasively for the need for formal work-ups of quality measures, including their statistical characteristics, feasibility, and cost-effectiveness. Our data bear out his observation that "most organizations that promote measures do not work them up rigorously. . . . Those that try to are severely hampered by the cost" (Eddy 1998, 20). More than half of the measures were insufficiently developed for implementation. Few measures had been tested for reliability or validity.

For more than two-thirds of inventoried measures, case mix adjustment was judged to be necessary but unavailable. Adjustment was particularly important for measures using utilization data to draw inferences about the appropriateness of clinical practices. For example, several measures used data from paid claims to determine the adequacy of medication treatment, for example, visit frequency or duration. However, patient noncompliance with scheduled visits can affect these results and individuals with severe mental illness have been reported to have "no-show" rates of greater than 40 percent (Smoller et al. 1998). Case mix adjustment based on such factors as illness type and severity may be useful in using these measures to compare performance among providers or plans.

In addition to describing the current status of quality measures and identifying areas for further development, the inventory is intended to contribute to the development of core measure sets. Currently, providers are asked to implement, report, and improve on different measures by each payer and managed care organization with which they contract and every accreditor and regulatory agency by which they are overseen. Even when two groups request the same measure, often they require different specifications.

Consensus is needed on a core set of the best measures available to decrease the burden on providers and allow them to use limited resources to move beyond reporting and focus on improvement. If adopted by a wide range of organizations, core sets can provide comparable data on providers, facilities, and plans, and can lead to the development of regional and national norms. Fewer measures also will allow for resources to be concentrated on their further development and testing, which most measures still need.

Several organizations have sought to develop measure sets, and some, including the American College of Mental Health Administration and the National Association of State Mental Health Program Directors, have worked to develop consensus-based core sets. None of these efforts has produced definitive results thus far. The reasons are varied and include political, technical, and resource-related issues, exacerbated by the tendency for each group to start this work anew. An intention of this inventory is to draw together previous work from these organizations, services research, and implementation initiatives to provide a foundation for subsequent initiatives.



The third purpose of this inventory is to provide a searchable database of mental health quality measures that will allow potential users to identify those meeting their needs. Following the model of CONQUEST, the information will be made available via an Internet-based, user-friendly relational database ([www.challiance.org/cqaimh](http://www.challiance.org/cqaimh)).

As health care moves more firmly into a competitive marketplace with ongoing pressures to lower costs, raise efficiency, and improve quality, formalized methods of quality assessment and quality improvement are becoming increasingly important. Continuous Quality Improvement (CQI) and similar models can lead to change, but only if based on accurate measures of important processes and outcomes (Hermann et al. 2000). Put another way, measurement-based quality improvement can be only as good as its measures. Our inventory indicates that mental health care has begun to operationalize measurement of important areas, but further testing, development, and consensus will be needed to realize the possibilities of CQI and related efforts.

## REFERENCES

- Agency for Health Care Policy and Research Depression Guideline Panel. 1993. *Depression in Primary Care: Volume 1. Detection and Diagnosis*. Washington, DC: U.S. Department of Health and Human Service.
- Berwick, D. M., A. B. Godfrey, and J. Roessner. 1990. *Curing Health Care*. San Francisco: Jossey-Bass.
- Blumenthal, D. and C. M. Kilo. 1998. A Report Card on Continuous Quality Improvement. *Milbank Quarterly* 76 (4): 625-48.
- Chassin, M. R. and R. W. Galvin. 1998. The Urgent Need to Improve Health Care Quality: Institute of Medicine National Roundtable on Health Care Quality. *Journal of the American Medical Association* 280 (11): 1000-5.
- Chisholm, M., P. B. Howard, M. A. Boyd, J. A. Clement, M. J. Hendrix, and B. Reiss-Brennan. 1997. Quality Indicators for Primary Mental Health within Managed Care: A Public Health Focus. *Archives of Psychiatric Nursing* 11 (4): 167-81.
- Donabedian, A. 1980. *Explorations in Quality Assessment and Monitoring: The Definition of Quality and Approaches to Its Assessment*. Vol. 1. Ann Arbor, MI: Health Administration Press.
- Donelan, K., R. J. Blendon, C. Schoen, K. Davis, and K. Binns. 1999. The Cost of Health System Change: Public Discontent in Five Nations. *Health Affairs* 18:206-16.
- Eddy, D. M. 1998. Performance Measurement: Problems and Solutions. *Health Affairs* 17 (4): 7-25.
- Gabbard, G. O., ed. 1995. *Treatments of psychiatric disorders*. Vol. 1. Washington, DC: American Psychiatric Press.
- Hermann, R. C. 1996. Variation in Psychiatric Practices: Implications for Health Care Policy and Financing. *Harvard Review of Psychiatry* 4:98-101.

- . In press. Linking Outcome Measurement with Process Measurement for Quality Improvement. In *Outcome Measurement in Psychiatry: A Critical Review*, edited by W. IsHak, T. Burt, and L. Sederer. Washington, DC: American Psychiatric Press.
- Hermann, R. C., R. A. Dorwart, C. W. Hoover, and J. Brody. 1995. Variation in ECT Use in the United States. *American Journal of Psychiatry* 152:869-75.
- Hermann, R. C., J. L. Regner, D. Yang, and P. Erickson. 2000. Developing a Quality Management System for Behavioral Healthcare: The Cambridge Health Alliance Experience. *Harvard Review of Psychiatry* 8(5).
- Joint Commission of Accreditation of Healthcare Organizations. 1989. Characteristics of Clinical Indicators. *Quality Review Bulletin* 15:330-39.
- Kessler, R. C., K. A. McGonagle, S. Zhao, C. B. Nelson, M. Hughes, S. Eshleman, H. U. Wittchen, and K. S. Kendler. 1992. Lifetime and 12-Month Prevalence of DSM-III-R Psychiatric Disorders in the United States: Results from the National Comorbidity Survey. *Archives of General Psychiatry* 5:8-19.
- Lawthers, A. G. and R. H. Palmer. 1997. In Search of a Few Good Performance Measures: CONQUEST and the Typology of Clinical Performance Measures. In *Models for Measuring Quality in Managed Care: Analysis and Impact*, edited by J. Seltzer and D. Nash, 121-50. New York: Faulkner and Gray.
- Mark, T., D. McKusick, E. King, H. Harwood, and J. Genuardi. 1996. *National Expenditures for Mental Health, Alcohol and Other Drug Abuse Treatment*. Rockville, MD: U.S. Department of Health and Human Services.
- McFarland, B. H., R. A. George, W. Goldman, D. A. Pollack, J. McCulloch, S. Penner, and R. H. Angell. 1998. Population-Based Guidelines for Performance Measurement: A Preliminary Report. *Harvard Review of Psychiatry* 6:23-37.
- National Advisory Mental Health Council. 1995. *Basic Behavioral Science Research for Mental Health*. Bethesda, MD: National Institutes of Health.
- National Advisory Mental Health Council's Clinical Treatment and Services Research Workgroup. 1998. *Bridging Science and Service*. Bethesda, MD: National Institutes of Health/National Institute of Mental Health.
- President's Advisory Commission on Consumer Protection and Quality in the Health Care Industry. 1998. *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: Government Printing Office.
- Rosenheck, R. and D. Cicchetti. 1998. A Mental Health Program Report Card: A Multi-dimensional Approach to Performance Monitoring in Public Sector Programs. *Community Mental Health Journal* 34 (1): 85-106.
- Salzer, M. S., C. T. Nixon, L.J.A. Schut, M. S. Karver, and L. Bickman. 1997. Validating Quality Indicators: Quality as Relationship between Structure, Process, and Outcome. *Evaluation Review* 21 (3): 292-309.
- Shortell, S. M., C. L. Bennett, and G. R. Byck. 1998. Assessing the Impact of Continuous Quality Improvement on Clinical Practice: What It Will Take to Accelerate Progress. *Milbank Quarterly* 76 (4): 593-624.

- Smoller, J. W., R. Y. McLean, M. W. Otto, and M. H. Pollack. 1998. How Do Clinicians Respond to Patients Who Miss Appointments? *Journal of Clinical Psychiatry* 59 (6): 330-38.
- U.S. Department of Health and Human Services. 1999. *Mental Health: A Report of the Surgeon General*. Rockville, MD: U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institutes of Health, National Institute of Mental Health.
- Young, A. S., G. Sullivan, A. M. Burnam, and R. H. Brook. 1998. Measuring the Quality of Outpatient Treatment for Schizophrenia. *Archives of General Psychiatry* 55 (7): 611-17.