

A Brief Overview of The Forum on Performance Measures in Behavioral Healthcare

Background

Over the past 15 years the extensive changes in the structure and financing of healthcare in this country and the proliferation of new evidence-based approaches to the recognition and treatment of diseases, including the behavioral health disorders, have made issues of quality and accountability of paramount importance (IOM *Crossing the Quality Chasm: A New Health System for the 21st Century* 2001). A central component of all effective programs that monitor and improve quality and foster accountability in the delivery of appropriate healthcare services is the development and implementation of empirical measures of non-financial performance. Recently, the need for such empirical measures of accountability and quality has been strongly re-emphasized. One example of this renewed interest is the recent attention being given to the establishment of standardized measures of non-financial performance by the IOM, both in its 2002 *Leadership by Example* and in its current project to adapt the principles and recommendations of the *Crossing the Quality Chasm* report to behavioral healthcare,. In addition, performance measurement has been given a central role within the SAMHSA matrix, as evidenced by the establishment of key domains of performance for agency programs and the development of the National Outcomes Measures (NOMS) program. Therefore, the need for the development and successful implementation of these measures has only increased.

Unfortunately, efforts in performance measurement in behavioral healthcare have, over the past decade, often lagged behind efforts made in physical healthcare. Although the various centers within SAMHSA have devoted significant resources over the past 10 years to the area of performance measurement (including support for the MHSIP Survey, the *Proposed Set of Consensus Indicators for Behavioral Healthcare* of the American College of Mental Health Administrators, the 1st Forum on Common Performance Measures at the Carter Center in 2001, and the consistent support provided by CSAT to the Washington Circle since 1998), these efforts have only partially succeed in establishing standard measures for their respective fields. Most often, development efforts have stalled at the domain and indicator level (e.g., perception of timeliness of access to care), with no agreement on standardized and common specifications for measures that would support comparability and aggregation.

On March 12th & 13th of 2001, representatives of a variety of interested groups from the treatment and prevention fields came together at the Carter Center in Atlanta for a meeting to assess the progress made to date on the development and implementation of field-specific common performance measures. The meeting highlighted both the considerable progress and the shortcomings made to that point by a variety of groups, both public and private, in the development of empirical measures of access, appropriateness and outcomes of care.

Employing a process that alternated between working groups and plenary sessions, the meeting was successful in making explicit the tremendous overlap in both content and process that has guided the efforts of the various groups working in this field to date. It was the consensus of the attendees in the final plenary session that the forum constituted an important and worthwhile venue for the coordination and mutual support of the various individual efforts already underway in the field and that, in fact, it take on a unique and ongoing identity as the Forum on Performance Measures in Behavioral Healthcare and Related Service Systems. The Forum was given the charge by SAMHSA to become the vehicle for the coordination of these various efforts at identifying and specifying *field-level* (e.g., for adult substance abuse) common indicators and measures.

Why Common Indicators?

There has been tremendous recent growth in measurement of outcomes and other aspects of performance. Much of this work is of very high quality in technical, clinical, and policy terms, making use of rigorously developed tools to improve services and systems in relevant ways, and some of it also incorporating the important perspectives of those receiving care. This orientation to performance measurement is an unquestionable strength of the field. However, the diversity and resulting fragmentation has also limited the capacity of the field as a whole to speak with clarity and authority on the issue of quality.

Cost, rather than quality, often dominates behavioral healthcare resource decisions, in part because the field lacks consensus on how to demonstrate the quality of care, including the attainment of outcomes. Diversity in indicators and measures impedes comparability. There are significant areas of agreement and overlap in performance indicators, but without commonly used operational measures of these indicators the performance of organizations and systems of care cannot be effectively compared.

Quality is much more difficult to define and demonstrate than cost. In the absence of generally accepted evidence of quality, it is difficult to counter effectively the proposition that cheaper is better, and resources continue to drain from behavioral healthcare. To be significantly more effective in advancing quality as a legitimate, central factor in decisions, the field must work in concert to establish, accept, and implement methods for measuring quality.

In addition to strengthening the position of quality in behavioral healthcare policy decisions, adoption of common performance indicators and measures will:

- Generate compatible performance measurement efforts across all organizations to facilitate appropriate comparisons for accountability to consumers and purchasers, benchmarking for organizational quality improvement, more informed decision support for consumers and purchasers selecting treatment and/or health plans;

- Provide guidance on critically important dimensions of performance to those behavioral healthcare organizations that are in early stages of measuring performance;
- Encourage collaboration in data-sharing for benchmarking and quality improvement purposes;
- Reduce redundancy in requirements for performance data by accreditation, regulatory, and purchaser organizations, thereby increasing efficiency and reducing costs.

Why Common Measures?

In discussing performance measurement it is important to maintain the distinction between indicators and measures. An *indicator* is a quantitative specification, typically expressed as a ratio (e.g., percentage), of a selected aspect of performance. A *measure* represents the methodology for deriving and calculating quantitative results that may be used in an indicator. There may be many different ways to define and collect data to be used in calculating an indicator. Thus, indicator is a more general concept than measure, and this distinction is important to maintain in order to be clear about the potential for comparability in using indicators.

A call for common measurement cannot be taken lightly: the inertia of existing investments in practices and information systems creates a significant challenge to change in most settings. It would be far easier to achieve consensus about indicators, leaving to implementers the detailed decisions about how performance would actually be measured. Some earlier plans for shared indicator sets have stopped short of recommending common measures. Although this approach allows wide participation in a general report card framework by minimizing the amount of adaptation required, the difficulty is that results produced by different measures cannot be compared with adequate precision. Variation in instrumentation across settings or populations makes any comparisons between groups largely speculative; performance measurement systems using unique measures are inevitably local systems. The goal of the Forum is to establish consensus that will move the field toward a concise but truly national performance measurement capacity. Despite the challenges to implementing common measurement, therefore, the adoption of common measures is a crucial, corollary component of a common indicator strategy.

Mission Statement

The mission of the Forum is to improve the delivery of behavioral health treatment and prevention services by supporting the development and adoption of broadly applicable indicators and measures to assess organizational performance and consumer outcomes. These indicators and measures should be designed to serve the needs of both external accountability as well as internal quality improvement. The Forum provides an ongoing venue for collaboration, coordination, and communication between the various initiatives, both public and private, which are already working separately to measure

service access & delivery, quality, and outcomes. The Forum also fosters the sharing of information and experiences of provider, government, employer, consumer, and accreditation groups in implementing performance and outcome measurement practices and initiatives.

Functions

The functions of the Forum include the following:

- ***Identification & synthesis*** of issues which are faced by the field of performance measurement in common, irrespective of area of focus (e.g., a common vetting process for measure development; model database architectures: and cross-cultural implementation)
- ***Coordination and communication*** of efforts and progress among the various separate initiatives already working in the field
- ***Representation*** of the overall Forum and the field of performance measurement to the field of behavioral healthcare at large and to other audiences
- ***Focus*** of the overall efforts of the field on the successful implementation of performance measures within delivery systems, both public and private.

Initially, the Forum was intended to focus on common issues of the *process* of measures development...issues such as what constitutes an appropriate pilot test design, how to deal with rate-based measures in populations where the denominator is not clear, etc. However, over the last two years, as the work groups and their leaders have worked within the Forum, the goals have expanded to include consideration of not just common *process* issues but also common *content*. In fact, over the last two years, a small set of measures have been considered and adopted by all the treatment fields represented within the Forum – constituting in effect a set of common measures for behavioral healthcare. Within the Forum we sometimes call this initial set of common measures the “down-payment” set, since it represents both a significant advance for the field and, at the same time, a limited effort in terms of scope. In order to have been included in this initial set of common measures, any given measure needed to be approved as meaningful, measurable, and feasible by all the treatment workgroups for administrative measures or developed within the Modular Survey initiative for primary consumer data measures.

Administrative Data-Based Process of Care Measures

Within the initial set of measures are three originally developed by the Washington Circle, a group convened by the Center for Substance Abuse Treatment in March 1998. The initial focus of the Washington Circle’s efforts was the development of performance measures to promote accountability and improvement in the recognition and treatment of addiction at the level of delivery systems (i.e. managed care organizations, state Medicaid programs, etc.). As a strategic framework for its development work, the Washington Circle linked its development work to an evidence-

based and clinically appropriate overarching process of care defined by the domains of prevention/education, recognition, treatment, and maintenance.

Originally seven measures, some based on widely-available administrative data elements and some based on planned consumer surveys, were conceptualized across these four domains. Early specification and pilot-testing work focused on the administrative data-based measures and three quickly proved to be not only feasible and measurable but also quite meaningful in the field of adult substance abuse. Subsequent work within the Forum has established these measures as having similar characteristics across the other treatment fields (i.e., adolescent substance abuse, child and adolescent mental health, and adult mental health). The three measures are:

- ***Identification Rates***, defined as the number of cases per 1000 members who were diagnosed or who received treatment services for a range of specified diagnoses (the exact diagnoses depending, obviously, on the field being measured).
- ***Initiation of treatment services***, defined as the percentage of individuals with an index diagnosis within the appropriate range of diagnoses who receive at least one additional service within a specified time range (e.g., 14 days for adult substance abuse).
- ***Treatment engagement***, defined as the percentage of clients with an index diagnosis that receive additional services beyond that required for initiation within a specified time frame (e.g., at least 2 additional services within 30 days after initiating treatment for adult substance abuse).

At the 2nd Forum on Common Performance Measures meeting in mid-April 2004 at the Carter Center in Atlanta, the exact specifications for these measures (e.g., diagnostic ranges and time frames for the various covered populations) were distributed and the rationales and evidence base for the specifications were addressed. Since then empirical testing of both the feasibility and validity of the measures beyond the adult substance abuse field has been conducted and supports their utility as truly common measures in behavioral health care.

Consumer Survey-Based Perception of Care Measures

In addition to the administrative-based measures just described, the common measure set presented and discussed at the 2nd Forum meeting contained a set of measures derived from the collection of primary data from consumers about their perception of their experience of care. Regard for the consumer point of view has been long-established as an important policy direction within the fields of both adult and child/adolescent mental health, as evidenced by the prominence of a variety of nationwide survey-based initiatives such as the Mental Health Statistics Improvement Program (MHSIP) survey and the Experience of Care and Health Outcomes (ECHO™)

survey. The importance of the consumer point of view in evaluating the quality of care is increasing in the substance abuse field as well.

The Forum's Adult Mental Health Workgroup (AMHW) identified key concerns and indicators from the adult mental health consumer perspective and items that could measure them. Building on that work, representatives of the AMHW joined with representatives from the other Forum treatment workgroups and the MHSIP Report Card Workgroup, as well as outside experts, in the Modular Survey Initiative in April of 2003 with the goal of identifying a small set of concerns and related items that could be considered as meaningful across a broad age range of consumers within both mental health and substance abuse treatment settings.

The intent of the initiative was to generate broadly applicable item modules addressing perceptions of care along the domains of access, quality/appropriateness, and outcome/improvement. The Modular Survey is designed to be both applicable across respondent groups and service and payer settings and specific to each. This is accomplished by architecting and building a linked cascade of modules that moves from the most broadly applicable (i.e., common to consumers of all ages from mental health or substance abuse treatment settings) to relatively less common (i.e. for specific age groups and treatment settings such as adult mental health) to more respondent-specific (i.e. for consumers with serious mental illnesses or from inpatient psychiatric units only). It is not intended to be a comprehensive survey of the target domains, but rather to be comprised of item measures of concerns that meet the test of commonality at some level.

The work of the Modular Survey Initiative began with the identification of current thinking on the important domains and key consumer concerns. Two workgroups, one for Adults and one for Children/Adolescents, then equated the consumer concerns with relevant items, and grouped these items into modules. The items were derived from established national behavioral health consumer surveys (e.g., MHSIP, ECHO, YSS, etc.). Linkage to such existing and well-established survey measures is important not just for efficiency and effectiveness of item selection but also to provide the opportunity to potentially benchmark performance in common areas across behavioral health and healthcare consumer surveys.

The Carter Center Meeting – April 2004

In mid-April 2004 a meeting of over 150 individuals representing a wide variety of interests and stakeholders was held at the Carter Center in Atlanta. At this meeting both the details of this Forum process and its initial results were presented and discussed. In brief, the initiative identified through a consensus process 11 items at the highest level of commonality – in other words, items common for all age groups of consumers and for all treatment fields within behavioral healthcare. In addition, items at the next highest level of commonality – those common within but specific to either the adult population or the children/adolescent population – were also identified – five for adults and a separate five for children & adolescents. During the summer and fall of 2005 these items were pilot-tested in Cincinnati, using the United Way agencies that provide behavioral health

services as pilot sites. Over 1,000 individual respondents participated in the pilot test, representing a wide range of diagnoses and levels of severity across mental health and substance abuse. These results were analyzed in conjunction with data from an additional 20,000 respondents from the MHSIP initiative (many of the items in the Modular Survey are derived from MHSIP items) using IRT analyses conducted by Ann Doucette, Ph.D. As a result of the pilot testing, a number of recommendations were made by Dr. Doucette's technical group to the Modular Survey Steering Group, the coordinating body for the initiative. These included retaining only 12 items (seven for quality, five for perceived outcomes) and collapsing Levels 1 and 2 into a single level common to all ages and all fields.

Progress Since the Carter Center Meeting

At the close of the April 2004 meeting much work remained to be done on the proposed common measure set. For example, the detailed specifications for the administrative data-based measures of identification, initiation, and engagement proposed at the meeting for fields and populations beyond the original one of adult substance abuse have subsequently been extensively pilot tested. The specifications proposed at the meeting for adult mental health treatment have been found to be feasible and meaningful and at least one public entity is using them as core measures for its system transformation efforts under the President's New Freedom Commission report (Attachment 1). In a similar fashion, the joint specifications presented at the Carter Center by the representatives for child/adolescent mental health and substance abuse treatment were also found to be both measurable and meaningful (Attachment 2). For the last 3 years the substance abuse measures have been included in the reporting set of the National Committee on Quality Assurance's HEDIS measures for health plans. They have also been submitted to the National Quality Forum for endorsement as national standards of measurement. In addition, all three of the measures are being used by the Veterans Administration to evaluate the quality of substance abuse care at the facility level. Finally, since the Carter Center meeting, the measures, originally specified for health plans, have been re-specified for use in the public substance abuse treatment sector and have been found to be both feasible and of value. In fact, in a paper accepted for publication in the *Journal of Substance Abuse Treatment*, Deborah Garnick and her co-authors demonstrate that the process measures of initiation and engagement of treatment are predictive of outcomes in the criminal justice area (arrests and incarcerations), an important development for the ongoing policy discussion concerning the appropriate roles of process and outcome measures in performance measurement.

On a parallel path, great progress has been made on the pilot testing and continued improvement of the Modular Survey, a consumer perception of care measurement tool first presented at the Carter Center meeting. There, a total of 24 items chosen by a national consensus process involving over 50 stakeholders were discussed. Since the meeting extensive pilot testing and development has occurred on that 24 item set, which addressed the domains of access, quality/appropriateness, and perceived outcomes. A round of pilot testing involving primary data collection from a variety of mental health and substance abuse treatment programs affiliated with the United Way in

Cincinnati yielded approximately 1200 responses. In addition, secondary data from a variety of sources (if you remember, all the items presented at the Carter Center meeting were chosen from existing, non-proprietary surveys) was used, with the final pilot testing database having over 20,000 unique respondents. An IRT (Item Response Theory) analysis was conducted on this sample and, using the results of that analysis, the workgroups and the Steering Group chose a total of 11 items for all three domains to comprise the 1st phase of the Modular Survey. Interestingly enough, all 11 items worked equally well for all fields (mental health AND substance abuse treatment) and for all populations (children, adolescents, and adults) and therefore, in keeping with the expanded mission of the Forum, represent true common items for the behavioral healthcare field.

Subsequent to the pilot testing of the common items, the Center for Substance Abuse Treatment supported the development and pilot testing of a set of substance abuse treatment-only items. Since there existed no consumer perception of care survey for substance abuse treatment clients, a work group was established to develop new items. This development effort was co-ordinate with both the on-going Mental Health Statistics Improvement Program's (MHSIP) revision of its mental health treatment-only and the development of the National Outcomes Measures (NOMS), which began in the year after the Carter Center meeting. A total of 35 items addressing the areas of treatment relationship, self-recognition of problem, recovery/relapse prevention, and social connectedness were developed by the work group. These 35 items have now gone through two rounds of pilot testing in substance abuse treatment clients, one round focusing in adults and the second on adolescents. The data from both these rounds will be available for analysis this summer and in the fall final recommendations for items for the substance abuse treatment module will be made by the original work group.

Already, however, there is significant interest in some of the items from both the common module and the substance abuse treatment only module. As part of the development of NOMS within SAMHSA, a Technical Consulting Group was convened in 2006. This group examined a wide range of items for potential inclusion in the final NOMS set and finally a total of seventeen items addressing the areas of social connectedness and client perception of care were recommended for further consideration. Of these seventeen, eight were from the Modular Survey, four times as many items as from any other survey tool considered by the group.

In summary, then, the original work of the Forum on Performance Measures on a small set of common measures for the behavioral health care field continues to be developed and the choices of measures reinforced by further use and adoption. By providing an ongoing platform for discussion and sharing of ongoing, high quality work within the various fields, the Forum has progressed beyond consensus about domains and indicators to first, consensus about process, then to consensus about content, and now to consensus based on empirical evidence. In the future the Forum will continue to develop the evidence-base for the feasibility and utility of a small set of common measures for our field, with the goal of fostering both strategic accountability and the measurable improvement of the quality of care we provide to our clients.