

CCCMHA's Recommended Framework for Outcome and Performance Evaluation
for Adults and Older Adults with Severe and Persistent Mental Illness

There are two critical issues that must be addressed in any performance measurement system that attempts to evaluate provider effectiveness in serving adults and older adults with severe and persistent mental illnesses:

1. The impending integration of primary care and behavioral healthcare and the overall healthcare system's focus on the triple aim: Better quality (improved population health), better care (improved patient experience), and lower cost.
2. The importance of taking a holistic view of health which is quality of life and recovery-based rather than simply symptom reduction-focused.

Measuring effectiveness (quality) – The importance of a quality of life approach

In most discussions of the measurement of physical health outcomes, the focus is on the measurement of specific symptoms of the disease. For example, in the treatment of diabetes, the monitoring and control of the client's A1c level is regarded as the primary means of measuring the quality of care. Even on the behavioral health side, when attempts are made to measure the quality of care, the focus is usually on the monitoring and control of the symptoms of the disorder. For example, the IMPACT model for the treatment of depression measures at every session whether the symptoms of depression are improving or getting worse.

This tendency to see the person as the disease no doubt results at least partially from the fact that virtually all evidence-based practices, both in primary care and in behavioral healthcare, are *disease-specific*. While nobody would argue that it is unimportant to monitor the A1c level of a client with diabetes or the level of hopelessness of a client with depression, the question of how these symptoms relate to the overall quality of life of the client is rarely asked. While it is clear that having an A1c level that is within normal limits is preferable to having one outside of normal limits, how does accomplishing this result in meaningful improvement in the client's overall quality of life, including their physical and mental health? Does an improved A1c level result in a better quality of life or does a better quality of life result in a better A1c level? Or both? How are these quality of life indicators linked to the more traditional quality measures in primary and behavioral health care? Most primary care providers do not see these kinds of questions within their "scope of practice" and this has significant implications for the integration of primary and behavioral healthcare.

Most behavioral healthcare providers have long understood that the disease is not the person. Recovery and an overall improved quality of life, not the mitigation or elimination of symptoms, has long been the dominant paradigm in mental health and substance abuse services. Decreased symptoms and increased functioning, the traditional outcome domains of interest to the mental health system, are inadequate to describe what we mean by "recovery-oriented" outcomes because they are driven solely by the individual's illness. Since recovery implies the re-claiming of all aspects of one's life, "recovery-

oriented” outcomes should include all components of what would generally be considered to be a “meaningful” and “satisfying” life, *regardless of the presence or absence of a mental illness.*

Therefore, we must identify, define and measure the domains that make up our clients’ quality of life. The following is a possible list of quality of life domains that our behavioral health programs should measure:

1. Residential
2. Employment (includes volunteering)
3. Education
4. Financial (includes payee issues)
5. Independence (Includes conservatorship)
6. Physical Health
7. Substance Use / Abuse
8. Legal
9. Family / Intimate / Social relationships
10. Mental Health
11. Leisure / Recreation (includes hobbies, other group or solitary free time)
12. Spirituality

Because of the Mental Health Services Act, Full Service Partnership (FSP) programs are already required to collect and report on many of the domains included in this list. We have the infrastructure in place to measure and report on the quality of life of our clients in domains such as residential, employment, and educational status. We must make the argument that focusing on the symptoms of an illness, while important, will be insufficient to measure our impact on the overall health and quality of life of the people we serve. Through this we bring great value to the healthcare integration effort because we have both the expertise in recovery philosophy and the structural means to implement it. What is missing, however, is a system that enables the use of these data for quality improvement purposes. Our public mental health system needs the ability to compare results across programs and systems (counties).

The current FSP data collection and reporting (DCR) system is designed primarily as a longitudinal evaluation. The progression of individual clients can be documented over time, demonstrating trajectory with regard to critical indicators such as housing, employment, incarceration, and psychiatric hospitalization both prior to and during enrollment in Full Service Partnership programs.

Longitudinal measurement systems lend themselves to a pre-/post-treatment evaluation strategy. A pre-/post-treatment design allows one to describe an individual’s status in a variety of domains prior to participation in a treatment or intervention, and then again after participation in the intervention. The inherent problem with this in the current measurement process is that FSP programs are not characterized by a specific intervention or treatment model, but rather adherence to general principles and treatment components; thus, the post-treatment point is not clearly articulated or well-defined. By default, the current assessment process is essentially a pre-/post-enrollment design, as opposed to a pre-/post-treatment design. Furthermore, while enrollment in an FSP program is not expected to last

indefinitely, there is no expected or anticipated duration. Length of enrollment is likely to vary by a number of factors including age group, the target of the specific FSP program, characteristics of the enrollees, and so forth. In addition, there are a number of reasons that partners disenroll from FSP programs, which may or may not be related to the attainment of their articulated goals at the time of enrollment.

In addition to longitudinal analyses, the DCR data lend themselves to cross-sectional analysis. A cross-sectional evaluation strategy allows one to describe the characteristics and experiences of individuals being served at a single point in time. For example, counties as well as the state have the ability to describe characteristics and circumstances of partners at various time points, such as at the time of enrollment, in the year prior to enrollment, and within the first year after enrollment.

In this regard, probably the most important cross-sectional analysis available in the DCR data is the opportunity to observe whether there are disparities in our services based on such factors as gender and ethnicity. For example, it is a relatively easy matter to determine whether male or female participants in FSPs have different rates of independent living or employment. Similarly, it is possible to determine if FSP participants of different ethnic groups or backgrounds have higher or lower rates of hospitalization or incarceration. While one should not assume that all differences found in outcomes between genders or ethnic groups are the result of disparities in treatment and services, identifying if differences in outcomes even exist would seem to be the logical first step in determining whether there are also disparities in the services that FSP participants receive.

Regardless of the evaluation strategy, the interpretation of FSP outcome data raises a number of challenges. There are currently no criteria against which FSP outcomes are evaluated; in other words, there are no guidelines or parameters to suggest the extent to which an FSP program is positively impacting the lives of partners. Two different comparison strategies could be considered to interpret FSP outcomes: *referenced-benchmarking* and *criterion-reference*. A referenced-benchmarking approach would compare an individual FSP program's outcome to the average of that same outcome across the population of FSP programs. Using this approach, each program's success is evaluated against the success of the whole. Another comparison strategy would be to employ a criterion-referenced approach. A criterion-referenced approach would compare an individual FSP program's outcome to a pre-determined standard. Using this strategy, each program's success is evaluated against the level of change expected in a particular indicator based on theoretical and/or empirical knowledge of expected or intended improvement.

Regardless of the point of comparison, peer-referenced or criterion-referenced, outcomes have to be understood not only within the context of the population being served, but also within the context of the specific services and supports being provided. There is a link, implicit or explicit, between the needs of partners, the services that are provided, and the outcomes achieved. In addition, knowledge of the specific characteristics of the partners enrolled and the services they are receiving is critical to interpreting program outcomes. This information would allow for similar FSP programs to be grouped together in employing either a peer-referenced benchmarking or criterion-referenced comparison approach to interpreting the observed outcomes.

Another critical element of measurement that is missing from the existing FSP outcomes and data collection strategy is the articulation of FSP intervention model(s). It is important to associate program performance measurement and outcome evaluation methods with service delivery approaches and the specific types of mental/behavioral health interventions that are provided. Given the diversity of populations served and intervention strategies employed, within and across age group-specific FSPs, there is a need to articulate the key elements of FSP programs.

Given the preceding discussion, we are making four recommendations regarding the FSP program performance measurement system.

The first three recommendations relate to maximizing the use of the current FSP outcome measurement and data collection system. First, the accuracy of the data should be assured. Only by knowing the validity of data entered into the DCR can there be any confidence in the validity of the outcomes reported. This recommendation relates not only to the accuracy of the data entered, but also to the extent to which complete data are entered for any given partner.

Second, counties should be trained more extensively on how to use their DCR data downloads. The current training available through CDMH assists counties in obtaining their FSP data, but there is little to no guidance on what to do with the data once received. Counties would benefit greatly from training in how to analyze and report FSP data downloaded from the DCR system.

Third, efforts should be made to clearly articulate the critical components of specific FSP programs, and data elements that define these components should be entered into the DCR system. Currently, each FSP program has up to three variable fields for county-specific data on both the KET and the 3M forms. These variable fields could be used to code elements of FSP program service approach and delivery for each county, and would assist in the interpretation of observed outcomes.

Each of the ways in which the current FSP outcome measurement and data collection system could be maximized would require a resource investment. It is recommended that the Behavioral Health Division of the Department of Healthcare Services create a stakeholder group with the expertise and authority to provide direction and oversight to these tasks and their necessary funding. This group would be charged with creating and maintaining a system for ongoing review of the performance of FSP programs, ensuring diverse stakeholder input.

Our specific recommendations are as follows:

Recommendation #1: Take steps, including providing adequate resources, to ensure the accuracy and reliability of data entered into the Data Collection & Reporting System.

Recommendation #2: Take steps, including providing adequate resources, to assist counties in cleaning, analyzing, and generating reports on the Full Service Partnership data they download from the Data Collection & Reporting System.

Recommendation #3: Take steps, including providing adequate resources, to assist counties in articulating the key elements of their Full Service Partnership programs so that appropriate data reflecting these elements can be entered into the Data Collection & Reporting System.

Recommendation #4: Establish a new advisory stakeholder group with the expertise and authority to serve as the statewide body charged with ongoing direction and oversight of the Full Service Partnership (FSP) performance measurement system. This group will create and maintain a process for the ongoing review of the performance of FSP programs, including issues related to the specific age groups (children, transition age youth, adults and older adults). It will also determine the feasibility of implementing a statewide “level of care assignment system” that matches client’s level of need with the appropriate level of care.

Recommendation #5: Charge the advisory stakeholder group to investigate which FSP program performance indicators and outcomes measures have application to programs serving non-FSP clients and begin the process of identifying indicators and outcome measures that are appropriate for non-FSP clients while paying attention to alignment with new federal and state healthcare reform requirements.

The place of recovery in performance measurement

While it could certainly be argued that “recovery” could be defined as the adequate attainment of the quality of life domains currently measured in the DCR, it should be recognized that certain of those domains are less robustly measured than others and some (e.g., spirituality) are not currently measured at all. SAMHSA has recently released a definition of recovery that consists of four domains:

- **Health:** overcoming or managing one’s disease(s) as well as living in a physically and emotionally healthy way;
- **Home:** a stable and safe place to live;
- **Purpose:** meaningful daily activities, such as a job, school, volunteerism, family caretaking, or creative endeavors, and the independence, income and resources to participate in society; and
- **Community:** relationships and social networks that provide support, friendship, love, and hope.

While the DCR is fairly good at measuring the domains of health and purpose, it is less adequate when it comes to measuring the domains of health and community. There are many outcome measures available that measure these other domains, such as the Milestones of Recovery Scale (MORS) and the Level of Care Utilization System (LOCUS).

Recommendation #6: Require that FSPs employ a standardized measure of recovery that addresses the SAMHSA domains (particularly Health and Community) and require that programs administer the measure on at least a quarterly basis.

Measuring the perception of care – Improving the client experience

Our current system is lacking a meaningful measurement feedback system that informs us of our clients' perceptions of their care. A Measurement Feedback System (MFS) requires frequent periodic assessment of client perceptions of the care they receive as well as their perceptions of the outcomes of that care. Generally, the practices being evaluated are not population-specific and are intended to assess the overall quality of the therapeutic relationship and/or the organizational culture.

Currently, the only statewide effort to examine program and/or system practices is the annual Mental Health Statistical Improvement Program Survey (MHSIP). Unfortunately, the MHSIP suffers from several shortcomings:

1. Insufficient frequency of administration: Because the MHSIP is only administered annually, too much time elapses before measurement again takes place. Therefore, even if quality improvement efforts are put in place based on the MHSIP results, there is no means of determining if the efforts are having the intended effect. Indeed, given the long period between administrations, it is possible that QI efforts might have an effect and then revert to prior performance levels due to a lack of timely feedback to programs and systems that would allow them to assess whether their outcomes and/or practices have improved.
2. Lack of a feedback loop to direct service staff: The lowest level of aggregation at which data are provided is at the overall provider agency level. This makes it impossible for an agency to determine even whether there are performance differences in their programs, let alone among their direct service staff.
3. Lack of a feedback loop to clients: While there are significant questions regarding whether the data collected in the MHSIP would be useful or informative to clients, there is no attempt to provide clients with any useful feedback based on their responses.
4. No connection between practices and outcomes: Currently there are no efforts to determine whether there are actually any relationships between the practices and outcomes identified on the MHSIP and any other programs or system outcomes.

Recommendation #7: Consideration should be given to creating a measurement feedback system that corrects these shortcomings and enables FSPs to incorporate client perceptions of their care into a comprehensive performance evaluation framework.