

The MHSIP Quality Report Toolkit



May 2004

Prepared by:

The Mental Health Statistics

Improvement Program (MHSIP)

Quality Report Workgroup



Human Services Research Institute

2269 Massachusetts Avenue

Cambridge, MA 02140

www.tecathsri.org



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

Substance Abuse and Mental Health Services Administration

Center for Mental Health Services

www.samhsa.gov

The Mental Health Statistics Improvement Program (MHSIP) Quality Report Toolkit was developed by the MHSIP Quality Report Workgroup with support from the Evaluation Center@HSRI, under the direction of Dow A. Wieman, Ph.D., Senior Research Associate, Human Services Research Institute, Cambridge, MA with the assistance of Neera Jain, Research Assistant, Human Services Research Institute.

The *Evaluation Center@HSRI* is a technical assistance center funded by the Substance Abuse and Mental Health Services Administration (SAMHSA) Center for Mental Health Services (CMHS) and operated by the Human Services Research Institute (HSRI).

MHSIP is a program of CMHS designed to improve the quality of mental health program and services delivery decision making at all levels of government through guidance and technical assistance on the design, structure, content, and use of mental health information systems.

The MHSIP Quality Report Workgroup:

Preface	1
<i>What is the MHSIP Quality Report Toolkit?</i>	1
<i>Phases of a Quality Report project</i>	1
<i>A Note on Terminology</i>	1
Phase I. How Do I Plan a Quality Report Project?	3
<i>Step 1. Assess Organizational Readiness</i>	3
<i>Step 2. Plan for Transition from Version 1.0 (if applicable)</i>	3
<i>Step 3. Develop a Plan For Internal Marketing</i>	4
<i>Step 4. Establish a Performance Measurement Steering Committee</i>	6
<i>Step 5. Determine Steering Committee Functions</i>	6
<i>Step 6. Decide on the Composition of the Steering Committee</i>	8
<i>Step 7. Determine Steering Committee Structure</i>	9
<i>Step 8. Address Human Subjects Concerns and Regulations</i>	10
Phase II. How do I Implement a MHSIP Quality Report Project?	15
<i>Step 1. Determine Schedule of Administrative and Survey Data Collection</i>	15
<i>Step 2. Assess Capability for Collecting Administrative Data</i>	16
<i>Step 3. Determine Method of Consumer Survey Administration</i>	16
<i>Step 4. Decide on Data Collectors/Interviewers</i>	20
<i>Step 5. Establish Procedures for Obtaining Consent</i>	21
<i>Step 6. Establish Policy on Inducements and Incentives</i>	22
<i>Step 7. Decide on Sampling Methodology</i>	22
<i>Step 8. Determine Sample Size</i>	24
<i>Quality Report Recommendations for Survey Implementation</i>	27
<i>Data Management</i>	28
<i>Step 9. Designate a Data Manager</i>	28

<i>Step 10. Establish Data Security Procedures</i>	28
<i>Step 11. Create data tracking logs</i>	29
<i>Step 12. Establish Data Entry Procedures</i>	30
<i>Step 13. Establish Data Quality Controls</i>	32
<i>Step 14. Address Multi-level Data Management Issues</i>	33
<i>Step 15. Establish Procedures for Dealing with Missing Data</i>	33
Phase III. How do I interpret Quality Report Results?	35
<i>Step 1. Assess and Account for Bias</i>	35
<i>Step 2. Decide on Approach to Comparison</i>	36
<i>Step 3. Address Psychometric Issues</i>	38
<i>Validity and reliability</i>	38
<i>Step 4. Control For Differences In Populations Served (Case-Mix Adjustment)</i>	39
Phase IV. How do I Use and Report MHSIP Quality Report Information? .	41
<i>Step 1. Determine the Purpose of your Quality Report Project</i>	41
<i>Step 2. Establish a Planning Process for the Report</i>	42
<i>Step 3. Identify and Engage Your User Groups</i>	43
<i>Step 4. Determine the Content of Your Report</i>	44
<i>Step 5. Establish Standards for Reporting</i>	44
<i>Step 6. Determine the Format for Presentations</i>	46
<i>Data Representation</i>	46
<i>Step 7. Decide on Time and Frequency of Reporting</i>	54
<i>Step 8. Market Your Results</i>	54
<i>Quality Report Recommendations for Reporting Results</i>	56
Phase V. How may I Modify and Enhance the MHSIP Quality Report?	57
<i>Step 1. Consider Issues Involved In Modification of The Quality Report</i>	57

<i>Step 2. Decide on Enhancements to the Quality Report.....</i>	<i>57</i>
Phase VI. How do I Evaluate and Improve a MHSIP Quality Report Project?	59
.....	
<i>The Process Evaluation</i>	<i>59</i>
<i>Step 1. Define The Purpose Of The Evaluation.....</i>	<i>59</i>
<i>Step 2. Define the Evaluation Questions.</i>	<i>60</i>
<i>Step 3. Assess the Quality Report Implementation Process</i>	<i>60</i>
<i>Step 4. Assess the Quality of the Data Collected</i>	<i>61</i>
<i>Step 5. Assess the Cost/Burden of the Performance Measurement System.....</i>	<i>61</i>
<i>The Impact Evaluation</i>	<i>62</i>
<i>Step 6. Define The Purpose Of The Impact Evaluation.</i>	<i>62</i>
<i>Step 7. Define the Impact Evaluation Questions</i>	<i>62</i>
<i>Methods for evaluating a performance measurement system.....</i>	<i>64</i>
<i>Quality Report Recommendations for Evaluating the Performance Measurement Product</i>	<i>65</i>
Appendix I. Links for Additional Resources	66
<i>Related initiatives and organizations</i>	<i>66</i>
<i>Methodology</i>	<i>66</i>
<i>Standards for Publicly Reporting Survey Methods:</i>	<i>66</i>
<i>Guidelines for designing health care quality reports.....</i>	<i>66</i>
<i>State MHSIP Report Card Reports On-line.....</i>	<i>66</i>
Appendix II: Data Matching.....	68

Preface

WHAT IS THE MHSIP QUALITY REPORT TOOLKIT?

The purpose of this Toolkit is to provide guidance for those who have chosen to use the Mental Health Statistics Improvement Program (MHSIP) Quality Report to assess and report on the quality of behavioral health service systems in which they have an interest.

The Quality Report is a modular performance measurement system designed to assess and report on the quality and efficiency of mental health services. Along with a set of **universal indicators** applicable to all persons receiving mental health services in any treatment setting, the Report Card consists of modules with **indicators for specific populations and treatment settings**. Specific populations are adults with serious mental illness (SMI) and children with serious emotional disturbance (SED). Specific treatment settings are: 1) Comprehensive behavioral health systems (e.g. managed behavioral health organizations); 2) Hospitals; 3) Residential programs; and 4) Outpatient (i.e. ambulatory) treatment. For more information on the Quality Report see the Final Report.

The Toolkit represents the Workgroup's step-by-step recommendations for how to conduct a Quality Report project. Thus, the Toolkit is organized as a set of user questions regarding each phase of a Quality Report project .

PHASES OF A QUALITY REPORT PROJECT

Phase I: How do I plan a MHSIP Quality Report project?

Phase II: How do I implement the Quality Report?

Phase III: How do I interpret and use the results of the Quality Report?

Phase IV: How do I report the results of a Quality Report project?

Phase V: How may I modify and enhance the Quality Report?

Phase VI: How do I evaluate a Quality Report project?

Within each phase, the Toolkit addresses the sequence of steps required to complete that phase.

A NOTE ON TERMINOLOGY

Consumers. The Toolkit throughout usually refers to persons who use mental health services as “consumers.” Occasionally, based on context, the Toolkit uses other terms such as “patient” or “client.” These choices are used interchangeably and are not intended to reflect differences in meaning.

Ownership. In referring to behavioral health organizations, the Toolkit sometimes distinguishes between “public” and “private” sector systems. Given the complexity of the financing and organization of mental health services today, this distinction is somewhat of an oversimplification, but for practical purposes such as this toolkit the difference is clear from the context, for example a discussion of the stakeholder role of state mental health commissioners. Most of the discussion applies to either type, however, in which case we use the terms

“mental health system” or “organization” to denote a formal organization with recognized financial and management control over a more or less comprehensive system of behavioral health services. In some instances, we differentiate between an organization of this type and “providers” to denote individuals or groups who provide services under the direction (e.g. through subcontracts) of the organization.

Versions. As discussed in the Final Report, the name of the revised MHSIP Quality Report represents a slight modification of the title of the first version, known as the “MHSIP Report Card” (or sometimes the “Consumer-Oriented Report Card). For purposes of clarity and consistency, this Toolkit refers to the earlier version as “the Version 1 MHSIP Report Card” and to the current revised version as “the MHSIP Quality Report,” or simply “the Quality Report.” Similarly the Toolkit differentiates between the groups responsible for development of the first and second versions as, respectively, the “MHSIP Report Card Taskforce” and the “MHSIP Quality Report Workgroup.”

Phase I. How Do I Plan a Quality Report Project?

PHASE I STEPS

- Step 1. Assess Organizational Readiness
- Step 2. Plan for Transition from Version 1.0 (if applicable)
- Step 3. Develop a Plan For Internal Marketing.
- Step 4. Establish a Performance Measurement Steering Committee.
- Step 5. Determine Steering Committee Functions
- Step 6. Decide on the Composition of the Steering Committee
- Step 7. Determine Steering Committee Structure
- Step 8. Address Human Subjects Concerns and Regulations

This section of the toolkit covers the organizational preparations required to implement the MHSIP Quality Report. The steps identified here apply to organizations implementing a performance measurement system for the first time. Many organizations implementing the MHSIP Quality Report, however, will be familiar with this or some other performance measurement activity, and may have much of the necessary infrastructure in place. This may simplify the planning and implementation processes to some extent, but they will still need to anticipate and plan for factors related to systems change, and in fact, transitioning from an existing system (as opposed to starting fresh) raises some special issues, as discussed below.

Additionally, organizations transitioning from established systems may find this is an opportunity to review and improve upon their performance measurement infrastructure in accordance with the following recommendations. Consequently, we recommend that all organizations, regardless of their experience with performance measurement, at least review all of the following steps.

Step 1 Assess Organizational Readiness Those responsible for planning and implementing the Quality Report should recognize and acknowledge that the purpose of the initiative is system change. The project will significantly influence the operations of a mental health care organization, and those in charge of the process should anticipate all the apprehensions about change, and the obstacles that arise in all types of organization and plan accordingly. A body of research on assessing organizational readiness for change is now available, (Backer 1995), and you may find that a review of this literature will help you to anticipate obstacles.

The process of preparing for a Quality Report project process requires early and careful attention to marketing the concept within the organization, enlisting key sources of support, organizing a Steering Committee, and settling on its functions, structures, and composition. These issues are addressed in subsequent steps.

Step 2 Plan for Transition from Version 1.0 (if applicable) Systems with well-developed performance measurement and reporting programs will have already addressed many of the issues described below. The

introduction of new measures, however, are likely to raise issues about organizational change similar to those that may have occurred with the system was first introduced. Recovery and cultural competence measures in the MHSIP Quality Report may create uncertainty for organizations that performed well as measured by the Version 1 Toolkit or some other system. Also, modification of the performance measurement system may cause concern about the burden of change among those accustomed to established methods of collecting and analyzing information. Finally, when a performance measure system has become successfully integrated into an organizational culture, individuals or groups may have come to depend on the information it provides. When changes are made, they may be concerned that they will lose important tools for carrying out their functions.

Accordingly, even if your organization is familiar with the Version 1 Report Card or some other approach to performance measurement, you should assess the need to pave the way for implementation of the new Quality Report.

Consider Parallel Measures During Transition. In developing the Quality Report, the Workgroup sought to introduce new measures called for in the field, but also to preserve those with demonstrated utility. Consequently, there is a good deal of continuity between the first and the second. We suggest that organization transitioning from the Version 1 Report Card to the Quality Report do so by overlapping the two sets of measures for one or two years, i.e. add the measures developed for the Quality Report while retaining for a measurement cycle or two those in Version 1 that have been dropped.

This serves several purposes. First, it provides more continuity for benchmarking purposes, both internally for comparisons over time, and externally for comparing with other organizations that may have a different timetable for adopting the new Quality Report. Second, it reduces the apprehension about organizational change described above by introducing innovation within the context of the familiar and customary practice. Finally, it sequences the steps required to modify data systems over the transition period, as new measures are added in one measurement cycle, the others are dropped in subsequent cycles.

Step 3 Develop a Plan For Internal Marketing. The value of performance measurement and reporting is not readily apparent to all mental health system stakeholders, and may in fact, create apprehension among many. In addition, given the costs of these activities, they must compete with other priorities for resources. Thus, you should make a committed effort to “sell” the idea of performance measurement to all levels of the organization. To the extent that you gain broad acceptance and support, or at least reduce misunderstanding and apprehension, resources expended in planning and technical work will be used more efficiently and effectively.

One approach that has developed in the public health field in recent years is the concept of “social marketing” (Center for Medicare Education 2002). This is a set of principles and techniques adopted from commercial marketing, except instead of getting “customers” to buy a product, the goal is to persuade them to accept a new practice or adopt new behaviors. The social marketing approach is discussed in more detail in Phase IV (How Do I Report and Use Quality Report Information?)

Engage Critical Supporters: Ultimately, the success of your performance measurement system will depend on the support and cooperation of a long list of stakeholders. However, there is a "short list" who are critical to this effort. These are individuals who control resources, provide leadership and confer credibility within the system. In the parlance of organizational change theory, they are "opinion leaders." Obtaining their active support will greatly enhance your capacity to move the performance measurement system forward, and failing to achieve it may eventually result in derailing the project entirely.

In organizations where the Version 1 Report Card, or another performance measurement system is already established, their ongoing participation will not be so critical as it was initially. You should, however, re-establish connection early on in the change process, insure they are informed and address any questions they have. If ongoing participation is not indicated, and you should at least obtain their endorsement of the change.

A more detailed discussion of stakeholders in the performance measurement process is presented below in the section on the Steering Committee. The following is a brief description of key individuals whose support you should gain at the very earliest stage of planning.

Top-level administrators. These are the individuals with the authority to allocate the existing resources (both dollars and staff) that are needed for the performance measurement system, and they can provide the leadership necessary to obtain support from other stakeholders. In public sector systems, these individuals might be the state Commissioner of Mental Health; in private systems it might be Chief Executive or Operating Officers.

Persons in these positions are often very receptive to performance measurement systems and initiatives to improve them because such systems provide information they need to effectively manage the operation of the system. In public systems, this information helps them in advocating for additional resources to enhance services.

Consumers. The meaningful participation of consumers insures that the project will maintain focus on final objectives and identify obstacles that might otherwise divert the process from these objectives. In addition, without meaningful stakeholder involvement, particularly involvement of consumers, the project will lack credibility among other critical sources of commitment. Without this type of credibility, you will be much less likely to garner support from other system stakeholders.

Service providers. In most performance measurement systems, providers are important in supplying data in the form of clinician ratings, administration of surveys, and/or in keeping high quality records (e.g., claims) that can then be used for performance measurement purposes. Since the quality of data is of paramount importance and collecting high quality data will burden providers to some extent, it is important to obtain the commitment of providers and to ideally establish collaborative relationships with them.

State legislators. Support from members of the state legislature can be critical, especially if additional resources are needed to implement the performance measurement system in public sector systems. Their support is also necessary when implementation requires changes in state law (e.g., to permit sharing data across state agencies or between local providers and the state or county mental health agency). State

legislators are typically receptive to performance measurement systems because they are viewed as a means of increasing accountability for the taxpayer dollars they allocate to fund services.

Step 4 Establish a Performance Measurement Steering Committee. To draw on the full range of expertise needed to design and implement an effective performance measurement system and to develop "buy-in" from those who must support this effort, you should establish a Steering Committee that includes representatives from all major stakeholders. This committee should be established early in the process to ensure stakeholder input on all critical decisions. The Steering Committee will also help to prevent inaccurate information from being spread among key stakeholder groups that may increase resistance to your efforts.

You will need to make several decisions regarding the Steering Committee: 1) the role of the committee; 2) its composition; and 3) its organizational structure. Each of these is discussed below.

Step 5 Determine Steering Committee Functions. The Steering Committee may perform a variety of functions in the development and implementation of your performance measurement system. Several recommended roles are described below, but the adoption of other roles will be influenced by the characteristics and structure of your organization and the desires of the Steering Committee itself.

Advisory vs. decision making function. The first decision to be made regarding the Steering Committee is whether it is to be "advisory" or if it is to be empowered to make decisions. There are advantages and disadvantages to each role, and authority may be given to the committee to make certain decisions and not others. The approach that you take will be determined by a number of factors including, in the public sector, the relationship between the state mental health agency (SMHA) and various stakeholders (e.g., does the SMHA contract for community services or does it directly operate local service providers), the current political climate, and the leadership style of the SMHA and its commissioner. Similarly in the private sector, decision making power should reflect the management structure and style and other organizational characteristics such as the relationship to a larger health care entity.

Generally, providing the Steering Committee with as much decision making authority as possible will result in greater commitment to the success of the system, greater participation, and more willingness to help "sell" the system to others. In any case, it is most important that this aspect of the role of the Steering Committee be stated clearly when making invitations to serve on the committee. A lack of clarity about limitations in the authority of the committee will lead to frustration and conflict.

Liaison to key stakeholders. Perhaps most importantly, Steering Committee members should serve as liaisons between the performance measurement project and key stakeholder groups. In this role, committee members should represent their constituencies at committee meetings, assuring that their unique perspectives are represented in all discussions and decisions. Similarly, committee members must be responsible for sharing the results of meetings and the rationale for particular decisions with their constituencies—they must keep them informed. This latter role is particularly important in preventing inaccurate information and misperceptions, which would complicate the task of generating support for your effort. Requesting Steering Committee

members to report to the Committee names and affiliations of people with whom they have discussed the project is one way to reinforce the liaison function.

Setting purpose and goals. As described in the next section, performance measurement systems may serve several purposes (e.g., accountability, quality improvement, and management), which will largely determine its design and other major decisions. It is therefore important that the Steering Committee have opportunity to provide input in defining the purpose(s) of the system. Even when the primary purpose of the performance measurement system is externally mandated (e.g., a legislative mandate for performance-based accountability), the Steering Committee may want to identify other purposes the system is to serve.

Designing the performance measurement system. Designing a performance measurement system is a highly complex task requiring a large number of technical decisions regarding such issues as protocols for data collection, management, and analysis; selection of standardized instruments; and development of software applications. How you choose to involve the Steering Committee in these activities will have to be determined. For example, you may want to establish technical sub-committees or work groups to address these activities, or you may want your organization's technical staff to do the design work and submit it to the Steering Committee for review and comment.

Using sub-committees for the technical tasks works well when at least several members have the needed skills and you have more time to design your system. When neither of these is true, you should probably have your organization's evaluation and data processing staff assume major responsibility for designing the performance measurement system and submit draft materials to the Steering Committee. Even in this role, the Steering Committee can be invaluable in identifying practical and political problems that must be addressed.

Patient safety and privacy issues. As discussed below, performance measurement and quality improvement activities can lead to complex legal and ethical issues related to the uncertainty of differences between these activities and research involving human subjects. These issues are especially sensitive in the behavioral health field, due to the reality of stigmatization. One function of the Steering Committee therefore is to ensure that these considerations have received sufficient attention.

Marketing. Throughout the development process, but particularly when implementation begins, the Steering Committee should play an active role in "selling" the performance measurement system to their constituencies and others. Many stakeholders (e.g., associations of private providers and consumer and family member organizations) also have political ties and will be free to lobby state legislators for funds to support the system. When committee members feel that they had a hand in shaping the system and now have a sense of ownership, they will be more committed to promoting it among their constituencies. This activity is described in more detail in Chapter 5: How do I Use and Report Quality Report Information?

Report Design. As discussed in Phase IV, planning for reporting the results of the Quality Report should begin early in the project. You may wish to designate a taskforce within the Steering Committee for this purpose.

Evaluation. This is another component that should be planned and initiated early on. Phase VI describes the evaluation process.

Step 6 Decide on the Composition of the Steering Committee Having determined the major functions to be performed by the Steering Committee, you are now in position to make decisions about the membership. Given the functions described above and the fact that performance measurement systems may have far-reaching impacts on the policies and operation of the organization, the Steering Committee should consist of more than just technical staff. The governing principle is usually to insure that all major stakeholder groups are represented. Although the level of participation will probably differ from one person to another, it is usually important to include policy makers and those who control the resources needed to make the performance measurement system work, those who must implement components of the system, those who will be directly affected by the system, and those who have the technical skills required to design an effective system. Given this, listed below are the key stakeholder groups who should be represented on the Steering Committee:

- Consumer organizations
- Family member organizations
- Provider organizations (public and private)
- Management
- Commissioner and management staff (public-sector)
- COO, medical director (private sector)
- Affiliated organizations, e.g.:
 - State Medicaid agency
 - Health care system administrators (integrated systems)
- Technical staff (in house or consultant)
 - Research and evaluation design
 - Statistical analysis including sampling
 - Data processing/information technology

Other stakeholders. Depending on the organization and structure of your mental health system and state politics, you may also want to include representation from several other stakeholder groups. For example, in county-based systems in which public providers are part of local government, you may want to include local government officials. In states where legislative support is critical to the success of the system (e.g., new funds are needed to support its operation), strong consideration should be given to including a key legislator or legislative staff members. Where the performance measurement system will result in new requirements for managed care organizations, these should be represented. When a private organization operates under contract with the state, appropriate agency officials should be included. And, finally, if it is expected that other state agencies (e.g., Medicaid, education, criminal justice) will serve as the source of performance data, you may want to consider inviting their participation as well (See Appendix II: Data Matching).

Diverse attitudes. Finally, you should carefully consider selection of particular individuals to represent the various stakeholder groups. While it is important to have a committee whose membership generally supports

the goal of establishing a performance measurement system, selecting only enthusiastic advocates of performance measurement has some disadvantages. First, it is important to hear the views and concerns of those who do not entirely support such efforts. Such individuals are likely to raise important questions that must be addressed if the system is to be effective—questions that supporters might not think to raise. Second, when the Committee includes only persons who support the performance measurement system, their shared enthusiasm can result in a system that is unrealistically large and complex. Therefore, it is advisable to include some skeptics on the Steering Committee who will add an element of realism to the endeavor. If their concerns can be effectively addressed, it is more likely that the system will be successful when implemented in the field.

Consumer participation. The issue of consumer representation in performance measurement projects requires special mention. Version 1 of the MHSIP Report Card set a precedent in its emphasis on being “consumer-oriented”. That is, it was values-based in addressing specific concerns about how services benefit and affect consumers, and the choice of measures was influenced by extensive consumer input. Since the appearance of Version 1, the importance of consumer participation in performance measurement and quality assessment has become widely recognized throughout the health care field. The Institute of Medicine, in its seminal report *Crossing the Quality Chasm* (2001), identifies as one of the six major aims for improvement, a system that is “patient-centered” i.e. that focuses on patients’ expressed needs and desired outcomes. In mental health care, the recovery movement has become a major influence as expressed the concept of “system-transformation” espoused by the President’s New Freedom Commission report (*New Freedom Commission on Mental Health* 2003). The Final Report for the MHSIP Quality Report describes the way in which these trends are reflected in the new version, beyond that which was already present in Version 1.

As in the development of measurement systems, consumer/patient participation is equally important in the implementation process. Moreover, it is important that this participation be *meaningful*. That is, membership should include a representative selection of persons who use the organization’s mental health services, and consumer members should have equal rights in any consensus process. The organization may consider paying consumers for their contribution.

Finally, consumer representation is especially important in the Steering Committee’s function, described above, of insuring the safety and privacy of consumers participating in the performance measurement project.

These principles will not only promote the credibility of the performance measurement system, but they will also enhance its utility, for instance in identifying effective methods of encouraging survey response and of reporting results.

Step 7 Determine Steering Committee Structure Once you have decided on the functions and

membership of the Steering Committee, you will need to determine how the committee will be structured. Generally, given the need to include policy makers and advocates, you will find that the full Steering Committee is not the place for detailed discussions about technical issues related to measurement, data processing, and so forth. Such discussions in the full committee are likely to discourage continued participation by many members. Therefore, it is usually preferable to establish one or more technical work groups under the Steering Committee to do the detail work of reviewing and selecting measures, designing data collection protocols, etc. Representation from multiple stakeholders on these work groups is still

desirable whenever possible, but these may be technical staff from the stakeholder groups rather than the policy-level staff who serve on the Steering Committee.

Given a structure involving technical work groups, the Steering Committee would establish overall direction for the effort, define principles for the development of the system (e.g., stakeholder involvement and consensus), and establish policies for the system (e.g., policies regarding the use of performance data and the protection of consumer confidentiality). The technical work groups would ensure that the system produces reliable and valid information that addresses the concerns and priorities identified by the Steering Committee.

An active Steering Committee that represents all key stakeholders is critical to the success of your performance measurement system. Their input will result in a higher quality system having stood the test of addressing the concerns raised by persons with different perspectives. In addition, their active support will be a major factor in moving the system off the drawing board and into the field.

Step 8 Address Human Subjects Concerns and Regulations The distinction between what constitutes research (thereby requiring review of protections for human subjects) and quality improvement (which presumably does not require review) has never been entirely clear (Lynn 2004). The issue has become even less clear over the past few years, as federal and state agencies and academic institutions have become increasingly diligent in their oversight activities in this area. Further, the introduction of the Health Insurance Portability and Accountability Act (HIPAA), beginning in 1996, has added additional levels of complexity.

The federal government has provided guidance on the implications of HIPAA standards for research (available at <http://privacyruleandresearch.nih.gov>). It is unclear, however, to what extent these apply to performance measurement and quality improvement activities. Presumably the difference is between individual health information that is used for internal operational purposes and that which is published in journals, etc. The NIH defines research for HIPAA purposes (described below) as “a systematic investigation...designed to develop or contribute to generalizable knowledge” (http://privacyruleandresearch.nih.gov/pr_04.asp). But what of situations where the information from performance management and quality improvement is reported to the public, as with the Quality Report—is informed consent and other protections required for projects of this type? Also, performance measurement and quality improvement activities can pose hypothetical risks to patients, thereby indicating a need for some form of oversight.

We do not know of any single source that offers, for compliance purposes, a concise and concrete distinction between performance measurement and quality improvement on the one hand, and research on the other. Nor do we suggest that it would be productive for you to attempt to become expert in this vast and ever-changing area of policy. Our recommendation is that, early on in your Quality Report project, you consult with your local Institutional Review Board (IRB) or in non-academic settings, HIPAA-mandated Privacy Board, to obtain guidance on this issue.

Finally, whether or not a review is legally required for collecting, managing, using and reporting health information, you should give careful consideration to all aspects of patient safety and privacy of personal health information collected. As noted in the discussion of the Steering Committee’s functions, this is especially important in the behavioral health field, where stigma continues to be a major reality in people’s lives.

The following comments on HIPAA, IRBs and Privacy Boards are not, therefore, intended to be comprehensive reviews of the relevant policy, but rather some guidelines for you to understand what kind of information the IRB or Privacy Board will need to know in order to offer guidance.

The Health Insurance Portability and Accountability Act (HIPAA). HIPAA, Public Law 104-191 which passed in 1996, includes among other provisions a set of “Administrative Simplification” provisions that require the Department of Health and Human Services (HHS) to adopt national standards for electronic health care transactions. Besides promoting efficiency in the health care system, however, such technological advances also create a host of new concerns about the privacy of health information. Consequently, Congress incorporated into HIPAA provisions that mandated the adoption of Federal privacy protections for individually identifiable health information. Accordingly, HHS published the “Privacy Rule,” which became effective on April 14, 2001, with compliance date of April 14, 2003 (with a year extension for small health plans). Failure to implement and subsequently observe these standards can result in either civil or criminal penalties. If you are uncertain whether or how HIPAA regulations apply to your Quality Report project, check with your organization’s Institutional Review Board (IRB) or Privacy Board, as described below.

Institutional Review Board (IRB) approval. IRBs existed prior to HIPAA and continue to serve mandated functions extraneous to HIPAA requirements. To these, HIPAA has conferred some additional functions, primarily involving the authority to grant certain types of waivers of the Privacy Rule for purposes of research. An IRB proposal and approval, even prior to HIPAA, has been necessary for research projects involving human subjects when State, Federal, or other public funds and/or patients, consumers, students or employees are involved. The federal requirements are referenced in the Code of Federal Regulations 45 CFR 46 (applicable only when federal funds are used), 21 CFR 50, and 21 CFR 56. (available at:

<http://ohrp.osophs.dhhs.gov/humansubjects/guidance/45cfr46.htm>

IRBs are responsible for holding the investigator accountable for compliance with these regulations based on its own judgment about the benefits and risks of the project. These issues will differ somewhat based on the method of sampling and data collection and for the survey and the type and source of data for administrative measures. The Federal Government’s guidelines and regulations governing the functions of IRBs are available at <http://www.fda.gov/oc/ohrt/irbs/default.htm> .

The Office for Protection of Research Risks in the Department of Health and Human Services provides a set of decision trees for determining when a research project is subject to IRB approval for purposes of federally-funded research projects. These are available at <http://ohrp.osophs.dhhs.gov/humansubjects/guidance/decisioncharts.htm>. Although most performance measurement activities are not funded as federal research projects, these guidelines will help with ensuring the protection of consumers’ rights in any context.

Privacy boards. Privacy boards are constituted by HIPAA for health care organizations that that have not served as a research setting such that they would have been required to establish an IRB. Privacy boards serve a somewhat analogous function to that of IRB, but one that is more circumscribed involving only the authorization of waivers to the Privacy Rule for purposes of research. Privacy Boards perform none of the activities conducted by IRBs in connection with the Common Rule.

For the relationship between IRBs and the HIPAA Privacy Rule see <http://privacyruleandresearch.nih.gov/irbandprivacyrule.asp>.

Information that IRBs and Privacy boards may require:

- A description of data collection procedures to ensure confidentiality;
- A description of the process to obtain consent including interviewer scripts and forms;
- Examples of any materials that will be sent to consumers such as postcards, letters, etc.;
- A description of how the interviewer would access help in an emergency;
- Examples of confidentiality forms to be signed by all interviewers and project staff;
- Information about how to access mental health services.
- A 6th grade literacy level for all materials (for Medicaid consumers, HCFA requires this as well).

REFERENCES

- Backer, T. (1995). Assessing and enhancing readiness for change: implications for systems change transfer. Reviewing the Behavioral Science Knowledge Base on Systems change Transfer. T. Backer, S. David and G. Soucy. Rockville, MD, National Institute on Drug Abuse.
- Center for Medicare Education (2002). "Marketing Your Program." Issue Brief 3(5): 1-2.
- Lynn, J. (2004). "When does quality improvement count as research? Human subject protection and theories of knowledge." Qual. Saf. Health Care 13(Feb): 67 - 70.
- New Freedom Commission on Mental Health (2003). Achieving the Promise: Transforming Mental Health Care in America. Final Report. Rockville, MD, DHHS Pub. No. SMA-03-3832.

QUALITY REPORT RECOMMENDATIONS ON PRIVACY AND CONFIDENTIALITY

To address the privacy and confidentiality issues related to the Common Rule and the Privacy Rule in planning your Quality Report project, we recommend the following:

- Become knowledgeable about privacy protections in both federal and state law. Assess both federal and state confidentiality laws and educate all parties involved in the project.
- Decide whether the data will be anonymous—this decision will affect the extent of your planning for processes to ensure confidentiality.
- Include IRB/Privacy Board requirements and confidentiality considerations in your planning *as early as possible*.
- First, assess whether an IRB/Privacy Board review is necessary. This can be done by consulting with lawyers and/or Board members.
- If a review is necessary, adjust your timeline accordingly to prevent delays in implementing your project. Boards may meet at relatively infrequent intervals, and require submission of materials prior to the meeting. There may also be a limit to the number of proposals considered each month on a first-come first-serve basis. Assume that you will be required to return for at least one subsequent meeting to demonstrate that recommended revisions are adequate. A Board may also require the final versions of instruments and documents such as consumer letters, consent forms, return postcards, etc.
- Decide whether the data will be anonymous—this decision will affect the extent of your planning for processes to ensure confidentiality.
- Clearly articulate the benefits of your performance measurement system to communicate their importance. (This step is important to ensure that the Board has full information to assess whether the benefits outweigh the risks).
- Include rationales that you may have assumed obvious, such as why the consumer perspective is important, how the data can be used for service planning, the benefit of understanding the characteristics of consumers who have optimal and less than optimal outcomes, etc. Remember that the level of expertise and sensitivity related to mental health may vary among board members.
- Establish procedures for reporting information such as harm to self or other, and child abuse.
- Include confidentiality in your training of project staff.

Phase II. How do I Implement a MHSIP Quality Report Project?

PHASE II STEPS

- Step 1. Determine Schedule of Administrative and Survey Data Collection
- Step 2. Assess Capability for Collecting Administrative Data
- Step 3. Determine Method of Consumer Survey Administration
- Step 4. Decide on Data Collectors/Interviewers
- Step 5. Establish Procedures for Obtaining Consent
- Step 6. Establish Policy on Inducements and Incentives
- Step 7. Decide on Sampling Methodology
- Step 8. Determine Sample Size
- Step 9. Designate a Data Manager
- Step 10. Establish Data Security Procedures
- Step 11. Create data tracking logs
- Step 12. Establish Data Entry Procedures

With all the elements of the planning process in place, you are now ready to proceed with your Quality Report Project. This section discusses various factors to consider and provides guidance for addressing issues that may arise in the course of the project: decisions about the measurement period for both administrative and survey measures, the use of administrative information such as claims and encounter data for performance measurement, survey administration and sampling methodology.

Step 1 Determine Schedule of Administrative and Survey Data Collection The first step in implementation is to determine the cycle for data collection. Data may be collected at different time intervals. Possible options are:

- Monthly
- Quarterly
- Semi-annually
- Annually

There are several advantages of **shorter intervals**. These include greater retention and recall of intervening events for surveys, and the opportunity of identifying problems and improving services more rapidly.

The disadvantages are greater expense, and for surveys, greater burden and intrusiveness.

Data collected at **longer intervals** is less expensive, and in the case of survey data less burdensome and intrusive to collect.

Its disadvantages are that survey data is less likely to be accurate, it is less likely to be timely, and harder to interpret since it is less likely to reflect specific events. Also, if the look back time is set shorter than the data collection interval, a period of time will not be assessed.

The question regarding how often to collect data is a balance between cost and on the one hand, and precision and timeliness on the other. Policy makers and those who will make use of the consumer surveys will need to weigh the various advantages and disadvantages of the timeframe of the surveys.

Step 2. Assess Capability for Collecting Administrative Data As described in the Final Report, the MHSIP Quality Report relies on both administrative and consumer self-report data. Here we discuss the methodological issues related to using administrative data, which can include enrollment and encounter/claims data, case records, and other records that are routinely kept by a service provider or management organization. The collection, management and analysis of administrative data presents numerous logistical and methodological challenges. As noted in the previous section (How Should I Plan my Quality Report Project) you should early on become very familiar with your organization's processes for collecting MIS data in order to anticipate and develop plans to address complications.

Administrative data, by definition, is intended for purposes other than performance measurement, e.g. management, accreditation, regulatory requirements, and contract compliance. The MHSIP Quality Report administrative data measures were designed with the awareness that their usefulness would depend in part on the extent that they made use of existing data elements developed for other purposes. It may be, however, that required data are not readily available or not available at all. Data requirements may exceed the capacity of existing information systems and administrative infrastructure or may create an insupportable burden for clinicians, administrators, and support staff.

In addition, inevitably, a degree of tension results from efforts to use data for multiple purposes. In such cases, truly the "devil is in the details." That is, minor features in the specification of data elements for management purposes can have profound implications for the results of performance measurement. An example of this is when the way in which "membership" i.e. the enrolled or eligible population is identified for administrative purposes is not congruent with the appropriate specification of the denominator for a performance measurement. These inconsistencies can profoundly complicate both the interpretability of the data and the usefulness for purposes of comparison or benchmarking. For these reasons, *active participation by IT staff* throughout the implementation is essential.

Further guidance regarding these issues is presented in the section "Data Management" below, and in the following Chapter, the section "Data analysis."

Step 3 Determine Method of Consumer Survey Administration One of the key lessons learned from use of the Version I Report Card is that organizations are likely to vary widely in their method of administering the survey, to the extent that a) reliability in any particular case is very difficult to determine and b) the potential for using the information for almost any purpose, whether benchmarking or other comparisons, quality improvement, performance monitoring, planning, etc., is often compromised by the

uncertainty about the extent of bias introduced by the method of administration. This problem is amplified by lack of standardization in reporting results, discussed in Phase IV.

The method of survey administration can affect the quality of data by:

- The number of persons who respond
- The extent to which persons express their true beliefs or feelings (bias)

The most commonly considered methods of survey administration are:

- Mail
- Telephone
- Point of service
- In-person interviews
- Combinations of the above

Telephone Surveys. One advantage of telephone surveys is that they usually produce higher response rates. Another is that this method requires no reading ability and the persons making the survey call can answer simple questions regarding its completion.

Several disadvantages must be weighed against the advantages. First, not all potential respondents will have telephone numbers. Second, telephone recruitment can violate privacy unless a protocol is developed for the situation in which someone other than the respondent answers the phone. Third, the respondent may not be able to find a private space in which to respond and the respondent may view the interviewer as violating their privacy. Finally, phone interviews are more expensive than simple mail-outs, especially when repeated call backs are necessary, as they often are.

Telephone interviewing also has the disadvantage of requiring interviewer training. Just a few of the areas in which interviewers must be trained are:

- Recruiting respondents
- Avoiding biasing responses
- Accurately recording responses
- Reassuring respondents who are upset by questions

Mailed and self-administered. Mailed surveys have the advantage of potentially reaching the largest number of possible respondents. While some people you may want to survey may not have phones and many may not be coming in for service, most will have mailing addresses.

Another advantage of mail surveys is that unless someone's mail is opened by another person, the request for participation is private. Also, the respondent can choose a private time and place to respond. Another advantage of mailed surveys is that if cash or other incentives are being used, they can be contained in the envelope.

Finally, mailed surveys are relatively inexpensive to administer. However, since the response rate for surveys can be low, the cost per completed survey can be higher than it might first appear, especially if multiple mailings and telephone follow-up are employed.

The primary disadvantage of mailed surveys is poor response rates, typically between 20 and percent. An additional disadvantage is that mailed surveys can only go to persons with addresses. You may also experience difficulties if you want to survey persons who reside in treatment facilities. These persons may not have private and confidential mail access.

Mail-out methods may utilize a variety of strategies for increasing response rates and/or accounting for biases due to non-response. The following list describes some features of these strategies:

- Anonymous versus identifying information: identifying information permits follow-up for non-respondents and adjustments for overt bias as described above
- Parallel postcard to be returned if survey returned: This strategy calls for enclosing a postcard along with the survey and asking respondents to return the postcard when they return the completed survey. The postcard contains identifying information but the survey itself does not. This allows for tracking and adjustments while maintaining strict anonymity of survey responses.
- Use of postal permit to reduce cost
- Non-identified envelopes to protect confidentiality of service use (i.e., envelopes containing surveys do not have "Department of Mental Health" return address)

A full discussion of these and other options can be found in the following references:

Salant, P. & Dillman, DA., (1994) How to conduct your own survey. London: Wiley.

Dillman, DA (1978). Mail and telephone surveys: the total design method. New York: Wiley.

Dillman, DA (1991). Mail surveys: a comprehensive bibliography, 1974-1989. Chicago, Ill.: Council of Planning Librarians.

Internet Surveys. Online survey administration is a rapidly growing field of technology, and many health plans now provide this capability for their enrollees. CMHS, through the Decision Support 2000+, has developed a prototype for completing the MHSIP Consumer Survey online (<http://www.ds2kplus.org/DS2K/main.aspx>). The feasibility and utility of this approach are being tested in

several state mental health systems. Subsequent editions of the Toolkit will provide full guidance on the use of these systems when they have been fully tested and implanted.

Automated Data Capture Systems. Software packages are available for collecting survey information by fax or internet and automatically entering into databases. These have a number of advantages including reducing the cost of data entry, speeding up the data collection process, reducing data entry errors, and building upon existing IT capabilities. Major disadvantages are the requirements for technical expertise and convenience for consumers.

Prior to making a commitment to this approach, you should assess costs and benefits according to the usual procedures for any IT investment.

Administration at Point of Service. One advantage of point of service interviews is a high response rate, particularly if the respondent is asked to complete the survey before leaving the point of service. Another is that the person distributing the survey can answer simple questions regarding its completion.

One disadvantage is that a point of service survey only reaches persons who appear for service. This can be partially addressed by using a wide window for data collection. A second disadvantage is that the method requires someone to distribute and receive the survey. There is consensus that this should not be a service provider or anyone associated with service provision, particularly when it comes to receiving surveys, as this might influence consumer to respond more positively than they feel or worry about the treatment they receive. Alternatives to providers are clerical staff or those involved in quality monitoring. To repeat, the most ideal person would be someone viewed as outside the service system.

In-person Interviews. In-person interviews have the advantage of having the highest response rate. This method also requires no reading ability and allows for persons to be helped to complete the survey; i.e., interviewers or surveyors can play a more or less active role in facilitating survey responses depending on the individual needs of a respondent. In-person interviews also allow for interviewer ratings. For example, with proper training, an interviewer can complete a rating of consumer level of functioning or symptomatology, if this measure is also included as a performance measure. Short of this, interviewers can provide basic information that may shed light on whether the respondent appeared to understand the survey process.

In-person interviews have the disadvantage of being the most expensive method. Like the telephone interview, they require resources for training interviewers as well as survey administration. It is also the most intrusive, and possibly the most burdensome method. Additionally, in-person interviews require careful consideration to issues such as interviewer safety, where interviews will take place, and respondents' comfort levels with persons of the opposite sex. All of these specific examples have emerged as difficult issues in one or more performance measurement projects.

Combined Methods. Combined methods can be used to minimize some disadvantages of single methods. However, they also combine certain disadvantages such as risks to privacy and expense. An example of common combined method is following a mail-out of the survey with phone calls to non-respondents attempting to complete interviews over the phone.

Most MHSIP Report Card efforts to date have used mailed surveys. At least one project used consumer interviewers and one used point of service administration. To date, no single method of administration has become the standard, but rather individual projects have sought to weigh the advantages and disadvantages of the various choices in a local context.

Step 4 Decide on Data Collectors/Interviewers If you decide to manage a point of service, phone, or in-person survey internally, you will have to decide who will distribute and administer the survey. Whom you choose to do so may affect your response rates and the quality of your data.

The most frequent options for survey distribution and/or administration are:

- Providers
- Consumers
- Third parties

Providers. Data collection by providers often has the advantage of being convenient. Providers are likely to have frequent contact with survey respondents, they know the respondents and may be trusted by them, and they are more likely to have necessary interviewing skills.

However, data collection by providers has the very important disadvantage that persons may be reluctant to be completely honest with them because they are part of the system being assessed. Consumers have expressed fear of retaliation in some cases and more general feelings of vulnerability when providers are involved in collecting evaluative information. It can also interfere with provider time designated for service provision. This may result in low response rates and/or spurious responding. In addition, consumers may not feel free to choose to decline to participate when asked by providers. *All in all, the disadvantages of using providers for data collection from consumers clearly outweigh the advantages in all but exceptional cases. Thus, we strongly recommend separating the data collection process from service providers and the service provision process.*

Consumer interviewers. Evidence suggests that distribution and administration of surveys by consumers will result in a higher response rate and more reliable responses. Consumer interviewers also adds empowerment value to the process.

A potential disadvantage of using consumers is that they usually have less experience than other persons who might be employed to conduct interviews. Consequently, training needs may be greater, particularly when compared to the option of using a professional survey research firm. Despite these disadvantages, a number of states have developed significant infrastructure to support this approach, including contracting with consumer-run organizations and providing benefits counselors to address issues related to the ramifications for income-based eligibility programs.

Third parties. Increasingly organizations have turned to commercial survey research firms (or in some cases academic research centers) in order to enhance the quality of their consumer survey methods. A professional firm will likely bring a great deal of experience and infrastructure for collecting data to the project, but can also

be prohibitively expensive. On the other hand this may be a cost-effective investment, as a poorly designed survey project is still costly, but the results are of limited value.

An alternative to survey vendors is to hire and train interviewers yourself. Experience suggests that you can probably find people, such as college students looking for part-time work, with appropriate skills and experience in the labor market. These decisions require careful assessment of your organization's resources for a project of this nature.

One advantage of third parties is that they may not cause respondents to alter their responses for fear of consequences since they are likely to be perceived as separate from the mental health system. A disadvantage of third party interviewers may be that they do not have credibility with some consumers. Some consumers may decline to be interviewed or provide data that is spurious. More generally, persons who do not have experience with the mental health system or with persons with mental illness will need to be trained extensively. You will need to ascertain that they understand the nature of the population you are addressing and the characteristics of the service system. This communication will require some understanding of survey research methodology on your part to be able to determine whether their approach to sampling, survey administration, etc. effectively addresses the unique circumstances of your organization and the consumers it serves.

Step 5 Establish Procedures for Obtaining Consent Successfully collecting survey data from consumers identified by the sampling strategy chosen is necessary to ensure the desired precision for measures. For this reason it is important to consider how to solicit consumers to participate, how to conduct the informed consent process, and what incentives and inducements can be used to increase participation.

Despite the importance for technical reasons of collecting data from all targeted consumers, it is worth emphasizing that you should approach every individual in a non-coercive way and provide ample guidance through the consent process. Ultimately, it is up to the individual to decide whether or not to complete the survey; it is your job to ensure that the consumer is provided with all the appropriate information so that he or she can make a fully informed choice.

Several different procedures for obtaining consent have been used in different settings. These include:

- Contact by a QI person or ombudsperson to obtain consent followed by contact by interviewer or data collector
- Consent process and data collection conducted by interviewer
- Inclusion of consent form in mailing containing survey
- Initial contact by consumer organization followed by referral to interviewer
- Consent and data collection conducted by a peer

No one process for obtaining consent has shown itself to be preferable to all others. In devising your consent process it is important to involve stakeholders, particularly consumers, and to consult requirements or regulations as a result of approval by an IRB.

Recommendations for obtaining consent:

- Make sure the consumer understands what the study is about, and what they are being asked to do.
- Explain the importance of the survey and what it is trying to find out.
- Describe how data might help to improve services.
- Answer any questions that he or she might have about the study.
- Review your procedures regarding privacy, anonymity and/or confidentiality.
- Offer the report and/or opportunity to attend meetings to discuss results at end of study.

Step 6 Establish Policy on Inducements and Incentives You can employ non-coercive strategies to increase participation from consumers. We refer to these strategies as inducements (i.e., actions that might lend credibility to the project and thus raising the appeal of participation), and incentives (i.e., actions that provide more direct compensation for participation).

Examples of **inducements** include letters supporting the project from important stakeholder groups.

Incentives can be cash, coupons or other means of compensation. Some have suggested that cash or money orders are preferable to checks, since not all respondents have bank accounts.

Step 7 Decide on Sampling Methodology Surveys are an economical means of determining characteristics of a population by observing a small sample of the population. Though possible, it is seldom feasible to survey the entire population of interest; in most cases you will need to select one of several sampling strategies.

Simple random sample. This is the most basic form of a sample survey, whereby every member of the population has an equal probability of selection. The value of this form of sampling is that, in general, the observed sample will have characteristics similar to those of the total population.

More complicated forms of surveys include stratified random samples and cluster (or multi-stage) surveys. These more complicated forms of sampling do not, in general, produce a sample that has characteristics similar to the total population. However, when combined with appropriate statistical techniques, they produce valid (i.e., unbiased) estimates of characteristics of the total population. The value of these more complex survey designs is that they produce more reliable estimates of population characteristics (i.e., less variability) and are often cheaper to implement.

Random vs. Stratified Sampling. It is important to decide whether to sample randomly from a population or whether to divide the population into subgroups or strata. Though many issues need to be addressed in

deciding on a sampling scheme (cost, precision, ability to implement), it is important to avoid what has been called "the tyranny of the large." Are there small subgroups of interest that conventional sampling won't allow us to find? For example, if one were to conduct a national survey, then people living in a frontier setting (<1 person per square mile) would be highly unlikely to appear in a simple random sample of 1,000 individuals. If this group were important to have in your sample, you should consider oversampling in this subgroup. Sometimes subgroups may be so small that it may be desirable or necessary to sample subgroups across certain providers or geographic boundaries.

Cross-sectional vs. Longitudinal Sampling. For measuring change over time, there are two broad categories of survey strategies: multiple cross sections and longitudinal designs. Within each of these categories, one can use any of the various sampling schemes mentioned so far, for example simple random samples or stratified samples. Multiple cross sectional samples (MCSSs) draw subjects from a group of people who meet certain criteria (e.g., received a service in last month). Each MCSS drawn over time might be composed of different persons. Longitudinal samples, conversely, identify a group of persons at baseline and follow these individuals over time. As a broad generalization, multiple cross sections are powerful methods for detecting system change, whereas, longitudinal designs are powerful methods for detecting individual changes.

A major advantage of **multiple cross sectional samples** is that they provide information about changes in population composition, characteristics, and attitudes over time. Another advantage of multiple cross sectional samples is that if specific subgroups are continuously represented in a population over time, change in these subgroups can be tracked. Examples of subgroups are:

- One or more systems
- One or more plans or providers
- One or more age groups
- One or more diagnostic groups

A major disadvantage of multiple cross sectional sampling is that, in general, it does not allow us to track individual change. For example, we may learn that functioning is improving in a population, but we cannot be sure whether this is because persons are becoming higher functioning or because a population is adding more higher functioning persons.

In **longitudinal samples**, one follows a specified group of individuals across time. A major advantage of longitudinal samples is that such samples give us information about individual and subgroup change. These changes can be discriminated from changes in population membership. The major disadvantage of longitudinal sampling is that it is difficult and expensive to carry out. The difficulty comes in finding persons for longitudinal follow-up. This is referred to as respondent "retention" in discussions of evaluation methods.

It is possible to combine cross sectional and longitudinal sampling. In these approaches, the cross sectional sample is larger, while the longitudinal design follows smaller groups of individuals intensively. This combined approach is certainly the most comprehensive, but is also the most resource intensive.

Some ideas for increasing the rate of retaining respondents are:

- Obtaining the names of friends or relatives to contact if a respondent moves
- Establishing a toll free phone number so that persons who move can contact a project when the project can't contact them

Additional ideas about retaining respondents are discussed in the following references:

Coen AS, Patrick DC & Shern DL. (1996). Minimizing attrition in longitudinal studies of special populations: an integrated management approach. *Evaluation and Program Planning*, 19:309-319

Ribisl KM, Walton MA, Mowbray CT, et al. (1996). Minimizing participant attrition in panel studies through the use of effective retention and tracking strategies: review and recommendations. *Evaluation and Program Planning*, 19:1-25.

Cluster, multistage. A cluster or multi-stage sample implies sampling at more than one level. For example, we may first draw a sample of providers, and then from the sampled providers, we may draw a sample of persons receiving treatment or services. The advantage of this approach is that if one is doing face-to-face interviews, it reduces the number of providers (or the number of interviewers) that need to be contacted, and thus reduces the cost of the survey. The disadvantage is that the survey data requires special computer software or trained sampling statisticians to construct estimates and confidence intervals from this type of survey data.

Inverse sampling. An inverse sampling techniques is a power method to estimate the frequency of rare or unlikely events. In this sampling design, one keeps sampling until reaching a pre-specified fixed number of rare events (e.g., suicide). This method allows greater precision for answering a specific question, but is not as good when one has many questions which need to be addressed.

Step 8 Determine Sample Size In performance measurement, it is possible to design projects that are more or less precise for estimating the magnitude of measures, more or less certain with respect to whether groups differ, and differentially able to detect group differences as a function of their size. One key factor in determining these things is sample size.

If your Quality Report project is to employ a sampling method, you will need to decide during the planning process precise, certain, and sensitive you want your measurements to be. Different policy uses require different levels of precision, certainty and sensitivity. For example, it is likely that measurements made for the purpose of quality improvement that will only be used internally require one level, whereas measurements made for monitoring contract compliance that might be used in legal proceedings require another.

Precision and certainty are directly linked to sample size. Of course, if your goal is to give all consumers a voice in assessing the service system, then it would be ideal to collect data on everyone in a target population. However, this may be impossible for practical reasons. Or it may be possible, but only if some inexpensive method, such as mailed surveys, is used that may result in low response rates and increase the potential for both overt and hidden biases. Therefore, it may be preferable sometimes to seek a smaller sample from which you can obtain a higher response rate.

The advantages of **smaller samples** are that they are less difficult and costly to obtain. It is also easier to do the data management associated with smaller samples. Their disadvantages are that they are less precise, offer less certainty, and are likely to detect only large group differences. Smaller samples also limit the potential for subgroup analyses.

The advantages of **larger samples** are that they offer more precision and certainty, and are more likely to detect smaller group differences. They are also more likely to permit subgroup analyses. However, the larger a sample, the more difficult and expensive it is to collect data and more difficult to manage data.

Sample size and Precision. There are two quick rules for the relationship between sample size and precision, one for proportions and the other for means. In general, there are two basic ways in which data are reported: the proportion of cases with a certain attribute (e.g., percent dissatisfied), or the sample mean (e.g., average score on the for the Outcomes domain). In samples of about 100 persons the precision in estimating the proportion is within plus or minus 10 percent. This represents a so-called 95 percent confidence interval. Thus, if you observe that 32 percent of the sample is dissatisfied, you would say that you are 95 percent sure that between 22 percent and 42 of the total population was dissatisfied. If one had reported the mean, then the 95 percent confidence interval would be plus or minus .2 times the standard deviation. These methods work for the total sample as well as subgroups within the sample.

If you have had samples of 400 per group of interest, the 95 percent confidence intervals for a proportion are within plus or minus 5 percent, and for the mean 0.1 times the standard deviation. It should be noted that to make the 95 percent confidence interval half the size (that is, to go from plus or minus 10 percent to plus or minus 5 percent we needed to increase the sample four-fold. To make the 95 percent confidence intervals range plus or minus 2.5 percent therefore, would generally require a sample size of 1,600.

Finite Population Correction. The only exceptions to the above rules are instances when the sample represents a large portion of the population. This is known as the finite population phenomenon. To correct for this, one uses a finite population correction factor. Essentially, this correction factor implies that if you have seen a large percentage of a population, then your precision is greater than if you had only seen a small percentage. The formula for the finite population correction factor is the square root of $(1 - n/N)$, where n is the number of respondents in your sample, and N is the number in the total population. Using the above as an example, we have said that if you have 100 respondents, then your precision is plus or minus 10 percent. Now if there were only 200 persons that could be observed, your precision would increase to plus or minus 7 percent (10 percent times square root of one half).

Power. When designing a research study, investigators will often perform a power calculation. In the research setting, power is interpreted as the chance you would miss a statistically significant difference due to the size of your sample. Thus, as the number of subjects in a study increases, so does the power of the study. In performance measurement systems, it may also be necessary to compare groups. The performance monitoring system differs from the research setting in that there may be many groups to compare, and more than one measure to compare. It is often the case that we may wish to compare one program to another and we may wish to compare both programs to a common benchmark.

As an example, recall that with 100 respondents, the proportion is known to within plus or minus 10 percent. Thus, if you are comparing a program with 100 respondents to a benchmark, if the program was more than 10 percent away from the benchmark, you would say that the program is significantly different (at the 0.05 level) from the benchmark. If you were to compare two different programs, each with 100 respondents, then if the programs differed by more than 14 percent you would say that the two programs were statistically different from one other. As stated above, if you had 4 times as many individuals in each program, you would be able to say that differences of 7 percent were statistically different.

These are statistically significant differences, not clinical or policy relevant differences. As the sample size in your data set grows, so will the number of statistically significant differences. It is important to decide how large of a difference is necessary to be considered important. This discussion will help frame the question regarding how large a sample one needs to draw. The calculations above assume similar populations or adjustments for case mix differences have already been done. They also assume a single comparison is being made; multiple comparisons invite opportunistic chance findings and the interpretation of levels of significance is more difficult.

QUALITY REPORT RECOMMENDATIONS FOR SURVEY IMPLEMENTATION

- It is useful to obtain identifying information as opposed to sending out wholly anonymous surveys to allow you to account for bias caused by non-response
- If you wish to compare groups you need to have larger group sample sizes than if you only wish to estimate the magnitude of a measure for one group
- We recommend not changing response categories, because doing so will make your data less sensitive to individual differences and make it more difficult to compare with the data from other MHSIP projects
- It is important to consider how to solicit consumers to participate, how to conduct the informed consent process, and what incentives and inducements can be used to increase participation
- It is desirable to collect data from all consumers targeted; each individual should be approached in a non-coercive way and guided through an informed consent process
- We strongly recommend separating the data collection process from service providers and the service provision process
- We suggest that you have the appropriate expertise within or available to the group that is to do translation work of the MHSIP instrument
- Back translation is normally recommended; the process does require greater investment of resources
- Interviewers should be well grounded in the cultural world of the respondents, and they should be very familiar with culturally appropriate conduct in social situations
- Before making any comparisons across or between systems, you should carefully examine the comparability of the definition of data elements, data sources, data collection, case mix adjustment and sampling strategies, as well as actual methods of calculation.

DATA MANAGEMENT

The following steps address procedures for collecting data, ensuring quality and security (confidentiality) and methods for dealing with missing data. It addresses issues related to both *administrative data already stored* and *data requiring entry* (DRE), such as consumer surveys.

Step 9 Designate a Data Manager Before undertaking data management procedures, an organization administering the MHSIP Quality Report should designate a single individual to act as data manager. Investing responsibility in one person facilitates organization of data and reduces the probability of errors caused by miscommunication. Although the data manager should hold primary responsibility, others should be familiar with the database and data management procedures in case the data manager is unavailable at any time. The data manager should maintain a codebook listing all data management decisions, procedures, and operations performed. This will ensure consistency in the data across time and data managers.

If resources require that responsibilities be distributed among two or more individuals, it is essential that all use identical techniques. They should be trained together and should meet frequently to plan data management procedures to ensure that data managed by different individuals is compatible. A codebook, such as that mentioned above, also allows different data managers to monitor work done by others and to make sure that their own work conforms.

Step 10 Establish Data Security Procedures It is crucial that you take measure at each stage of data management to ensure the confidentiality of the data. The most important consideration, and the most extensive guidance, for protecting the privacy and confidentiality of consumer information is HIPAA, specifically the Privacy Rule and The Security Rule (for more information see: <http://www.hhs.gov/ocr/hipaa>).

In addition, you should find out what laws govern consumer confidentiality in your locality. You should also review what statements regarding confidentiality were made in proposals, disclosure agreements, and informed consent documents. Make sure that all practices conform to the laws and the stated policies. All staff should be thoroughly trained to concord with these confidentiality guidelines.

Policies for handling outside requests for data should be consistent with HIPAA requirements. Any site administering the MHSIP Report Card must decide whether data will be shared with others. If data will be given to other organizations, steps must be taken to ensure that this transfer of information does not breach confidentiality.

Several technical measures should be taken to ensure confidentiality.

- Internet transmissions containing any potentially identifying information should be encrypted and sent over a secure connection to an email address accessible only to individuals who need to see the data.

- Computer file containing any potentially identifying information should be saved as an encrypted file and password protected. Only individuals who require access to this information should be aware of the password.
- All raw data containing any potentially identifying information should be stored in locked file cabinets. A single individual who is aware of project staff who might require access to this information should hold the key and monitor access to the files.
- Current HIPAA requirements for data security should be reviewed and any not included above should be implemented.

Step 11

Create data tracking logs As an essential part of the database codebook, the data manager should create logs that delineate each step in data management procedures. There are three types of data tracking logs: 1) an “interview log”, applicable only to collected by survey, 2) a “data entry log”, applicable to all DRE, and 3) a “data checking” log, applicable to both DRE and administrative data.

Interview Logs. An interview log is a record of steps in the administration of a survey and will vary according to the method of administration. For example, a column for date of interview would not be applicable for a mail-out survey administration. An example “interview log” for a face-to-face interview methodology would include the following information:

- Identifying information for the interviewee (See the section on confidentiality below)
- Contact information for the interviewee
- Back-up contact information in case of invalid address or phone number
- Identifying information for the interviewer
- Contact information for the interviewer
- Date interview scheduled
- Date interview completed
- Reason for late interview

Data entry logs. All DRE should have a data entry log. This should contain the following information:

- Date of receipt of raw data
- Data entry person or service
- Date of completion of error checking review of raw data
- Date sent for data entry
- Date of receipt of entered data
- Date of completion of error checking of data entry
- Incidence of data entry errors
- Date data entry errors corrected

- List of errors that could not be corrected
- Incidence of missing data
- Miscellaneous information for a batch of data that should be noted for future batches (e.g., a survey that should be excluded from analysis because the respondent rescinded consent)

Data checking logs. Both DRE and administrative data should have a data checking log. This type of log should contain the following information:

- Date of completion for each electronic error checking procedure
- Incidence of errors for each electronic error checking procedure
- Date errors corrected
- List of errors that could not be corrected
- Miscellaneous information for a batch of data that should be noted for future batches

These logs will maintain an institutional memory of data management and assist the data manager in monitoring the data. They should enable the data manager to ensure that important methodological issues are being addressed such as:

- Are sufficient interviews being scheduled?
- Are interviews being completed on time?
- Are sampling procedures being reflected in the data (i.e. does the data appear as expected)?
- Is data entry consistently accurate?
- Is there a high incidence of errors in the data?

Step 12 Establish Data Entry Procedures There are several options for data entry, such as a data entry person or service, data entry software, or scanner or fax software. No matter which option is used, a database structure must be created before data can be put into electronic form. The best approach for this task is to create a data key.

Creating a data key. Examine the document that is going to be entered. For each item to be included in the database, define the following parameters:

1. **Decide on a variable name.** In some database software requires label to be eight characters or less.
2. **Define the format of the variable.** Examples of variable formats are numeric (containing only numerals), string (containing characters or numbers), and date (containing a combination of characters and numbers representing a date).

3. **Define how proper responses should be coded.** For each item with response categories, each category must be assigned a value. For example, an item inquiring about sex might be coded with a value of “0” for male and a value of “1” for female.
4. **Define how improper responses should be coded.** Often responses such as “doesn’t know”, “no response”, “not applicable”, or “refused” will appear. Values must be assigned to each of these responses to allow monitoring of the frequency with which they occur.
5. **Define the length of the variable,** i.e. the number of characters or numbers of which it will be composed. For items with response categories, this is simply equal to the number of characters required for each category. For open-ended questions, this is the maximum number of characters you wish to allow for a response.
6. **Define variables for other information.** In addition to the items being entered from the document itself, variable information should be defined for any other data of interest, such as unique identifier, date of administration, and date of entry. Writing this information on a master document next to the item to which it corresponds makes for easy reference in later stages of data management and analysis.

Entering data. Once a data key has been created, you are prepared for entering the raw data into electronic form. Irrespective of data entry method, it is advisable to enter data in small batches as it is received (assuming it is not received all at once). In addition to making the task more manageable, it is useful in preventing large-scale errors. If a mistake occurs, it can be corrected in a small batch and be averted in future batches.

You should begin by organizing the raw data. If there are different DRE documents being entered (e.g., consumer surveys, family member surveys), you should group each type together. It is useful to organize each group by a unique identifier so that you can more easily navigate the raw and electronic data in the event of any problems.

The data may be entered manually into an ASCII file (a generic file with no particular format) or directly into a particular database manually, or it may be entered with data entry software. If you have entered the data into an ASCII file, you must then read this file into a particular database format. This procedure is built into database programs and varies according to platform. Although data entry software requires that you set up entry screens and lay out the database prior to entry, this method can restrict responses to legitimate values and thereby reduce data entry errors. Both of these approaches to data entry are resource intensive as they require a person to type all the information into a computer. You must provide the data entry person with information regarding the data entry process. A copy of the data key is usually sufficient for this purpose. This will show how each item on the document will be created in a database.

Scanning. An alternative to the time-consuming method of entering data manually is to use scanning software. This type of program uses a scanner or fax to read the information from the document, eliminating the need for someone to type the information into a computer. Although this method can save time, it requires more advanced technical expertise and can be prone to technical difficulties.

Step 13 Establish Data Quality Controls The simplest and most straightforward means of maintaining high-quality data is to thoroughly train all individuals who will be involved in data collection and data management procedures. If all data managers are familiar with protocol and use identical procedures, potential problems will be averted. Beyond proper training, several other techniques can be employed to increase the accuracy of data.

Checking for data entry errors. For DRE, data managers can perform initial checks upon receipt of raw data prior to data entry. A sample of documents to be entered should be checked to see that forms were completed and items were filled out properly. Any errors that are discovered should be corrected prior to entering the data.

When the data have been entered, the data manager should perform several checks to ensure that no data entry errors occurred. A list of unique identifiers in the database should be printed and compared with the raw data to make sure that all data was entered. A random sample of the electronic data (e.g. 10 percent of cases) should be compared with the raw data to make sure responses were coded accurately. Any errors that are detected should be corrected. The incidence of errors should be recorded to gauge the accuracy of data entry. Another method of gauging data entry errors is to enter a random sample of the raw data in duplicate by alternate means and then compare the data. Any inconsistencies should be checked with the raw data and corrected, and the incidence of errors should be recorded.

Electronic error checks. Other data checking procedures are conducted on the electronic data and are applicable to both DRE and administrative data. These entail examining frequency distributions for range checks and outliers and checking for logically impossible combination of responses.

The frequency distribution should be examined for all variables in the database to check that all responses fall within the realm of possible answers. While rare, it is possible that data entry personnel have entered an out-of-range value, e.g., 3 for a “sex” variable that should be coded 1 or 2. In addition, sometimes respondents may enter information that is invalid, but correctly key-entered. A frequency distribution will reveal any of these out-of-range values. Whenever such errors are found, the data manager should refer to the raw data and correct the value. If the erroneous response is not attributable to data entry error and appears on the actual document, the data manager must decide whether he or she can reasonably recode it into a valid value. If not, the item should be recoded into a system missing value.

Outliers or improbable values. A review of the frequency distribution also enables the data manager to check for outliers or improbable values. In some cases, range and logic checks may not detect values that are possible but unlikely. For example, it is possible for an individual to be 100 years old; however, it is advisable to verify such extreme data values. This type of error check is particularly useful in examining reported rates of service utilization.

Logic errors. Wherever possible, data should be examined for logic errors, or logically impossible combinations of responses. For example, an individual who does not receive medication should not report involuntary muscle movements resulting from the use of psychotropic medications. These errors should be

corrected by referring back to the raw data. In cases where this is not feasible, decision rules must be developed for recoding the data to values that are more likely to be accurate or to code the fields as missing.

Depending on the database software being used, these checks can be automated. If the software accepts programming, syntax files can be written to conduct them. To do so, the data manager would need to create flags for each error. For example to check for outliers in age, a variable, *agecheck*, could be created and assigned a value of zero. Then, a conditional statement, such as *if age > 100 or age < 18*, could recode the variable to a value of one. This flag marks each case where an aberrant value for age appears. The same procedure can be implemented to check for out-of-range values and logic errors.

Step 14 Address Multi-level Data Management Issues Special issues arise when data must be

collected from multiple sites for analysis at a higher level. The first task that the higher-level data manager must address is merging the data files from different sites. This can be problematic if sites are using different database software or different variable formats in the same type of database. Problems of database software incompatibility can be addressed by using software packages that convert databases from one format to another. Many database programs also allow users to save databases in a variety of formats.

Problems related to differing variable formats can be more difficult to remedy. If for example, one site saves date of interview as a date format and another saves it as an eight character string format, it may require manual correction in certain database programs. This may be unavoidable with administrative data. However, these can be averted in DRE by comparing each sites data key prior to data entry. Making sure that all data files have similar structure and will therefore be compatible will greatly facilitate database merges later on. Familiarity with the data management techniques of each site will enable the data manager to provide technical assistance to sites by sharing the approaches of the others. Periodic on-site reviews of a random sample of sites can help the data manager to accomplish these tasks as well as to check on data quality.

A higher-level data manager must employ some of the same techniques used by site data managers. Upon receiving data from a site, the state-level data manager should re-run error and logic check procedures on the data and inform the site of any errors that are found. This will ensure that no errors went undetected. Similarly, frequency distributions should be re-examined to identify out-of-range values, missing values, and inconsistent data. The site data manager should be informed of any aberrations that these tests reveal. Although these operations are performed at the site level, it is prudent to conduct them again. This re-examination also allows for the comparison of data from different sites, which can identify anomalies that might not be evident in a single site's data.

Step 15 Establish Procedures for Dealing with Missing Data If data have been entered and error

checked and there are still missing responses, procedures must be developed for handling this missing data. This is extremely important, as for many statistical analyses, missing data on a single item results in the entire survey response being discarded. This could drastically reduce the number of subjects available for further analyses. Interviewers should make every effort to get respondents to answer every item. For some analyses, "not applicable" may be a legitimate response, but it cannot be included in analysis of a scale and must

therefore be assigned a missing value. The frequency of occurrence of missing values for each item should be recorded and included in any technical reports or appendices.

A frequency distribution of the items will reveal the degree to which respondents did not answer an item (often including “not-applicable”, which for this discussion will be considered missing). *We recommended that only respondents who provided valid responses for more than half of the items be kept in the analysis.* Whatever the strategy employed, however, it should be clearly described in the codebook. Tables can be constructed with the numbers of respondents with high numbers of missing item responses.

Imputation. For cases with items missing a response, it is possible to impute a value for the subject that has some rationale behind it. One method is simply to enter the sample mean of those that did answer the question. Another is to use the mean of items measuring similar content, often in the same subscale. Another more complicated procedure involves calculating a regression equation to predict a given item response from the responses to other items. Then, using the coefficients generated by the resulting equation to weight items for which there is a response, enter the predicted value for the person not answering the item.

All these techniques have consequences, however. They tend to reduce error variance and make differences appear larger than they actually might be. This situation can be improved by introducing “noise” into the prediction using patterns of variation from existing data. If you choose to do this, you may wish to consult with a local expert. This limitation should also be explicitly stated in the summary reports.

It is worth emphasizing that even the most technically advance methods for handling missing data have their limitations and drawbacks. Thus, the completeness of the data collection is absolutely critical in determining the type of conclusions you will ultimately be able to draw from the data.

Phase III. How do I interpret Quality Report Results?

PHASE III STEPS

- Step 1. Assess and Account for Bias
- Step 2. Decide on Approach to Comparison
- Step 3. Address Psychometric Issues
- Step 4. Control For Differences In Populations Served (Case-Mix Adjustment)

The purpose of any kind of performance measurement system is by definition to draw comparisons—whether comparison with past performance, some ideal, or some other organization. Such comparisons, however, raise a host of complex methodological and political issues that must be thoroughly addressed. Accordingly, the Quality Report Workgroup strongly recommends:

Before making any comparisons across or between systems, careful examination of the comparability of the definition of data elements, data sources, data collection, case mix adjustment and sampling strategies, as well as actual methods of calculation is essential.

The following steps identify issues to be considered in drawing conclusions from Quality Report information.

Step 1 Assess and Account for Bias Textbooks on survey methods typically assume that response rates will be 100%; i.e., all individuals selected for a survey will respond. In most real world settings, this is not the case. There will always be some number of individuals who will fail to respond for a variety of reasons. Response rates may vary by method of administration of the survey, inability to locate a consumer of services, anxiety in responding, or inability to understand the survey, to name just a few. Once a survey has any non-responders, no matter how few, the results of the survey are subject to two forms of bias: overt and hidden. Overt and covert refer to whether or not biases are, or can be, observed.

Identifying information. The value of responses varies with how much you know about who does and does not respond. Even very low response rates can provide sound estimates of population responses if we know something about the characteristics about who responded and who did not. Demographic, clinical and other information about respondents is also useful for analyzing response patterns among sub-groups for purposes of identifying performance improvement opportunities.

Proportions of respondents in various demographic dimensions should be inspected for departures from what would be expected in the client sample population, if that is known. It may also be possible to detect overt bias on the basis of theoretically critical issues, such as diagnosis or utilization. It is possible to perform this investigation only when there is some way to link survey responders to case records.

Post-stratification. If you have information on consumer's demographics, diagnosis, and service utilization, you can compare sample characteristics with that of the population. If the sample does not differ from the total

population, then there are no overt biases within the sample, and the data can be reported. If the sample and the total population do differ, then you will need to correct for overt biases.

One method is to stratify the sample on those characteristics that differ between the sample and the total, then weight the sample using the total population's weights. Such a technique is known as a *post-stratification* of the sample.

Propensity analysis. Another way to accomplish this is to perform a *propensity analysis*. Simply stated, a propensity analysis is a statistical model that predicts whether an individual will be a responder or a non-responder. This prediction model is constructed using logistic regression, with response status (responder/non-responder) as the dependent variable. Then findings are compared for persons with similar propensity scores in different groups.

Sensitivity analysis. Even a small percentage of non-respondents introduces the potential for hidden biases. A hidden bias cannot be controlled for with statistical adjustment since the characteristics that differentiates respondent and non-respondent samples are by definition unmeasured. Instead, one performs a *sensitivity analysis* to examine the impact of hidden bias. One way to do this is to assume the most extreme values (both positive and negative) for the non-responders. This extreme setting is equivalent to saying that there is a large correlation between the outcomes that are reported and the variables associated with non-response. A second approach is to explore how related other characteristics would have to be to influence the results. In this setting, the analyst examines various scenarios for the hypothesized non-respondent data and calculates the correlation coefficients necessary to produce this data. In this way, one can examine the sensitivity of the mean of the respondent sample to non-respondent hidden bias.

Step 2 Decide on Approach to Comparison Once you have computed results of the Quality Report, you can take one of two basic approaches, or a combination of the two, to make comparisons:

- Comparison of results across the organizations whose performance is being assessed by the performance measurement system in a relative manner,
- Comparison of results against benchmarks, derived prior to or separately from the collection of performance measurement data in this system.

Comparison of organizations to other organizations in the system. This is fairly straightforward and will be familiar to most. Very simply, this approach compares the results from one organization to another in a relativistic way. This type of comparison typically leads to ranking of organizations with respect to their competitors or peers.

Comparing scores from different groups involves comparing between-group differences (in averages or other measures of central tendencies) to measures of within-group variability. In general, larger between-group differences coupled with smaller within-group variability are more meaningful. Statistical methods for comparing groups are discussed in most statistic texts, and most statistical software programs provide a variety of methods for comparing group scores.

It is worth reiterating that comparisons between groups should not be made without carefully considering inter-group comparability and employing some form of case-mix adjustment for any differences observed or theorized to be of importance.

Comparison against benchmarks. Underlying the use of benchmarks to interpret performance measurement data is the assertion that a desired, minimum, or standard level of performance can be stated prior to collection of performance measurement data. Performance measurement data can then be compared to the benchmark—organizations that fall short of the benchmark are deemed unsatisfactory even if they compare favorably with other organizations. More complex benchmarks can include a range of acceptable values instead of a single reference point.

Benchmarks can be found or derived from many sources. No one source for benchmarks is likely to meet the needs of all indicators and measures within a performance measurement system—different sources of benchmarks may be more likely to support certain indicators.

Potential sources of benchmarks include:

- Humanistic values
- Public opinion
- Expert judgments
- Statistical norms
- Historical performance measurement data
- Scientific evidence of links to outcomes (i.e., for structure and process indicators)

Humanistic values. Humanistic values are used most often in setting benchmarks for indicators of untoward, sentinel events or critical incidents. For example, an indicator like number of consumers who are homeless, humanistic values may have a benchmark set at zero. Another case where humanistic values may be the basis for benchmarks is for indicators about fundamental rights; e.g., from a humanistic point of view, a system's performance should be questioned until all consumers have an individual service plan.

Public opinion. Public opinion may be used as a source for a benchmark when there is a strong and broad opinion about any aspect of service system performance. For example, the public may believe that every person should have a primary care physician, thus leading to a benchmark for this indicator at 100 percent.

Expert judgments. There is a wide range of methods for arriving at expert judgments, ranging from a single expert deciding based solely on his or her experience to scientific approaches like consensus panels on which several experts assemble the literature on a topic and, through facilitated discussion, reach a consensus judgment. Whenever possible, you should take advantage of instances where a topic found within your performance measurement system has been the focus of expert judgment derived by means of a systematic, scientific strategy.

Statistical norms Statistical norms are benchmarks derived from substantial testing of a measure on a population, often found with widely used survey instruments. It is important to recognize that statistical norms will often be derived for different populations and that it makes sense to adopt them as benchmarks in your performance measurement system only when your population is similar enough to the tested population. For example, an instrument may have norms for a healthy population and a disabled population. It would be important to understand the characteristics of the different populations to see if either would be appropriately applied to your performance measurement system.

If you know something about the comparability of your groups with the groups on which the norms are based, you can compare the values for MHSIP Consumer Survey measures. However, you must also consider the variability in the norms and your data. The more variable scores are within groups, the less meaningful any differences between the scores for those groups and norms or other groups. Most statistics textbooks explain how to compare scores to norms given measures of intra-group variability.

Historical Performance Measurement data. In the absence of statistical norms for many measures, historical performance measurement data can be used as benchmarks. A benchmark from historical performance measurement data can be simply how the individual organization performed in the prior year, can be the average performance across organizations in the prior year, or any other type of standard based on an organization's prior performance.

Step 3 Address Psychometric Issues

Validity and reliability

Validity refers to whether the measures obtained actually measure what they are intended to measure. Reliability refers to whether measures do so repeatedly and consistently. There are multiple types of validity and reliability.

If certain types of validity and reliability have been previously demonstrated for measures that are used without modification, it is usually not necessary to repeat these tests. Reliability as measured by internal consistency has been well established at acceptable levels for the Version 1 MHSIP Report Card Consumer Survey in a number of states. If you are using a measure for which no validity or reliability testing has been done, or if you have modified a measure, you may wish to do your own validation and reliability analysis. Below is a guide to various analyses you might undertake. Additional detail is given in the references provided.

Reliability Several types of reliability may be calculated.

Internal consistency is usually calculated using Cronbach's alpha, ranging from 0 to 1. Most researchers consider 0.7 as a minimally acceptable criterion. If a measure has more than one dimension or scale, it is usual practice to calculate alpha for each scale. Some statistical programs also calculate the correlation of each item with the scale total (having removed that item from the scale). This information allows the researcher to see if individual items are much less affiliated with the scale being investigated than others. The results of the internal consistency analysis may be compared with the results from other applications.

Test-retest reliability assumes stability of the construct being measured over some minimum amount of time, although that assumption may not be valid in all cases. The survey questionnaire is given a first time, and then after sufficient time has passed that respondents will not remember their answers, but their situation relative to the construct being measured is unlikely to have changed, the questionnaire is given again. Measures of test-retest reliability such as correlation coefficients are then computed. The size of the correlation you expect should depend on how stable you expect the survey to be. A prudent value for a test-retest measure of association would be .6 or higher.

Inter-rater reliability analyses are required if non-survey items are added to the Quality Report, such as information extracted from clinical case records or, in the case of surveys, open-ended questions, and those responses are to be coded so that Likert scale items correspond to those codes (See Phase V). If the criteria for each level of coding are spelled out explicitly in a scoring manual and raters are trained to use the manual, it is likely that different raters will score responses in a similar way. However, it is necessary to demonstrate that agreement with an empirical test. Often, some number of case records or surveys, minimally 25-30, are scored by two or more raters. Then, by performing an analysis similar to that used in repeated measures designs, it is possible to show that the variance due to raters is very small. As with the coefficient alpha, a statistic with range between 0 and 1 is calculated, where 1 means perfect agreement. If you do not have experience with this type of reliability testing, it may be useful to consult with a psychometrician.

Validity

Content validity of the MHSIP Quality Report has been established by including items based on the domains, concerns and indicators described in the Final Report of the MHSIP Quality Report Workgroup. In the case where an organization wishes to add or delete items, it will be necessary to show the correspondence between the revised survey and the domains, concerns and indicators. In some situations, local agencies may wish to expand certain sections of the survey because of special interest in these areas. In that case, the analyses described in the reliability section above must be performed.

Predictive validity of MHSIP scores might be tested by examining whether low scores predicted persons switching providers or plans or filing grievances.

Discriminant validity of MHSIP scores might be tested by examining whether persons who made complaints about their care scored lower than persons who did not, or whether scores for groups of persons served by different organizations differed.

Convergent validity of MHSIP scores might be tested by examining whether other performance measures (e.g., HEDIS mental health measures) correlated with MHSIP scores. For this type of test, it is important to be confident that the two performance measures are intended to measure closely related concepts.

Step 4 Control For Differences In Populations Served (Case-Mix Adjustment) If your performance measurement plan involves comparing entities such as health or behavioral health plans or provider groups with one another or with benchmark data, you should control for differences in persons served. Controlling for such differences is often referred to as "risk adjustment" or "case mix adjustment." Controlling for differences in

populations served will increase your confidence in attributing differences in performance indicator scores to the performance of the service provider organization or MCO. Some experts recommend only adjusting for variables that correlate with both group membership and dependent variables. We recommend controlling for any variables observed or theorized to differentiate among groups.

There are at least three major approaches to case-mix adjustment:

- *Subgroup analyses*: In this approach, only subgroups that are similar in theoretically indicated ways (e.g., diagnostic groups) are directly compared.
- *Regression (or analysis of covariance)*: These approaches use regression methods to compare populations by statistically controlling for the effects of variables that distinguish groups before comparing groups scores.
- *Propensity scores*: In this method, variables that distinguish groups are statistically identified using methods like logistic regression. Then individuals are given propensity scores (literally propensity to be in one group or another) based on their values for these variables. Then scores between groups are compared only for persons with similar propensity scores.

It cannot be emphasized too strongly that comparing entities on performance measures is usually not valid unless some form of case-mix adjustment is implemented. For more detailed information on risk adjustment methods, a companion to this Toolkit provides the necessary knowledge for someone with a basic understanding of statistics to conduct a case-mix adjustment analysis with data such as that provided by the MHSIP Quality Report. The toolkit, entitled *Case-Mix Adjustment in Behavioral Health Care* is available from the Evaluation Center@HSRI (www.tecathsri.org).

Phase IV. How do I Use and Report MHSIP Quality Report Information?

PHASE IV STEPS

- Step 1. Determine the Purpose of your Quality Report Project
- Step 2. Establish a Planning Process for the Report
- Step 3. Identify and Engage Your User Groups
- Step 4. Determine the Content of Your Report
- Step 5. Establish Standards for Reporting
- Step 6. Determine the Format for Presentations
- Step 7. Decide On Time and Frequency of Reporting
- Step 8. Market Your Results

The way in which Quality Report information should be presented depends upon the purpose for which you intend to use it. A lesson learned from experience with the Version 1 MHSIP Report Card is that organizations varied extensively in the formats of their reports and in the information contained in the report. Accordingly, the following section discusses, first, the possible uses of the Quality Report and then various considerations in representing the information for those purposes.

Step 1 Determine the Purpose of your Quality Report Project An early activity of the Quality Report Project Committee should be to determine the purposes for conducting the project and the stakeholders who will use the results or otherwise be affected by the Report. Among the potential uses of the Quality Report are the following:

Program evaluation: The Quality Report may be used to identify the performance of the system as a whole (comparison with benchmarks) or of components of the system. For example, a number of the measures from the Version 1 Report Card were used in the common protocol of the SAMHSA-funded multi-site study entitled “Impact of Managed Care on Vulnerable Populations.” When using the Quality Report for this purpose, however, you should very clearly specify the goals of the evaluation and probably should maintain additional sources of information.

Quality improvement: The Quality Report may be used for QI by identifying areas for potential quality improvement projects, examples of best practices, and benchmarking to assess the impact of quality improvement activities.

Accreditation: Most accreditation organizations employ some form of consumer survey. The National Association of State Mental Health Program Directors (NASMHPD) Research Institute has developed a version of the MHSIP survey for inpatient units that has been adopted by Joint Commission on the Accreditation of Health Care Organizations (JCAHO) ORYX system (<http://www.rdm.org/nripms>). It is likely that the modular design of the MHSIP Quality Report will enhance its suitability for use by accreditation agencies.

Performance based contracting: State Medicaid agencies and the federal Medicare program use consumer surveys for purpose of monitoring performance by contracting provider organizations. In some cases these have been tied to provider incentives. The MHSIP Quality Report is well suited for this purpose, especially for providers of services to the population of persons with serious mental illness, for whom it may capture elements of performance overlooked by performance measurement systems oriented toward a more general population.

Accountability: The MHSIP Quality Report is suitable for accountability at many levels of a mental health system, and Version 1 has been widely used for this purpose. At the federal level many elements of the MHSIP Report Card have been incorporated into CMHS' Uniform Reporting System.

Public health: Quality Report information can be useful for public health agencies and organizations for monitoring patterns of behavioral health services utilization and disease.

Advocacy: In recent years, a considerable amount of literature has appeared describing methods for effectively influencing policy. A consistent recommendation, based on research, is that policy makers respond most directly to credible data presented in a clear and useful way. The addition of measures of system support of recovery and cultural competency will contribute to the effectiveness of advocacy efforts in these areas.

Consumer choice: Consumer choice, based on performance reports, offers the potential to harnesses indirect but powerful incentives (market forces) to improve system performance. Since the publication of the version 1 MHSIP Report Card, researchers have produced a considerable body of information on the how to report performance measurement results in ways that enable consumers to choose care that is of the highest quality or the most appropriate for their individual circumstances and preferences.

Management and decision support: The MHSIP Report Card is an integral component of SAMHSA's Decision Support 2000+ initiative (<http://www.rdm.org/nripms>). When fully operational, this system will allow organizations to generate standard and custom reports to guide program planning, quality improvement and benchmarking. Quality Report information provides a means of monitoring over- or under-utilization of services and other forms of provider behavior and performance.

Policy analysis and legislation: The Quality Report provides valuable information to support policy makers in making decisions about resource allocation, public program structure, disparities in care for ethnic minorities, and a host of other policy issues.

Step 2 Establish a Planning Process for the Report As noted in Phase I, preparations for this step should occur early on in the Quality Report planning process. At that point, you may find it worthwhile to establish a Reporting Taskforce as a sub-group of the Steering Committee. Subsequent steps are analogous to those for the Steering Committee, described in Phase I, i.e. deciding on the Taskforce's functions, structure and composition. Finally, the Reporting Taskforce will need to make a number of decisions regarding the nature of the report(s) to be produced.

The following are some questions you and/or the Reporting Task force will need to address in planning for reporting results:

- What are the purposes of the report?
- Who will receive the report?
- Will the same report be sent to all audiences or will certain persons or groups receive specialized reports?
- If you are reporting results by organizations, will these have the opportunity to review the reports prior to public distribution?
- Are your results subject to immediate release at any time under freedom of information laws or regulations?
- Will you prepare a narrative report or just report data in the form of charts, graphs and/or tables?
- At what level of detail should results be reported?
- What will your policy be regarding requests for reports of additional analyses or variations on standard presentation? How will you obtain feedback on the report before full dissemination?
- How will you distribute the reports?
- How will you obtain feedback after dissemination, and what will you do with it?

Below we provide guidance in answering the above questions and others regarding reporting performance measurement results.

Step 3 Identify and Engage User Groups Identifying and engaging the groups that will use the information you will provide will help you focus on the essential goals of the report. It is important, therefore, to establish a process for obtaining feedback from these groups. The process should employ protocols that demonstrate you are hearing and responding to their recommendations.

Since performance measurement projects typically have multiple audiences, it is especially important to design a process and content for reporting results that is responsive to the particular needs and desires of important audiences. For example, health care organizations increasingly include consumers, family members, and other lay persons as active members of planning and quality improvement boards and committees. Lay persons increasingly demand reports that are clear, accurate, and simplified. The central role of diverse audiences also suggests that it is important to involve these persons in the planning of performance measurement reporting.

Focus groups are an effective mechanism for soliciting stakeholder needs and preferences by serving as sounding boards for your ideas on reporting results. Focus groups should be homogenous (i.e., composed only of consumers, or only of providers, etc and should meet periodically to review developments in your reporting plans.

Step 4 Determine the Content of Your Report Decisions about report content are crucial to the

success of a performance measurement initiative; therefore, decisions about report content should be considered early in the design phase of the Quality Report project. Mechanisms that might be helpful in eliciting ideas about useful content include key informant interviews with knowledgeable stakeholders from each group, focus groups of stakeholders, and information interviews with persons from other states who have more experience.

Decisions about report content should be driven by several considerations including:

- The use to which different groups will put the information (e.g. contract monitoring, CQI, advocacy, etc.)
- The major policy concerns that need to be addressed
- The relevant comparisons that should be featured (e.g. trends over time, performance vis-à-vis standards or benchmarks, relative performance across organizations, etc.)

Few readers will be able to make productive use of undigested information such as voluminous data tables they must wade through to get the specific information they seek. This calls for attention to what information should be included in reports intended for specific audiences.

The simplest data presentation will often meet the needs of most mental health system stakeholders. In most cases, brief, targeted presentations with clear summaries highlighting major points are more effective than long, detailed reports. Presentation of response rates as percentages by domain and by individual item is the most basic and understandable way in which to communicate survey results. The information may then be presented in more complex ways as the need arises. Reports of any kind should describe limitations including possible sources of bias, issues of data quality, and generalizability of results.

Step 5 Establish Standards for Reporting A key element in enhancing credibility and utility of MHSIP

Quality Report results is to provide clear and comprehensive descriptions of the methods used collecting and analyzing the information being reported.

Survey response rates. Adequate response rates can be difficult to achieve with the population of persons who use behavioral health services, particularly those with serious mental illness. Concerns about stigma may inhibit response. Reliable and adequate contact information is often difficult to obtain. Budget limitations often prevent the use of the most effect methods for enhancing response rates. A lesson learned with Version 1 is that, because of these and other factors, response rates may vary considerably in practice, and often have been quite low (though others have been very satisfactory). Calculating and reporting response rates can be a complex task. The key is to present explicit definitions for categories of non-response (refusals, bad contact information, partial completions, etc.). Even when response rates are less than optimum, clear and comprehensive presentation of the various categories will enhance the credibility and utility of your Quality Report project.

Survey methods: A second lesson learned with Version 1 is that organizations will choose a variety of methods for administering surveys, and that the details of these methods are often difficult to discern from public reports. Though it is likely that methods will be more standardized with the Quality Report, it is important nonetheless for reporting organizations to be clear about the methods that were used for conducting the survey.

Administrative Information. As discussed in Section V, data from an organization's information system is typically collected to serve multiple purposes, and performance measurement is usually not the original or primary one. Any performance report should describe characteristics and limitations of this data, and to present at least an overview of the modifications required to serve the purpose of performance measurement. Accordingly, the Workgroup recommends that, to be considered "official" reports of MHSIP Quality Report results describe at least the following:

- A description of the methods used in collecting the data, including survey administration and extraction from the information system
- A definition of the population the survey is intended to represent (universe)
- The sampling frame (list of eligible survey respondents), its origin, limitations and potential bias
- A description of sample selection procedures, sample sizes, response rates, and any known or suspected biases
- The time period for administrative and survey data collection
- The amount of missing data by domain and survey item
- Response rates, including the number of blank or incomplete forms, with method of calculation and definitions of categories
- The degree to which the data are considered generalizable across the state or local area
- A discussion of what level of difference might be considered *meaningful* for each variable or class of variables presented

Scoring. One area where lack of standardized methods limited the potential use of Report Card findings was in the methods used for scoring Consumer Survey responses. Accordingly, the MHSIP Report Workgroup offers the following recommendations for scoring and reporting information for the "official" MHSIP Quality Report.

- Both domain (summary) and individual item scores should be calculated (though not all may be reported).
- Reports should include specific information on the amount of missing data
- All "not applicable" responses should be treated as missing.
- Domain scores (means or percentages for access, appropriateness, outcomes and satisfaction) should be calculated only if at least two thirds of the items comprising any domain are complete.

- For quality improvement purposes, percentages should be calculated for individual items, regardless of the amount of missing data. The percentage of missing data associated with each item, should however, be reported.

Step 6

Determine the Format for Presentations

Having determined the content of each report, the next task is to design formats for data presentation that help to make your point. The science of how to report health care performance measurement information effectively has grown rapidly in the past few years, supported by the interests of government and business purchasers in promoting consumer choice as a means of driving quality.

Decisions about format should be based on considerations such as the following:

- Different formats are appropriate for different types of data.
- Different audiences will require varying levels of detail in their presentations. For example, audiences interested in highly condensed information may prefer a few striking bar graphs or pie charts. For those interested in item level analyses, however, tabular presentations may be preferable..
- Audiences will vary in the amount of experience they bring to statistical data; presentations should be tailored to the audience's level of expertise.
- Stakeholder needs may change over time as policy issues evolve and as various constituent groups gain more experience in understanding statistical information. Such changes should be monitored by providing opportunities for stakeholder feedback.
- Decisions about report content should be revisited in a systematic fashion. Periodic discussions with stakeholder groups can provide important feedback on whether the reports are serving their intended purpose.

DATA REPRESENTATION

The following section briefly describes common methods of presenting data, with advantages and limitations of each for representing performance measure results.

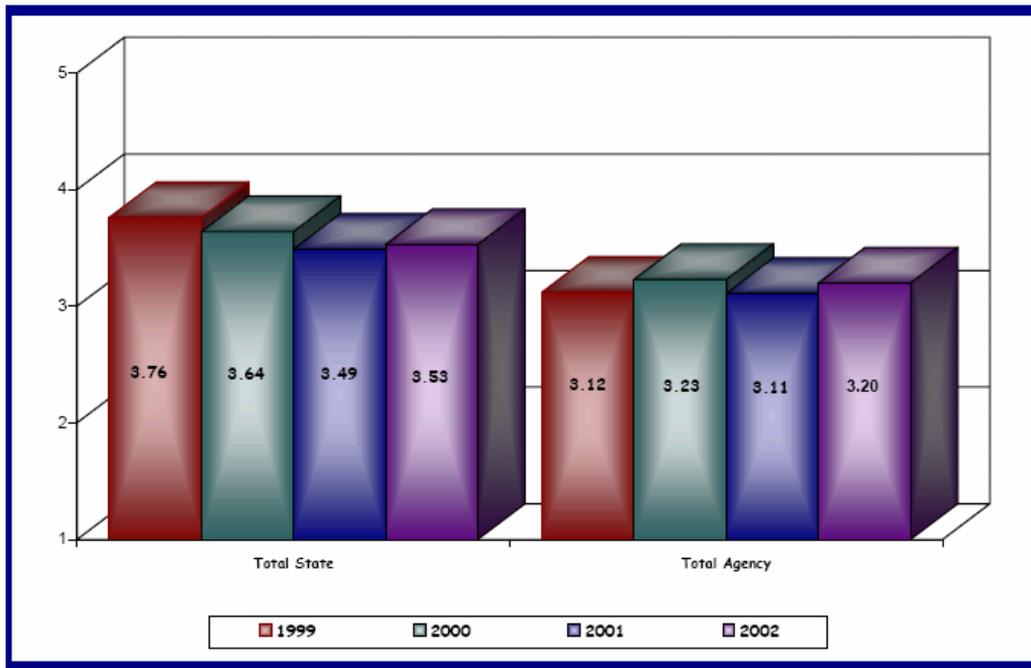
Bar Graphs. Bar graphs are useful for showing sums and averages and are often used to illustrate variable values, using spaces between the bars to differentiate among variables. Bar graphs are simple to read and to make. They can be designed to show both numbers and frequencies. They can vary in the amount of space between the bars, in the horizontal or vertical display of the bars in two or three dimensions, and in the order of the bars. They can display two or more categories of data by organizing the bars in a group, overlapping, or stacked chart. Horizontal bar charts can be used to address the following circumstances:

1. Variable values with long names - a horizontal graph is the only solution when the variable names will not fit under the vertical bar.
2. Many variable values - as few as 6 to 8 variables makes a vertical chart difficult to construct neatly. A horizontal graph easily makes more space for the variables.

You may wish to add explanatory summary text, especially when the information displayed on the bar graph is complex or potentially confusing.

*Example 1: Three Dimensional Vertical Bar Chart
(from Missouri Department of Mental Health at <http://www.dmh.missouri.gov>)*

Service Means Comparison of 1999, 2000, 2001, & 2002



Comparison of 1999, 2000, 2001, & 2002 Mean Ratings

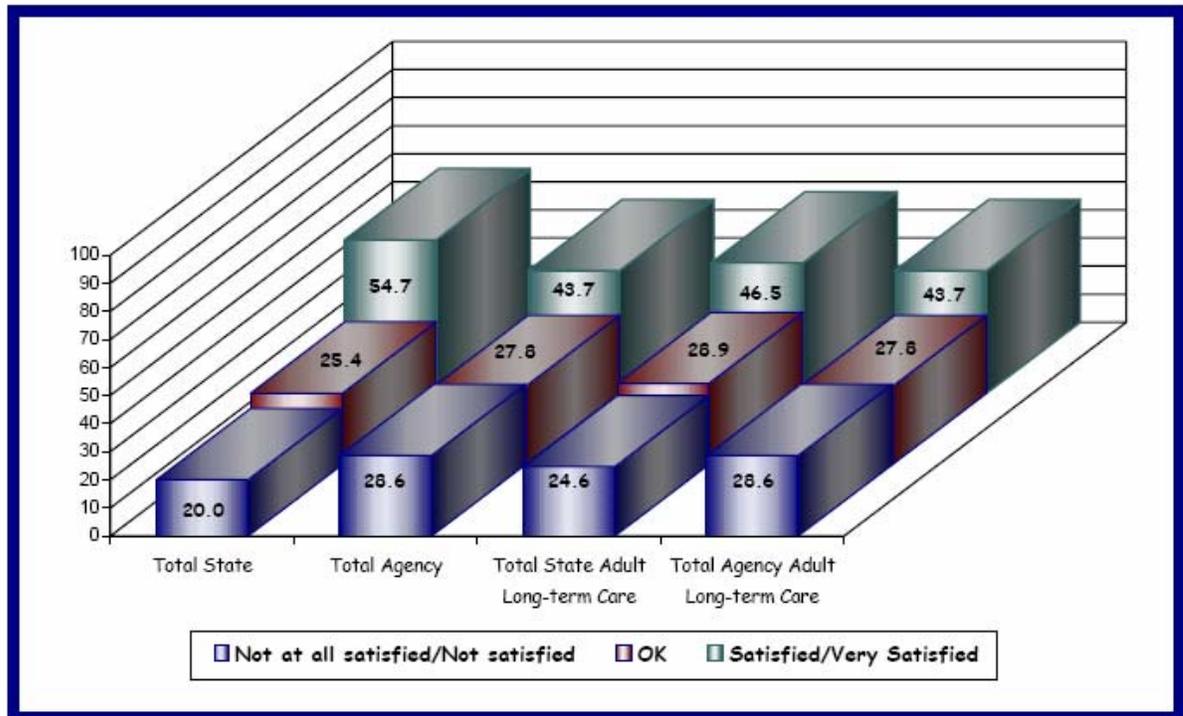
Some of the key findings were:

- For this agency, the mean of the responses to the question " How satisfied are you with the services you received?" was 3.12 in 1999, 3.23 in 2000 3.11 in 2001, and 3.20 in 2002.
- For this agency, the mean of the responses to the service question increased from year 2001 (mean= 3.11) to year 2002 (mean= 3.20).

Grouped/stacked Bar Graphs. Grouped bar graphs are used for displaying two or more categories at a time. Separate bars on the same axis represent the different categories.

Example 2: Three Dimensional Grouped Bar Chart
(from Missouri Department of Mental Health at <http://www.dmh.missouri.gov>)

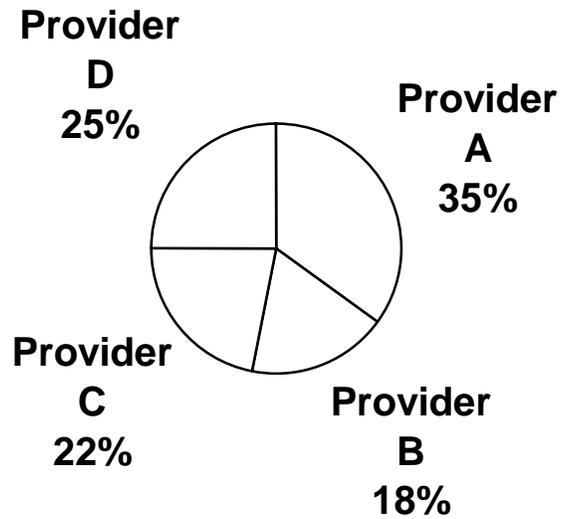
Overall Satisfaction with Services



Program Satisfaction: Percent of responses to the question "How satisfied are you with the services you receive?"

Pie Charts. Pie charts are used when you want to focus on proportions of quantitative variables. If you have several groups that you would like to compare then it is useful to draw a pie chart for each group.

*Example 3: Pie Chart
Percent of Total Enrollees by Provider*



Tables. Tables can be used to display performance results of multiple organizations or of a single organization on several domains. The major advantages of tables are their ease of construction, straightforward interpretation, and ability to read actual numbers. Tables are not terribly visually appealing, however, and can be overwhelming to a reader if they are large. In general, tables are most appropriate when the reader is interested in detailed results and/or there is a large number of organizations.

Example 4: Tables

(Florida Department of Health and Human Services at www5.myflorida.com/cf_web/myflorida2/healthhuman/substanceabusementalhealth)

**ACCESS 4:
The percentage of consumers who report that physicians, mental health therapists, or case managers can be reached easily.**

Target Population: Adults with Mental Illness

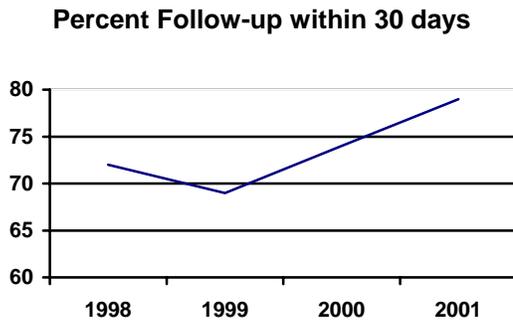
Measure: Response to question #25 "I was able to talk with staff when I needed to do so".

Source: Behavioral Healthcare Rating Scale (BHRS) data in the Alcohol, Drug Abuse and Mental Health Data Warehouse (ADMDW)

BHRS #25	Forensic	Severe and Persistent	In Crisis	Total
1 disagree strongly	1	184	130	315
2 disagree	2	225	174	401
3 disagree a little	3	330	226	559
4 agree a little	10	919	505	1434
5 agree	57	5273	2302	7632
6 agree a lot	47	3854	1917	5818
Number agreed	114	10046	4724	14884
Number disagreed	6	739	530	1275
Percent perceiving easy access	95%	93%	90%	92%

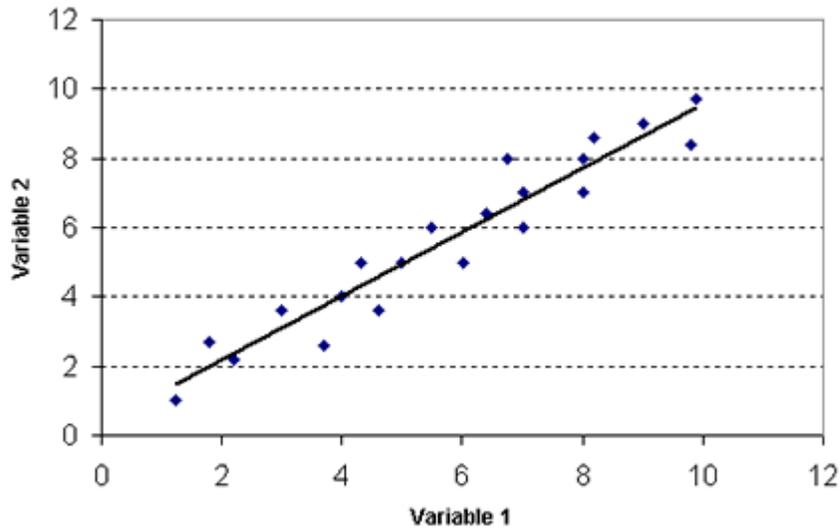
Time Series Charts. When you want to display data in a time series it is important to design a chart so that it is easy to make comparisons over time. It is difficult to detect patterns of change over time in tables. Line charts are suitable when you want display changes that are large and when there are turning points. Line charts are easy for the eye to follow and clearly display distinct changes over time.

Example 5: Time Series Chart



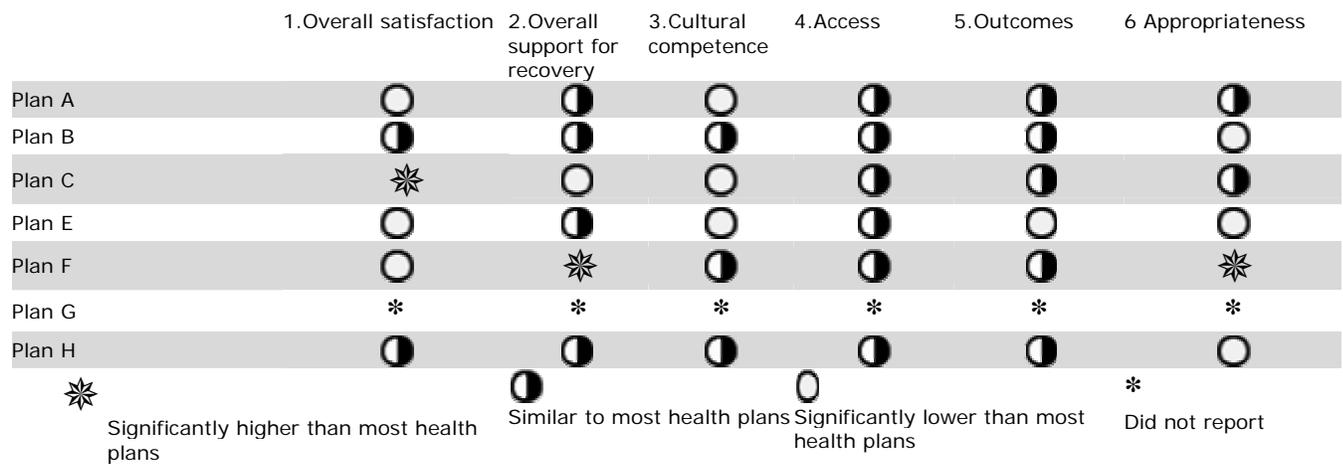
Scatter plots. Scatter plots are used to show the relationship between two quantitative variables. The data consist of paired coordinates (x,y) each indicated as a dot on the graph. A curve in the chart is often included to show an estimated regression. This gives an overall picture of the relationship between all the dots on the graph.

*Example 6: Scatter plot showing positive relationship between two variables
(From Statistics Canada <http://www.statcan.ca>)*



Stars or Other Symbolic Representations of Performance. This type of reporting structure is common among ratings of movies, restaurants, colleges and the like because of its simplicity and visually appealing qualities. Simply, this type of reporting is the display of a number of stars or other symbol next to a variable or organization based on the organization's performance or score on the particular variable. This type of report is familiar to most and is the basis of so-called "five-star" restaurants, resorts, etc. Different symbols can be used to denote different areas of performance. For example, stars may be used to represent overall performance, smiling faces may be used to represent consumer satisfaction, houses may be used to represent housing outcomes, and clocks may be used to represent waiting times.

Example 7: Star Chart



Step 7 Decide on Time and Frequency of Reporting Different user groups may have different needs as far as frequency and timing of reports. A legislator, for example, may be most interested in a report that is concomitant with the fiscal year and is released soon before the budget appropriations period. A program manager, alternatively, may be interested in more frequent reports for quality improvement purposes and in timing of reports that help with preparation of proposals.

Coordinate with data collection: Examine how the timing and frequency of reporting aligns with the planned frequency and timing of data collection. Without careful coordination of data collection and reporting, you may find yourself without any data to report at a time when a report is due.

Plan for production and distribution. Decide what your production schedule is and who is responsible for each area. Here a detailed chart and timeline would be appropriate and helpful. Determine who is responsible for:

- Data flow
- The report development schedule
- How the report will be used

Step 8 Market Your Results Research has shown that simple dissemination of quality information alone is seldom sufficient to stimulate quality improvement action. Without broad acceptance and support, even the most careful planning and skillful technical work will be ineffective. The value of performance measurement systems is not readily apparent to all who are involved in or affected by the reporting organization. In addition, given their costs, performance improvement projects must compete with other priorities for resources. Finally, performance measurement information, particularly if it is not fully understood, may create anxiety about system change, resulting in resistance. For these reasons, it is important that you develop a deliberate strategy for marketing your performance measurement system to system stakeholders.

One developing area of research for this kind of activity is “social marketing.” Originating in the field of health care communication, social marketing is simply the adaptation of techniques developed in commercial marketing to promote ideas, attitudes and behaviors. Kotler and Zaltman who developed the concept describe it as “differing from other areas of marketing only with respect to the objectives of the marketer and his or her organization. Social marketing seeks to influence social behaviors not to benefit the marketer, but to benefit the target audience and the general society.” (cited by Weinrech, NK “What is Social Marketing?” available at <http://www.social-marketing.com/Whatis.html>). A key element of social marketing is to systematically develop an understanding of the audience’s needs and interests and then adapt the information you provide in ways congruent with them. More information about these techniques is available from the following sources:

RESOURCES ON SOCIAL MARKETING

Weinrech, N.K. Hands-On Social Marketing: A Step-by-Step Guide (Sage Publications, 1999)

Medicare Issue Brief: Social Marketing <http://www.medicareed.org/content/CMEPubDocs/V3N5.pdf>

QUALITY REPORT RECOMMENDATIONS FOR REPORTING RESULTS

- Be clear about the actual purpose and use of the data and how it should be reported and disseminated.
- Identify and engage the groups that will use the information you will provide. It is also important to establish a method of obtaining feedback from these user groups.
- Consider focus groups as "sounding boards" for your ideas on reports. It may be advisable for them to meet periodically as you modify your reporting plans.
- Establish a mechanism to obtain constant feedback from the user groups as to the quality of the of the information being received and any changes that may be needed to fine tune the process.
- Develop a flow chart that indicates clearly how data will move through the system.
- Make decisions about report content in the design phase of the performance measurement project. Mechanisms that might be helpful in eliciting ideas about useful content include key informant interviews with knowledgeable stakeholders from each group, focus groups of stakeholders, and information interviews with persons from other states who have more experience.
- Include a discussion of the strengths and limitations of the data presented.
- Conduct periodic discussions with stakeholder groups to gain feedback on whether the reports are hitting the mark.
- Design formats for presenting information that help to make your point efficiently.
- Determine how often and when reports are needed.
- Examine how the timing and frequency of reporting aligns with the planned frequency and timing of data collection.
- Decide what your production schedule is and who is responsible for each area.

Phase V. How may I Modify and Enhance the MHSIP Quality Report?

PHASE V STEPS

Step 1. Consider Issues Involved In Modification Of The Quality Report

Step 2. Decide on Enhancements to the Quality Report

Step 1 Consider Issues Involved In Modification of The Quality Report The MHSIP Quality Report was developed for public use to benefit consumers and others interested in the quality of behavioral health services. Thus, no proprietary protection or prescriptive authority exists to regulate its use. One consequence of this approach is the potential for it to be modified in ways that would produce results inconsistent with its intended purpose.

Rather than attempting to exercise control over usage, the Workgroup's approach to this issue is simply to assert that only the full Quality Report as produced in the Final Report will be acknowledged as the "official" version. Accordingly, we recommend that the consumer survey be used without modification to insure the integrity of the underlying structure of domains and concerns, and to enhance the value for benchmarking and other types of comparison as discussed in section III of the Toolkit. Moreover, we strongly recommend that organizations implement the entire Quality Report (both the survey and administrative measures), as individual components designed to perform as parts of a comprehensive overview of a behavioral health system.

Step 2 Decide on Enhancements to the Quality Report Concerns raised by modifications of the Quality Report are much less significant in the case of enhancements, i.e. components added to address particular issues outside the focus of the Quality Report itself. This occurred quite extensively and successfully with the Version 1 Report Card. The following are some examples of ways in which the Quality Report may be enhanced to serve additional organizational priorities or stakeholder interests.

Qualitative evaluation: A major trend in the program evaluation field in recent years has been the development of sophisticated theory and practice of "mixed methods," i.e. the combination of quantitative and qualitative evaluation. Program evaluation is one potential use of the MHSIP Quality Report, but its use for this purpose probably requires supplemental information. Qualitative evaluation is an excellent means of obtaining this information. A number of textbooks and journal articles describe methods of qualitative evaluation to add context to performance measurement, understand exceptional (positive or negative) findings, and explore needs and opportunities for quality improvement first identified by performance measurement data.

Open-ended questions: Open-ended questions appended to a performance measurement system are one form of qualitative evaluation. Many organizations supplemented Version 1 of the MHSIP Report Card with opportunities for respondents to comment on various aspects of system performance. This method is particularly useful for addressing specific local concerns that may not be captured by the generic Quality

Report, such as recent policy initiatives or organizational changes. Responses to open-ended questions can be difficult to synthesize, and they can not be regarded as representative of the entire population, but if they are properly phrased to elicit specific information or opinions they can be very useful for program planning and quality improvement.

Reports of abuse, rights-violations and service complaints. A number of organizations using the Version 1 MHSIP Report card provided an option for reporting negative experiences with the mental health system. While no formal assessment of the value of this is available, many organization administrators feel that it provides an important alternative channel to formal grievance procedures, which some consumers may be reluctant to use for a variety of reasons.

Measures of symptoms and functioning: A number of organizations have supplemented the Version 1 Report Card with various measures of symptoms and functioning. These enhance the potential to use the Report Card for outcomes measurement, but they also add to the burden of completing the form and may dissuade some consumers from responding. The science of outcomes measurement is highly complex and still in a process of development. While it is not within the scope of this Toolkit to discuss all of the issues involved, you should carefully consider all the implications using this as a tool for outcome assessment beyond the measures included in the survey. For example, results are likely to be biased, since survey response is probably influenced by functional level or symptom severity. On the other hand, this information offers an opportunity to examine experiences of care among different patient sub-groups.

Phase VI. How do I Evaluate and Improve a MHSIP Quality Report Project?

PHASE VI STEPS

- Step 1. Define The Purpose Of The Evaluation
- Step 2. Define the Evaluation Questions
- Step 3. Assess the Quality Report Implementation Process
- Step 4. Assess the Quality of the Data Collected
- Step 5. Assess the Cost/Burden of the Performance Measurement System
- Step 6. Define The Purpose Of The Impact Evaluation.
- Step 7. Define the Impact Evaluation Questions

This chapter describes methods for evaluating both the process of developing a performance measurement project and the results of this process. These two overarching goals lead to two distinct types of evaluations of a performance measurement system: a *process evaluation*, and an *impact evaluation*. The process evaluation has as its overall goal to assess each step in the development and implementation of a performance measurement system in terms of inclusion, efficiency, resource expenditure, etc. The impact evaluation has as its overall goal to assess the degree to which the performance measurement system has achieved its goals, which may include things like assisting consumers in making informed health care choices, informing legislators in making resource allocation decisions, facilitating quality improvement within service provider agencies, etc.

As with any type of evaluation, the planning of an evaluation of a MHSIP Quality Report project should begin early, concurrent with the planning of the other aspects of the project. The scope and content of the evaluation will, of course, vary with each project. Below we present a structure for thinking about the evaluation purposes, evaluation questions that may be relevant, and some methods that might be used to address such questions.

THE PROCESS EVALUATION

Step 1 Define The Purpose Of The Evaluation It is crucial to begin the evaluation process with a clear and shared understanding of the purpose(s) for the evaluation. The defined purposes should shape the scope and content of the evaluation and ensure that evaluation resources are allocated efficiently. We present and discuss several potential purposes of a process evaluation.

Providing Feedback to Improve the Process Used in Developing the Performance Measurement System. This is usually the primary purpose of a process valuation. This purpose builds on ideas found in concepts like continuous quality improvement (CQI) and formative evaluations (Rossi & Freeman, 2003).

Assessing Stakeholder Satisfaction and Buy-in with the Process. Since stakeholder commitment is so central to a successful PM system, this should be a major goal of your process evaluation. Once again, it should be formative, providing constant feedback to the performance measurement system administrators so that fine

tuning can take place before small problems become large. Within this goal, you may want to examine at least two aspects of stakeholder satisfaction: 1) Involvement in the process, and 2) Quality of the products.

Modeling the Use of Information for Improvement. If performance measurement is based on the tenet that evaluative information can improve service delivery, then the performance measurement system itself should use evaluative information to improve its own processes.

Analyzing Cost-Benefit. The process evaluation may include an analysis of the Cost of Developing and Implementing the Performance Measurement System in Relation to Benefits Expected

Comparing the Development Process with National Models It may be useful to make this an explicit evaluation purpose to help remind yourself to take advantage of the thinking and work that has already been done.

Step 2 Define the Evaluation Questions The evaluation questions relevant to any particular process evaluation will vary depending on the characteristics of the performance measurement system being developed. Nonetheless, below we suggest some questions that might be addressed in a process evaluation.

Questions Related to Performance Measurement System Development:

- Are stakeholders satisfied with their level of involvement in the process?
- Is the development process efficient?
- Are there clear lines of responsibility for tasks?
- How much time does the process consume? (This should be examined for both research/evaluation staff and other stakeholder groups.)

Step 3 Assess the Quality Report Implementation Process

The evaluation should address the following aspects of the implementation process:

- *Technical assistance:* Is sufficient technical assistance provided to persons and organizations required to participate in the performance measurement system? Is it provided to the right individuals/organizations? Is it appropriate for its intended audiences?
- *Training:* Is sufficient training provided? Is it provided to the right individuals/organizations? Is the level of training appropriate for its intended audiences? Are there appropriate mechanisms to ensure that training diffuses through the system?
- *Barriers to implementation:* What are the barriers encountered in attempting to implement the system? How are these addressed?
- *Problem resolution:* How effectively does the system resolve problems related to implementation? Are problems addressed at the most appropriate system level?

- *Protocol Adherence*: Are protocols that have been designed for the performance measurement system followed rigorously?
 - Are samples drawn according to the sampling plan?
 - Are data collected during the specified time period?
 - Are data collected for all of the performance measures included in the planned system?
 - Are protocols related to privacy, confidentiality and informed consent followed?

Step 4 Assess the Quality of the Data Collected

The following are questions to address in assessing data quality:

- Have the psychometric properties of the data been assessed? Are the data reliable and valid?
- Do the data meet acceptable standards of completeness, accuracy and timeliness?
- For existing data sets, have appropriate quality checks been conducted to ensure accuracy and completeness of data?
- Have the methods used to integrate data across multiple organizational levels and data sources been checked to ensure that these procedures produce accurate data?

Step 5 Assess the Cost/Burden of the Performance Measurement System

The process evaluation should consider at least the following aspects of cost and burden:

- Resources required from the central office/providers for:
 - Training and technical assistance
 - Quality monitoring
 - Data collection
 - Data entry/management
 - Data analysis/reporting
- Costs associated with the process of developing the performance measurement system
- Burden placed on various stakeholder groups?
- Burden placed on staff/consumers?

THE IMPACT EVALUATION

Step 6 Define The Purpose Of The Impact Evaluation The purposes of an impact evaluation should follow from the specific goals of the performance measurement system as articulated early in the process of planning and development. Phrased in terms of evaluation purposes, they are:

- Assessing the degree to which the performance measurement project effected changes in service delivery
- Assessing the degree to which the performance measurement project provided information useful to legislators, government officials, and plan administrators for making resource allocation decisions
- Evaluating the utility of information from the performance measurement system for contract monitoring and management
- Evaluating the extent to which consumers used the information produced in the performance measurement system in selecting plans, providers, and/or services

An additional purpose of the impact evaluation, not specifically related to the goals of the performance measurement system, is related to the appropriateness of performance measurements selected:

- Evaluating the association of performance measures with other, typically longer-term, measures of outcome or system effectiveness

This type of evaluation is a kind of convergent validity testing mentioned in Phase III. Sequentially, this type of evaluation necessarily follows data collection and analysis. Although not strictly an evaluation of the performance measurement system's impact, this type of evaluation may shed light on whether some structure or process measures are good indicators. For example, Druss and Rosenheck (1997) evaluated an individual HEDIS measure of continuity of care, percentage of persons who received a follow-up appointment within 30 days of hospital discharge, by evaluating the correlation between this measure and future rehospitalization.

Step 7 Define the Impact Evaluation Questions Again, the specific evaluation questions will follow from the intended impact of the performance measurement system. We suggest some more general questions that might apply to many impact evaluations:

Assess the acceptability of the performance measurement system

- Do various stakeholders find the information generated by the performance measurement system meaningful and relevant?
- Does the information help consumers make choices about plans, providers, and services?
- Is the information produced culturally sensitive?

Assess the utility of the information produced

- Are the reports produced adequate?
- Does the content address the concerns of each stakeholder group?
- Is the format appropriate for the content and the level of expertise of various audiences/
- Are reports generated in a timely manner?

Assess measurable changes in the system with respect to:

- Program management and improvement
- Contracting
- Accountability for public funds
- Local planning
- State level planning
- Treatment planning
- System level policy changes

METHODS FOR EVALUATING A PERFORMANCE MEASUREMENT SYSTEM

Below, we present a matrix that suggests methods for addressing the questions that might be the focus of an evaluation of a performance measurement system. Two other strategies, not included in the matrix of methods, are case studies based on specific uses of performance measurement information and comparisons of systems with and without access to performance data. Case studies can provide a widely accessible report that highlights a particular outcome of the performance measurement system. Comparative qualitative or quantitative studies of systems, one with and the other without a performance measurement system may highlight advantages of implementing a performance measurement system that are undetectable without an external comparison.

Methods for Evaluating a Performance Measurement System, by Type of Evaluation Question

Evaluation Question	Focus Groups	Interviews	Surveys	Chart Reviews	Data Audit	Logs/project records	Statistical analysis of data
Design Process	X	X					
Level of involvement	X	X	X				
Efficiency of the process		X				X	
Quality of implementation	X	X				X	
TA provided	X	X	X				
Barriers encountered	X	X	X				
Problem solving effectiveness	X	X	X				
Protocol adherence		X			X		X
Sampling					X		
Timing					X		
Confidentiality	X	X	X				
Quality of the research	X	X					
Reliability/validity							X
Completeness, etc				X			X
Cost/burden		X				X	
Ease/burden	X	X	X				
Resources						X	
Acceptability	X	X	X				
Utility	X	X	X				

*QUALITY REPORT RECOMMENDATIONS FOR EVALUATING THE PERFORMANCE
MEASUREMENT PRODUCT*

- Begin the evaluation process with a clear and shared understanding of the purpose(s) for the evaluation
- Consider both process and impact evaluations
- Include as a major goal of the evaluation process to assess stakeholder satisfaction because of its importance to the success of the performance measurement system
- Use case studies based on specific uses of performance measurement information
- Conduct comparisons of systems with and without access to performance data

Appendix I. Links for Additional Resources

(NOTE: Links to websites frequently change. These will be confirmed and updated in subsequent versions of the Toolkit).

RELATED INITIATIVES AND ORGANIZATIONS

Mental Health Statistics Improvement (MHSIP) Online
<http://www.mhsip.org>

Substance Abuse and Mental Health Services Administration (SAMHSA)
<http://www.samhsa.gov>

Decision Support 2000+ Online
<http://www.ds2kplus.org>

The Forum on Performance Measures in Behavioral Health and Related Systems
<http://www.mhindicators.org/>

METHODOLOGY

Research Randomizer
<http://www.randomizer.org>

Sample size calculator
<http://www.surveysystem.com/sscalc.htm>

American Association of Public Opinion Research Guidelines for Reporting Survey Results
<http://www.aapor.org>

Toolkit on Risk Adjustment in Behavioral Health (Michael Hendryx, produced by The Evaluation Center@HSRI)
<http://www.tecathsri.org/pubs.asp>

STANDARDS FOR PUBLICLY REPORTING SURVEY METHODS:

American Association for Public Opinion Research (AAPOR):
www.aapor.org

National Council on Public Polls(NCPP):
<http://www.ncpp.org/disclosure.htm>

Council of American survey researchers (CASRO):
<http://www.casro.org/codeofstandards.cfm#clipublic>

GUIDELINES FOR DESIGNING HEALTH CARE QUALITY REPORTS

Agency for Healthcare Research and Quality (AHRQ) “Talking Quality”
<http://www.talkingquality.gov/>

STATE MHSIP REPORT CARD REPORTS ON-LINE

California:
<http://www.dmh.ca.gov/RPOD/adult-crp.htm>

Delaware MHSIP survey:

<http://www.state.de.us/dhss/dsamh/ccss.pdf>

Florida Consumer Survey:

http://www5.myflorida.com/cf_web/myflorida2/healthhuman/substanceabusementalhealth/publications/consumer.html

Hawaii Consumer Survey:

http://www.state.hi.us/doh/camhd/reports/satisfaction_survey.pdf

Kansas:

<http://www.srskansas.org/hcp/MHSIP/AdultReports.html>

Kentucky:

<http://dmhmrs.chr.state.ky.us/mh/outcomes/>

Minnesota:

<http://edocs.dhs.state.mn.us/live/DM-0004-ENG.pdf>

Missouri Consumer Survey:

<http://www.modmh.state.mo.us/pm2001/survey.htm>

Nevada:

<http://mhds.state.nv.us/pdfs/ConsumerSurvey03.pdf>

New Mexico:

http://www.state.nm.us/hsd/mad/pdf_files/SALUD/MHISPRpt2000_2001.pdf

North Carolina:

<http://www.dhhs.state.nc.us/mhddsas/manuals/index.htm>

Oklahoma:

<http://ww1.odmhsas.org/eda/cmhcpcindicatorsfy01.pdf>

Oregon:

<http://www.dhs.state.or.us/mentalhealth/publications/2002adultsurvey.pdf>

South Dakota:

<http://www.state.sd.us/dhs/dmh/>

Vermont:

<http://www.state.vt.us/dmh/Data/consumersatisfactionrpts.htm>

Wyoming

http://mhd.state.wy.us/infonetwork/mhsip_survey.html

Appendix II: Data Matching

(by Tracy Leeper, Rebecca Moore, Bernadette Phelan, John Pandiani, and Steve Davis)

The data sources for the MHSIP Quality Report are primarily surveys and mental health agency administrative datasets. However, another rich source of data about people receiving mental health services is often available to State Mental Health Authorities (SMHAs) seeking to evaluate the impact of the funds they allocate on the people they serve. Administrative data collected by non-mental health government agencies can provide useful, objective information about outcomes that are very important to consumers, e.g., employment, criminal justice system involvement and mortality. As state and local budgets shrink and priorities for spending focus more on service delivery than on data collection and reporting for evaluating those services, matching administrative datasets can be an economical alternative means to evaluate the impact of services while reducing the data collection burden on agency staff and consumers, and maintaining data privacy.

Data matching, as discussed in this toolkit, is also referred to as cross-database analysis, secondary data analysis, or administrative data analysis. It is usually performed using data collected by two or more separate agencies for administrative purposes other than those for which the data matching analysis is conducted. To perform data matching, it is necessary to have some form of record identifier that is common to two (or more) sets of data. Typically, the first dataset represents people receiving a service, the value of which stakeholders wish to determine. The second dataset represents events that have occurred in the lives of people, some of whom may be the people served in the first dataset. The aim of matching is to link or overlay these datasets to determine the extent to which people served and not served have had the event (outcome) of interest occur in their lives (or the extent to which the event occurred before and after a group was served).

This toolkit provides information about “what to do” about data matching and makes reference to other sources for descriptions of “how to do” various types of matching. Specifically, the toolkit addresses:

1. What types of data matching strategies are available,
2. What are the advantages and limitations of using data matching for program evaluation and other tasks,
3. What lessons have been learned by early adopters that will help you address obstacles you may face, and
4. What resources are available for reference and technical assistance.

Data Matching Strategies. Selection of a strategy for data matching depends on the level of detail in the data that is available and the restrictions on its use. Three approaches will be discussed here: deterministic (DM), probabilistic (PM) and probabilistic population estimation (PPE). Deterministic matching is the most straightforward of the three matching strategies. DM requires only that the two datasets use the same record identifier in the same format. For example, every state collects standard data about employment and income that is organized by Social Security Number (SSN). If SSN is collected as part of the mental health record and the appropriate interagency privacy and security prerequisites are met, then a straightforward match of the

two datasets can be performed using SSN as the key. All records in one dataset that have an exact match on SSN in the second dataset will be linked as pairs. If historical data are available in the employment dataset and the mental health service dataset, then it is possible to assess employment and income levels before and after treatment has occurred, or for various doses of treatment (Interstate Collaborative Study, 2003). The primary advantage of this approach is its simplicity. Disadvantages include the lack of datasets that include SSN and possible mismatches because of transposed numbers, children using their parent's SSN, or people with multiple SSNs.

Probabilistic matching takes advantage of more identifiers, if they are available, to match records between two datasets. Typically, name, date of birth, gender and sex are matched, but SSN or another identifier that is common between the two datasets may also be available. Because of multiple spellings of names, misspellings, aliases, transposed numbers in dates of birth and SSN, strategies must be developed to 'clean' data and to examine multiple identifiers for a single person in the matching process. Because some data may be more reliable than others, probabilities must be established and a judgment made about whether two records are a genuine match as each record in one dataset is compared to each record in the other (this process is sometimes simplified by using a reliable variable to first 'block' the datasets, e.g., only matching males to males and females to females). Some record pairs will be perfect matches and some will be obvious non-matches. For those for which there is some doubt, a 'threshold' must be established by examining a sample of record pairs. Perfect or partial matches on individual variables are given weights which are added. If the sum is above the selected threshold, then the pair is accepted as a match. For example, dates of birth may have different birth days, but because first, middle and last name and race are exact matches, the pair is accepted (Oklahoma TOPPS II Final Report, 2003?). An advantage to this method is that no one variable must be perfectly reliable in both datasets for matches to be made. Disadvantages include the time required to examine records and establish a threshold, and the limits imposed by unreliable reporting of data.

A third approach to matching is the probabilistic population estimation (PPE) method developed by Banks and Pandiani (1996). This method requires only date of birth and gender in the two datasets. Rather than match individual pairs of records, the PPE method uses the distributions of dates of birth to estimate the degree of overlap between the two datasets. The advantages of this approach are that fewer data elements are required to perform the analysis and issues of privacy can more easily be addressed. A disadvantage is that fewer analyses can be performed on the matched data.

Advantages and Limitations. Some of the advantages and disadvantages of the individual approaches to data matching have been described above. One advantage of these approaches in general is their sustainability. Once the data sharing relationships have been established, algorithms have been developed and data matching programs written, the continuation of the process is a matter of scheduling data extract and matching programs to be run. This is a far less expensive and burdensome approach to outcomes assessment than mounting a system-wide survey or follow-up study. Data matching avoids other limitations of surveys as well, e.g., high attrition rates, self-report bias, sampling design problems, and relatively short observation periods (Fowler, 1994). Using data matching, entire populations of clients can be studied using objective measures over several years.

The utility of analyses based on data matching is limited by the data available to be matched, e.g., access to school performance data is often restricted, but other sources that represent areas of concern to consumers and other stakeholders are often available, such as, criminal justice data and income data. A significant limitation of some administrative data is the timeliness of its availability. Although some datasets, such as employment data, may be available on a quarterly basis, it is often the case that administrative data are not compiled and made available until after the end of a fiscal or calendar year. Sometimes the compilation and reporting process may take six months beyond the end of the reporting period, so data may be a year and a half old before it is available for matching and further analysis. Another potential limitation to using data matching is the level of access to other agency databases of interest (see the discussion below).

Data Sources. Any administrative dataset is a potential source for matching if there are enough identifiers present that correspond to, or can be re-formatted to correspond to, identifying variables in the mental health dataset. Examples include substance abuse treatment data, DUI and other arrest data, prison and jail data, employment data, mortality (death certificate) data, hospital discharge data, TANF data, child custody data, and Medicaid service data.

Data Access. At least two important factors affect access to other agencies' data: statutory limits on access (e.g., HIPAA for health-related data, 42 CFR for substance abuse data, or FERPA for school data) and the capacity and willingness of other agencies to cooperate. The former can be influenced by how conservatively agencies interpret regulations and how imaginative would-be users are when designing protocols that meet privacy demands. The latter can be affected by state budgets, agency relations, Governor and Legislature leadership, and individual staff relations. It is important to consider how each of these factors may be at play in your state and address each one. Querying contacts made during previous work on interagency projects can be a useful way to begin to understand another agency's culture, policies about data sharing and lines of authority before drafting a request for data access. Sometimes informal requests can be made to information services staff; in other situations, formal requests from one agency director to another may be more fruitful. Initiation of a new program, or the occurrence of a crisis event that incites demands for more oversight, can provide the opportunity establish a data sharing agreement that was not present before. For example, a state starting a mental health court might pursue access to criminal justice data to monitor the impact of the new program, or a demand to eliminate duplicate spending by the Medicaid agency and SMHA could provide the opportunity to establish a data sharing agreement between the two agencies. Given the decline in state budgets in recent years, it is advisable to start with a narrowly focused request, rather than overwhelming an agency with a huge request. Ask what the critical questions are you want to address with the data. You may also want to consider using grant funds or some other source to support the contributing agency's data system changes, e.g., to extract data in the format you need.

You must be meticulous about addressing privacy, confidentiality and security issues in data sharing agreements. Thoroughly describe all protocols for data storage, management, re-release and destruction. Make use of your agency IRB, general counsel or HIPAA privacy officer to ensure all likely concerns are addressed. Attending to these issues in any data sharing agreement you draft will help demonstrate your commitment to keeping the contributing agency's data safe from misuse.

Uses of Matched Data. Using matched administrative datasets can be very productive. There is the potential to evaluate an entire population, rather than a sample. Since self-report data collected through surveys may have selection biases not found in administrative data, when parallel questions are addressed by both data sources, the latter can be used to validate and interpret the former.

Evaluating long-term outcomes of programs is a particular strength of matched data. Some treatment effects may not persist or may not manifest themselves immediately. The results of analyses using follow-up surveys to evaluate treatment may lose too many people to attrition or may be too expensive to maintain to assess long-term outcomes. But with longitudinal administrative datasets, it is possible to perform pre-to-post-treatment analyses (and identify large treatment and comparison groups) with extended pre- and post-treatment observation periods without additional cost, except perhaps for relatively inexpensive additional computer storage capacity or speed.

Another use of data matching is to simply demonstrate the overlap in two populations. Such an analysis can be used when there is a question of whether two agencies are duplicating services among people both serve. Another example would be to demonstrate the need for two agencies to better coordinate services for people they see in common, e.g., showing that many people going to jail have been served in the mental health system.

Resources. Data matching can be a fruitful and relatively inexpensive alternative for performing program evaluations. The preceding discussion provides a brief overview of issues to consider, but more examples and detailed information about performing matches can be found in the following sources:

References

- Baldwin, J.A., Acheson, E.D., and Graham, W.J. (1987) *Textbook of Medical Record Linkage*. Oxford: Oxford University Press.
- Banks, S.M. and Pandiani, J.A. (2001) "Chapter 12. Using Existing Databases to Measure Treatment Outcomes," *Developing Outcome Strategies in Children's Mental Health*, Edited by Mario Hernandez, Ph.D., and Sharon Hodges, Ph.D., Paul H. Brookes Publishing Co., Inc., ISBN 1-55766-520-6.
- Banks, S., Pandiani, J., Schacht, L.M. (1996) "The Probabilistic Population Estimator Applied to Estimating Statewide Mental Health and Corrections Community and Institutional Caseload Size, Overlap, and Outcomes." Presented at the National Conference on Mental Health Statistics, Washington, D.C.
- Bell, R.M., Keesey, J., and Richards, T. (1994) "The Urge to Merge: Linking Vital Statistics Records and Medicaid Claims," *Medical Care*, Vol. 32, No. 10: 1004-1018.
- Brown, M., Longhi, D., & Luchansky, B. (1997) "Employment outcomes of chemical dependency treatment and additional vocational services publicly funded by Washington State." Olympia: Washington State Department of Social and Health Services/Research and Data Analysis.
- Boussey, C.A. (1992), "Record-Linkage Methodology the Key to Epidemiological Surveillance," Department of Health and Rehabilitative Services, Interagency Office of Disability Prevention, Tallahassee, FL.
- Coffey, R.M., Graver, L., Schroeder, D., Busch, J.D., Dilonardo, J., Chalk, M., and Buck, J.A. (2001) "Mental Health and Substance Abuse Treatment: Results from a Study Integrating Data from State Mental

- Health, Substance Abuse, and Medicaid Agencies," U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Substance Abuse Treatment, Center for Mental Health Services.
- Copas, J.B. and Hilton, F.J. (1990) "Record Linkage: Statistical Models for Matching Computer Records," *Journal of the Royal Statistical Society*, 153: 287-320.
- Felling, I.P. and Sunter, A.B. (1969), "A Theory for Record Linkage," *Journal of the American Statistical Association*, 64(328), 1183-1210.
- Finigan, M. (1996), "Societal Outcomes and Cost Savings of Drug and Alcohol Treatment in the State of Oregon," Office of Alcohol and Drug Abuse Programs, Salem, OR.
- Fowler, F.J., Jr. (1993). "Survey Research Methods (2nd ed.)." Newbury Park, CA: Sage Publications, Inc.
- Gerstein, D.R., Johnson, R.A., Harwood, H.J., Fountain, D., Suter, N., and Malloy, K. (1994), "Evaluating Recovery Services: The California Drug and Alcohol Treatment Assessment" (CALDATA), California Department of Alcohol and Drug Programs, Sacramento, CA.
- Goerge, R.M. (1992), "Tracking Children's Careers Across Multiple Service Settings: Creating an Interagency Database on Children's Services," National Governors' Association Conference on Making Information Work, Washington, D.C., January 18-23, 15-29.
- Harrison, P.A. (1995), "Developing State Outcomes Monitoring Systems for Alcohol and Other Drug Abuse Treatment" (TIP Series 14), U.S. Department of Health and Human Services, Public Health Service, Substance Abuse and Mental Health Services Administration, Center for Substance Abuse Treatment, Rockville, MD.
- Hassard, T.H. (1986), "Writing the Book of Life: Medical Record Linkage", *The Fascination of Statistics* ed: Brook, Arnold, Hassard, and Pringle, 25-46.
- Institute of Medicine (1990) "Broadening the Base of Treatment for Alcohol Problems." Washington, D.C.
- Jaro, M.A. (1995), "Probabilistic Linkage of Large Public Health Data Files," *Statistics in Medicine*, 14: 491-498.
- Kerlinger, F.N. (1986) "Foundations of Behavioral Research (3rd ed.)." New York, NY: Holt, Rinehart, and Winston.
- Kestenbaum, B. "Probability Linkage Using Social Security Administration Files", Working Paper from the Social Security Administration.
- Longhi, D., Brown, M., and Comtois, R. (1994) "ADATSA treatment outcomes: Employment and cost avoidance." Olympia, WA: Washington State Department of Social and Health Services Planning, Research and Development, Office of Research and Data Analysis.
- Newcombe, H.B., Kennedy, J.M., Axford, S.J., and James, A.P., (1959), "Automatic Linkage of Vital Records", *Science*, 130(3381): 954-959.
- Newcombe, H.B. (1967), "Record Linking: the Design of Efficient Systems for Linking Records into Individual and Family Histories," *American Journal of Human Genetics*, 19: 335-59.

- Newcombe, H.B. (1988), "Handbook of Record Linkage: Methods for Health and Statistical Studies, Administration and Business." New York: Oxford University Press.
- Pandiani, J.A., Banks, S.M., Schacht, L.M., (1998) "Using Incarceration Rates to Measure Mental Health Program Performance." *The Journal of Behavioral Health Services and Research*, 25(3): 301-312.
- Pandiani, J.A., Banks, S.M., Schacht, L.M., (1998) "Personal Privacy vs. Public Accountability: A Technological Solution to an Ethical Dilemma," *The Journal of Behavioral Health Services and Research*, 25(4).
- Pandiani, J.A., Banks, S.M., Bramley, J. and Moore, R. (2002) "Mortality of Mental Health Services Recipients in Vermont and Oklahoma," *Psychiatric Services* 53: 1025-1027.
- Perrin, E.B. & Koshel, J.J. (eds.) (1997) "Assessment of Performance Measure for Public Health, Substance Abuse, and Mental Health." Washington, DC: National Research Council, National Academy Press.
- Roos, L.L. and Wajda, A. (1991) "Record Linkage Strategies, Part I: Estimating Information and Evaluating Approaches," *Methods of Information in Medicine*, 30: 117-123.
- Stern, D. L. (1994) "System Change and the Maturation of Mental Health Outcome Measurement. Outcome Issues in a Managed Care Environment." Boulder, CO: Western Interstate Commission for Higher Education.
- Substance Abuse and Mental Health Services Administration, Center for Substance Abuse Treatment (1999) Detailed Description of Linking Methods. <http://www.samhsa.gov/csatsat/idbse/idbmethod.asp>.
- TOPPS-II Interstate Cooperative Study Group (2003) "Drug Treatment Completion and Post-discharge Employment in the TOPPS-II Interstate Cooperative Study," *Journal of Substance Abuse Treatment* 25: 9-18.
- Whalen, D., Pepitone, A., Graver, L., and Busch, J.D. (2001) "Linking Client Records from Substance Abuse, Mental Health and Medicaid State Agencies," U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Substance Abuse Treatment, Center for Mental Health Services.
- Wickizer, T.M., Campbell, K., Krupski, A., Stark, K.D. "Employment Outcomes Among AFDC Recipients Treated for Substance Abuse In Washington State." *Milbank Q*; 78(4):585-608.