the EVALUATION@HSRI CENTER

The Performance Improvement Project:

A Technical Assistance Manual

October 13, 2004

Prepared by:

Dow A. Wieman, Ph.D.
The Evaluation Center @HSRI
Human Services Research Institute
2269 Massachusetts Ave

Robert Egnew, M.S.W., M.P.H
National Association of County Behavorial Health Directors

Ed Diksa, ScD
California Institute for Mental Health

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

CiMH
Helping You Make the Difference

TABLE OF CONTENTS

Preface

In 2003, the Centers for Medicare and Medicaid Services (CMS) promulgated new regulations requiring Medicaid managed behavioral health organizations to conduct annual Performance Improvement Projects (PIPs). This requirement applies to the county-sponsored authorities that provide mental health services under California's Medicaid wavier programs.

In order to comply with this new requirement, the California Institute for Mental Health (CIMH) and the National Association of County Behavioral Health Directors (NACBHD) requested the assistance of the Evaluation Center@HSRI to develop a technical assistance manual that would support the development and implementation of PIPs. The Evaluation Center @ HSRI is a national technical assistance center for the evaluation of adult mental health systems funded by the federal Substance Abuse and Mental Health Services Administration (SAMHSA) Center for Mental Health Services (CMHS) and operated by the Human Services Research Institute (HSRI).

The manual provides a step by step approach to all phases of developing and implementing and evaluating Performance Improvement Projects, with detailed information and resources presented in format and language designed to be readily utilized by local county-sponsored behavioral health authorities.

Acknowledgements

The following individuals contributed to the development of the Performance Improvement Project Technical Assistance Manual:

Sheila Baler	Darcy Johnson	Rita McCabe-Hax
Ed Diksa	Gail Kinnamon	Kenneth Meibert
Eydie Dominguez	Diane G. Koditek	Timothy Mullins
Robert Egnew	John Lessley	Sandra Naylor Goodwin
Fred Hawley	Beth Martinez	Madelyn Schlaepfer

Introduction

Basic Principles

Experienced managers, clinicians, consumers and others "in the field" usually know instinctively when behavioral health care is of the highest quality and, alternatively, when there are gaps between what is possible and what is currently provided. This Manual describes methods for converting this intuitive knowledge into effective action by means of Performance Improvement Projects.

This section describes the basic concepts underlying the Manual's approach to providing technical assistance with conducting Performance Improvement Projects (PIP).

1. What is the purpose of this technical assistance manual?

The manual is intended to support the efforts of Mental Health Plans (MHPs) to design and carry out PIPs that meet the Federal standards, as represented in the protocols for evaluation by External Quality Review Organizations (EQROs), and that are effective in achieving the goals of providing high quality care. Accordingly the manual addresses the issues identified in the EQRO evaluation protocols (that are in turn described in the authorizing legislation), and in addition offers recommendations based on best practices in performance measurement and quality improvement (QI) drawn from a wide range of sources and experiences.

(NOTE: Specific references to the protocol for EQRO evaluation and elaborations upon points in that document are identified parenthetically as "Protocol." A second important source is a presentation by Judy Ashley, entitled "Surviving and Thriving in the World of PIPs"; these references are identified as "Ashley." Other sources are similarly identified parenthetically.)

2. How is this technical assistance manual organized?

The manual is laid out according to a set of questions and answers about the activities necessary for planning, collecting data, implementing, reporting and evaluating an effective PIP. Each of these phases contains a series of steps, also represented in question and answer format.

The manual addresses the following questions that PIP teams are likely to ask various points in the PIP process.

How do I:

- Select the study topic?
- Define the study question?
- Select the study indicators
- Identify a representative and generalizable study population?
- Use sound sampling techniques (if sampling is used)?
- Collect reliable data?
- Implement interventions and improvement strategies?
- Analyze data and interpret study results?
- Plan for improvement?
- Sustain the improvement?
- Evaluate the improvement process?

3. What is a Performance Improvement Project (PIP)?

In general terms, a PIP is "a systematic process for improving the quality of care and service designed, conducted and reported in a methodologically sound manner" (Ashley).

To elaborate, the Protocol defines a PIP as "a set of related activities designed to achieve measurable improvement in processes and outcomes of care. Improvements are achieved through interventions that target health care providers, practitioners, plans, and/or beneficiaries."

In this context, "outcomes" are measures of client mental health, functional status or satisfaction following the receipt of care or services (Protocol).

(NOTE ON TERMINOLOGY: "Process" and "outcomes" along with "structure" are conventional terms in the field of health care quality measurement, used to describe dimensions of health care. Some have suggested that the recent and very influential report on health care quality from the Institute of Medicine (IOM), entitled *Crossing the Quality Chasm* indicates a need for revising this classification of measures (Institute of Medicine 2001). Appendix B presents the rationale for retaining the conventional terminology in this Manual.)

4. What is the difference between a PIP and a research project?

PIPs typically employ many of the features of clinical, services and epidemiological research, including design, methods and terminology. They differ most fundamentally, however, in the purpose for which they use these tools. Whereas the goal of the research project is to create *new knowledge*, the PIP uses existing knowledge (including that produced by previous research) to *enhance the effectiveness and efficiency of operations*. A PIP would not seek to test the effectiveness of a process of care or to develop practice guidelines; rather, it would promote implementation of processes already shown to be effective or adherence to established guidelines.

In practice, the distinction may be less clear. For example, interventions employed by PIPs will have varying levels of evidence to support their effectiveness, and one aspect of measurement may be to ensure that that the intervention is achieving the expected effect. Also, some improvement projects may have sufficient value in demonstrating new approaches to the provision of care and be sufficiently rigorous in their design and execution to warrant description in peer-reviewed research publications or other means of dissemination. (Appendix E presents several examples).

The following are some elaborations on the differences between PIPs and research projects (Ashley):

- Objectives of PIPs support real work or practical application of findings whereas research has no requirement for practical application.
- With PIPs, all participate in the improvement process vs. research use of experimental and control groups.
- Tests are observable in PIPs, vs. blinded in research.
- The topic of a PIP is adapted to changes, vs. the fixed hypothesis in research.
- A PIP entails many sequential tests, vs. one fixed test of a hypothesis in research.
- A PIP is usually not subject to the more rigorous requirements for review by Institutional Review Boards (IRBs) for protection of subjects and the privacy requirements of the Health Insurance Portability and Accountability Act (HIPAA).

5. What is quality improvement?

A variety of models and methodologies for improving the quality of health care have been promulgated over the years, identified as Quality Assurance, Total Quality Management, Continuous Quality Improvement, and Performance Improvement. While these differ in their particulars and to some degree in philosophy, they all share three general features:

They locate the source of most quality problems in the health care *system*, not individuals within it. In the words of the IOM, "exhortation, blaming and trying harder" will not fix most quality problems.

Their respective approaches are *data-driven*. Problems are identified and improvements are monitored through the systematic collection of information about organizational operations, providers and patients.

Their approaches are *cyclical*. Their models are based on the assumption that improvement activities overall will take place continuously, and that there is always room for improvement in every aspect of health care. Specific improvement projects are cyclical in that they involve some variation on collecting and analyzing information to identify improvement opportunities, developing an improvement strategy, and then measuring again. If improvement has not been satisfactory, a new cycle begins.

A Note on Templates:

Many of the recommendations presented in this Manual lend themselves to representation in the form of templates, checklists, worksheets, etc. Because each group is likely to differ in its preferences and requirements for how this information is organized, the Manual does not prescribe any specific model (though several sources for examples are presented in Appendix E). For various sections of the Manual, however, you may wish to develop your own templates for documenting and communicating decisions made and actions taken.

Phase I

How do I plan for a PIP?

Steps in this phase

- How do I select a PIP study topic?
- How do I prioritize study topics?
- How do I define the study question?
- How do I choose study indicators?
- How do I establish indicator criteria?
- How do I define the study population?

The effectiveness of your PIP will be directly proportionate to the care and effort that you invest in the planning process. PIPs represent a significant commitment of resources, and thorough planning will ensure that these are invested wisely. This phase covers all aspects of the planning process necessary to conduct an effective PIP.

Basic Principles:

The critical elements of a PIP

Data-based. A PIP is based on systematic and comprehensive data collection and analysis

Data elements. A PIP utilizes specific types of information about service system, providers and recipients: penetration (proportion of total enrollees using a service), utilization (number and type of encounters), socio-demographic characteristics: (age, gender, race based on standard definitions, diagnoses, disability or functional status, satisfaction, and treatment outcomes.

The Quality *Improvement process.* A PIP relies on the standard cycle of steps for quality improvement: Plan Do Check Act.

Real improvement. The gain in quality is measurable and would not have occurred without the PIP.

Lasting improvement: The improvement in quality is sustained by the organization as demonstrated by repeated data collection (e.g., one-year follow-up).

Step1 *How do I select a PIP study topic?*

A study topic is simply an area of concern in behavioral health care. Study topics may involve some aspect of processes of care, treatment outcomes, access to services, continuity of care, consumer satisfaction or many other features of behavioral health care systems where there is reason to believe that opportunities for improvement (gaps between actual and best practices) may be present.

What are Administrative and Clinical study topics?

Study topics may be classified according to two general types: administrative or clinical.

A clinical study topic would be one for which outcome indicators measure changes in mental health status or functional status. An administrative study topic would be one for which indicators measure changes in beneficiary satisfaction or processes of care.

Where should I begin looking for study topics?

The first, and perhaps most productive source of opportunities for improvement is your own personal experience and observation through site visits and discussions with clinicians, managers, staff and consumers. As noted in the Basic Principles cited in the Introduction, most people who participate directly in the system of care have an instinctive sense of where and how quality might be improved. This information will prepare you for the more systematic methods of data collection and analysis described below.

The following areas that MHPs are required by contract to monitor for quality improvement provides a framework for organizing your ideas, impressions and observations about opportunities for quality improvement:

- Service delivery capacity
- Accessibility of services
- Consumer satisfaction
- Service delivery system and meaningful clinical areas, including medication practices
- Continuity of care with physical health care providers and other human service agencies
- Provider appeals

What are some additional sources for study topics?

The hallmark of quality improvement is continuous data collection and analysis. You should routinely collect and analyze, on an ongoing basis, administrative information (e.g., claims and encounter data) in areas such as the following:

- Socio-demographic characteristics of the entire population of enrollees
- Patterns of service use. especially by vulnerable sub-populations (Examples: elderly, racial and ethnic minorities)
- Services or processes of care that are provided in high volume and/or offer the greatest potential for harm (Examples: medications, seclusion and restraint)
- Areas which are known to be prone to problems based on organizational experience
- Sentinel events, i.e., critical incidents involving death or serious injury, or a high risk of these, requiring immediate response

Study topics may be suggested by information received through other channels as well, such as:

- Results of existing performance measurement initiatives
- Consumer/Family member input: Consumer satisfaction surveys and grievance procedures are a means of identifying quality problems not apparent from administrative information.

- Provider input: Your organization should have an established process to solicit recommendations from individual about potential opportunities for improvement
- Findings of Compliance Officers
- Utilization management data (e.g., service authorization, notices of action , denials)

Finally, you may find suggestions for potential quality improvement opportunities by reviewing the trade and scholarly publications and conference reports in the behavioral health field

By combining information from these various sources, you will almost certainly identify more than one area at any given time in which there are likely to be opportunities for improvement.

Step 2 *How do I prioritize study topics?*

PIPs require a considerable investment of resources. It is important, therefore, that the topic you select will provide maximum return on your investment. There are two general aspects to consider in choosing among possible study topics: the *quality* of the topic, as determined by established criteria; and the *feasibility* (practicality) of the topic, as assessed on several dimensions. Each of these is discussed in turn below.

What are the characteristics of a good study topic?

A number of quality improvement protocols and guidelines have identified criteria for what makes a good study topic, based on importance and expected level of benefit. A good study topic combines considerations of the population affected, the prevalence or frequency of the clinical condition or system characteristic that is target and the importance of the topic for MHP members. These considerations include the following:

Relevance to the population. The topic should be based upon information about the characteristics and utilization patterns of the MHP members to insure that it is one that significantly affects the population. The topic of improved access for minority populations, for example, would be irrelevant for a membership that is relatively homogenous.

Prevalence of the condition, volume of services or extent of need. To fully understand the nature of a quality gap and to monitor progress in addressing it requires a sufficient number of instances to insure that they are typical and representative. An unusual or typical occurrence may be a cause of concern but it would not be a productive topic for a PIP, and if the event occurs infrequently it is impossible to know how like it is to reoccur. A possible example is suicide, which might be investigated more productively using methods of qualitative case study or root cause analysis (see Step 3).

Extent of risk. This consideration should be balanced with that of frequency: that is, study topics may have considerable potential for benefit if they focus on either high-risk processes that occur relatively less frequently or process that have lower risk but occur in higher volume.

Extent to which the intervention is likely to make a meaningful difference. Some services or administrative processes may occur frequently and/or affect a large proportion of enrollees without being prone to quality problems. For such as these an improvement process would produce little gain.

Extent to which the MHP can expect to improve care through a PIP, that is, the extent to which a particular outcome is under the control of the system or individual provider to control. This is often a more complicated determination than might first appear. Consider for example the adequacy of physical health care for enrollees in an MHP. The enrollees' primary clinicians may validly assert that health care lies outside their professional responsibility, area of expertise, command of resources, or all three. Alternatively, however, the MHP may establish a process for improving the integration of behavioral and physical health care, by insuring that each enrollee has an identified primary care physician and that mechanisms are in place to insure regular communication between clinician and primary care provider (or to record the enrollee's non-consent). In this example, an administrative, rather than clinical, intervention was the most appropriate, and this opportunity for improvement would have been bypassed had the organization assumed it had no influence over the process.

Demonstrating that an outcome is under the control of the individual clinician or the MHP is the most critical factor for establishing credibility of the PIP process and thereby for achieving stakeholder (especially clinician) buy-in. For this reason, it is very important to include individual providers or managers (depending on the topic) in the group responsible for selecting study topics and indicators.

What are the practical considerations in selecting study topics?

Once you have narrowed your list of potential study topics by means of the criteria above, of course no single topic will rank the highest on all considerations. You will therefore need to further prioritize topics according to practical considerations as well. Some of these are:

Ease of implementation. The choice of study topics should certainly take advantage of "low-hanging fruit" by selecting areas in which the intervention, while acceptable according to the criteria described above, also is likely to require relatively few resources, limited behavioral change, etc. Ease of implementation is likely to be determined by the additional following considerations.

Burden on existing operations. The PIP team may wish to defer a topic for which the intervention is likely to impose demands on individuals or the system as a whole, in favor of others that are easily integrated into existing practices. For example, a topic that requires changes in the MHP's information system or extensive additional data recording by clinicians to monitor performance is less likely to be successful than one that can be accomplished using existing sources of information.

Acceptability to all stakeholders (individual providers, members, administrators, support staff, etc.) This is an extension of characteristic of a good study, described above. Even if a topic will impose little additional burden, stakeholders are likely to be resistant if, for example, they see no utility in the change, or they fear it will draw resources from other activities important to them. Again, this is an argument for soliciting the input of all major stakeholders at some point in the topic selection process.

Relevance to specific program goals or priorities. An MHP may establish some goal regardless of evidence of any quality problem and wish to measure the extent to which the goal has been attained.

Amount of retraining required. An intervention involving changing behavior or practice that can be accomplished only by retraining will be more costly and more likely to meet with resistance than one in which change can be accomplished by less direct methods of influencing processes of care.

Availability of existing standards or guidelines. Practice guidelines are recommendations based on evidence or expert opinion to inform treatment choices. Though not necessary for every PIP project, and not available for many important topics, using tested guidelines or standards simplifies the PIP process and enhances the reliability and validity of indicators. The availability of national standards will provide benchmarks against which performance may be measured.

Availability of data. Perhaps one of the most important, because most often problematic, consideration is whether the data are available to measure current performance and progress toward quality goals. You may not be able to ascertain this immediately, and for any topic under consideration you should spend some time exploring the nature and quality of required data before proceeding further.

For the process of selecting study topics, you may want to develop a checklist to insure that all of the above considerations have been adequately addressed and to document the decisions of the PIP team.

What are the ethical considerations in conducting a PIP?

Quality improvement activities are generally not required to maintain the same degree of oversight as research projects on the grounds that they involve operations rather than direct patient treatment, impose little or no risk, and that in any case, requirements for research such as informed consent would be impractical for QI. This is also the view of the federal government as represented by legislation such as HIPAA.

Others, such as Perneger (2004), however, argue that research and quality improvement are less different than many assume. QI projects may in fact influence the type of care provided, and may impose at least some degree of risk to the extent that quality interventions may have unintended deleterious consequences which might be overlooked by those who are personally invested in

them. Moreover, though individual informed consent may be impractical, "community consent" may be obtained, for example by having consumer representatives participating in PIP.

MHPs are not likely to subject PIPs to the same standards of outside review as for research projects. Nonetheless, it is worthwhile to consider issues of patient protection explicitly throughout the PIP process.

Step 3 *How do I define the study question?*

Having identified the topic of the quality improvement study, you are now ready to formulate the specific problem to be addressed. The PIP protocols refer to this as the "study question."

This step may appear simple, but it is the most critical to the success of your PIP project, and the most common stumbling block in planning quality improvement projects. It is important, therefore, that the entire project group give it careful attention, preferably using a systematic approach such as Root Cause Analysis to define the problem to be addressed.

Root Cause Analysis

Root cause analysis (RCA) is a structured technique developed in industry for intensive analysis of the causes of isolated but serious accidents. The technique has been adopted in health care to respond to "sentinel events," and is now required for accreditation by the Joint Commission on Accreditation of Healthcare Organizations (JCAHO). JCAHO defines a sentinel event as: "An unexpected occurrence, involving death or serious physical or psychological injury, or the risk thereof. Serious injury specifically includes loss of limb of function. The event is called 'sentinel' because it should send a signal or sound a warning that requires immediate attention." (http://www.jcaho.org/pms/reference+materials/glossary.htm).

RCA was developed specifically to explore the causes of quality problems that are serious and call for immediate action but occur too infrequently to be identified through routine quality improvement data collection and analysis. However, this method may also serve to refine and focus PIP study topics generally.

RCA addresses a major challenge for PIP teams: "digging down" to identify "proximal" (i.e., immediate) causes of a quality gap as opposed to the "distal" (less direct) causes. PIP teams frequently fail initially to adequately map the causal sequences to identify the primary barriers to optimum care. The failure to perform this critical task results in a quality improvement strategy that is ineffective because it does not address the true source of the problem.

Following a data collection phase in which the team assembles all available information and data about the failure, the essence of RCA consists of "causal charting" i.e., identifying the entire chain of events and range of circumstances leading up to the sentinel event. This is accomplished by repeatedly asking "why" questions of each distal cause until finally arriving at the most proximal cause, to which the improvement or prevention strategy will be addressed. These are typically organized into factors, e.g., human, equipment, environmental, risk, barriers, etc.

This is only an overview of the basic concepts of RCA. A number of commercial, organizational and governmental publications are available, easily identified through a web search. Most of these offer various types of tools for conducting RCAs. An example is the JCAHO "Framework for a Root Cause Analysis and Action Plan In Response to a Sentinel Event" available in various software formats (Word, Excel, etc.) at:
http://jcaho.com/accredited+organizations/ambulatory+care/sentinel+events/forms+and+tools/framework+.htm

Stating the study question

The three most important criteria for the study question are that it be stated in terms that are: 1) *clear*; 2) *simple*; and 3) *answerable*. To ensure that the question meets these standards, it is useful to write it out for review by the PIP project team, and then submit it to be assessed by a variety of stakeholders, notably those whose activities are addressed by the question such as clinicians or managers.

The answerability criterion poses the greatest challenge for PIPs, most commonly due to the inadequacy or unavailability of the data necessary to answer the question. It is therefore worthwhile to invest substantial effort in two activities: first, specifying the nature and quality of the data required to answer the question; and second, assessing the availability of the necessary data within the organization.

Some examples of study questions:

Suppose the chosen topic is access barriers for sub-groups of enrollees. Potential study questions might be, "Does the location of clinics impact utilization of mental health services by Latino beneficiaries?" or, "Do clinic opening and closing hours have an impact on elderly members?"

Another topic may be hospital readmissions. A possible study question is, "Does the failure to make post hospital medication appointments increase consumers' chances of being readmitted to a hospital in a 12 month period?"

Basic Principles: Continue asking why

Whether the PIP team employs RCA or some analogous process, its members should recognize from the outset that *formulating a good study question is a highly iterative process*. The more intensively the question is reviewed from more perspectives in the planning phase, the more likely that it will address the true root causes of problem, and therefore the more effective the PIP will be. This process of revising and refining the study question should not end with the planning phase, however, but continue throughout the entire PIP process in response to better understanding, additional input, availability of data and other developments.

Step 4 *How do I choose study indicators?*

Study indicators are: defined, measurable variables which reflect a discrete event or a status, used to measure performance in the area of the study topic.

An example of an event would be screening for substance abuse. An example of a status would be level of functioning. Indicators can be few and simple, or many and complex, depending on the study questions and the availability of data and resources to gather the data (Protocol).

(NOTE ON TERMINOLOGY: The PIP review protocol states, "It has been found, in program evaluation literature, that 'outcome measures', 'measures', and 'indicators' appear to be used interchangeably." The Protocol has selected the term "Study Indicator," and this manual generally follows this precedent, except in some instances where the context calls for some alternative such as "performance measure" as more appropriate.)

To continue the example from the previous step, the PIP might be a study focused on assessing and improving the accessibility of services to a specific population, including reducing disparities between the level or types of services provided to this population as compared to other MHP enrollees. Indicators would be required to identify areas of needed improvement (in this case differences among groups in the services received) and to measure improvement over time. In this case, penetration rates (proportion of a group receiving a service) for the population being studied, and for the population as a whole as a comparison, would be measurable indicators in a study of access (Protocol).

Indicators, as in this case, are usually reported as rates; i.e., the ratio between the number of people who, for example, received a certain service (the numerator) to the number of people eligible for the service and for whom it is clinically indicated within a specified period (the denominator). An example might be the ratio between the number of all enrollees who identify them selves as "satisfied" or "very satisfied" with services received during the past year (numerator) and the total number who received services during that period (denominator).

Where can I find good indicators?

The preferred source for indicators is practice guidelines based on empirical evidence and tested in the field. The Schizophrenia PORT guidelines, for example, could provide indicators for interventions addressing patterns of antipsychotic medication providing or provision of psychosocial services and other evidence-based interventions (Buchanan 2002).

The Center for Quality Assurance and Improvement in Mental Health (CQAIMH) at www.cqaimh.org maintains a database of hundreds of behavioral health process measures of care. Measures are drawn from a wide range of sources including state mental health agencies, provider organizations, accrediting agencies, and research projects. Search categories include diagnosis, quality domain (e.g., prevention, access, continuity), clinical setting, level of evidence and others.

Alternatively, an MHP may choose to develop its own performance indicators. This is a fairly technical and resource-intensive undertaking, however, that requires careful planning, analysis to insure reliability and validity, and field-testing to insure feasibility. Some of the requirements for indicators are described below.

What are the characteristics of a good study indicator?

All indicators must meet at least the following criteria (Protocol). They must be:

Objective: The way in which data is collected, analyzed and reported will not vary, no matter who performs these activities. Objectivity depends highly on the extent to which the remaining criteria are met.

Clearly and unambiguously defined: How the measure is to be applied, and to whom, must be fully specified.

Based on current information: The data required for the indicator must reflect current practices and characteristics of the system.

Capable of measuring the area of concern; e.g., consumer outcomes, access to care, and satisfaction, or their proxies.

Reliable: Allows for tracking trends over time.

Valid: Measures what it purports to (see below for further discussion of reliability and validity).

Additionally, you may wish to ask the following:

- Does the indicator measure a standard or goal that should be achieved and maintained?
- Does the indicator represent an outcome where need for improvement is known (e.g., high rates of sentinel events such as avoidable hospitalizations)?
- Can the indicator measure achievement of a specific program goal or program priority (e.g., increase in system capacity)?
- Is there empirical evidence of the validity of the measure's value as a quality indicator?
- Is there evidence that the indicator provides meaningful information about the chosen study topic?
- Are there guidelines, benchmarks, performance measures or objectives set by professional organizations to use as standards?
- Is the indicator used with national and state data sets to serve as benchmarks for comparison?
- Is the indicator sensitive to change and can it measure effectiveness of health service programs?
- Is the indicator similar to those used in other geographic areas so state and national comparisons can be made?

These questions may be used as a check-off list in the written PIP plan.

Basic Principles:

Two Psychometric Concepts You Should Understand

Reliability. A "reliable" measure will give the same result every time that it is applied to the same aspect of health care. This can be compared to a thermometer, which will show the same temperature each time it is used to measure the temperature in a particular location in a living room if, in fact, the temperature has not changed.

Validity. A "valid" measure will be reliable and will also measure the intended aspect. This can be compared to ensuring that the reading on the thermometer measuring the average temperature in the living room is not instead measuring the temperature of the corner next to the hot fireplace.

(From AHCPR "Child Health Care Quality Toolbox" www.ahcpr.gov/toolbox.)

Reliability and validity are assessed using specific statistical tests, which require appropriate expertise. If you are developing your own measures and do not have the requisite expertise in-house, you should obtain consultation to ensure that these critical steps in measure development are successfully completed. For guidance, you may find discussion of the types of validity and reliability and methods of testing them in most textbooks on measurement and many online sites, several of which are listed in the Resources section.

Final Selection of Indicators.

Having developed a list of viable indicators, your PIP team is ready to make the final selection. This process may be unexpectedly difficult and time-consuming due to the inherent tension between necessarily including multiple stakeholder groups and finally reaching consensus among them. It may be that underlying differences in priorities will not emerge until the final decision process.

Step 5 *How do I establish indicator criteria?*

Indicator criteria (sometimes known as "measure specifications") are the set of rules describing how the indicator is to be applied.

For an indicator of penetration rates as a measure of access to a particular service, for example, criteria would include specific definitions of the population studied (e.g., diagnosis, demographic characteristics, length of enrollment, etc.), precise definitions of the service (e.g., outpatient medication management, but not counseling) and criteria for having received the service (e.g., one encounter in a three month period), and definitions and data sources for numerators and denominators to determine rates.

The written PIP plan should include a description of the criteria, including at least:

- Definition of the study population

- Description of data collection methods
- Definition of numerator and denominator
- Definition of units of measure (e.g., encounters, penetration)

Step 6 *How do I define the study population?*

The study population for a PIP should represent the entire enrolled population or some specified sub-group (e.g., a category defined by race, ethnicity, gender, etc.). For MHPs, therefore, it would include all Medi-Cal enrollees, or a subpopulation differentiated according to some clearly specified characteristic such as those receiving a particular service or those with a particular characteristic. In the latter case, it is important to insure that the available data resources can in fact capture all of those with the identifying characteristic for purposes of the study—an important function of pilot testing, as discussed below.

What are qualitative population data?

Qualitative data are characteristics or attributes by which persons or things can be classified; for example, sex, race, poverty level, or the presence or absence of a specific mental health diagnosis. Calculations of proportions and calculations of rates are the two most common qualitative measures.

What are quantitative data?

Quantitative data are concerned with numerical variables such as age or number of outpatient visits. Quantitative data require, at a minimum, simple descriptive measures of central tendency (i.e., mean, median, or mode) and measures of variability (i.e., range or standard deviation).

Where can I find information about the population?

The following are some potential sources of data for determining target population (PIP Protocol):

- Medi-Cal enrollment files on enrollment characteristics relevant to health risks or utilization of clinical and non-clinical services, such as age, sex, race/ethnicity/language and disability or functional status;
- Utilization, diagnostic and outcome information on Medi-Cal outpatient and inpatient encounters, services, procedures, medications and devices, admitting and encounter diagnoses; and patterns of referrals or authorization requests obtained from MHP encounter, claims, or other administrative data;
- Data from outside organizations such as local or national public health reports on conditions or risks for specified populations;
- Data from other DMH or MHP committees, such as the State and MHP Quality Improvement Committees and Cultural Competency Committees;
- Data from surveys, grievance and appeals processes;
- Data on appointments and provider networks;

Should I use the entire population or a sample?

See Phase III: "To sample or not to sample?" below.

Phase II

What should the data collection plan include?

Steps included in this phase:

1. What type of data do I need in order to conduct a PIP?
2. What are baseline data?
3. What should I consider for my data collection instrument?
4. When should baseline and follow-up data be collected?
5. Who should collect the data?

Step 1 *What type of data do I need in order to conduct a PIP?*

The type of data required for the PIP depends on the answers to three questions: 1) What does the *study question* require? 2) What does the *indicator* require? 3) What is *available*?

The first step in developing the data collection plan, therefore, is to review these three questions carefully, identify as precisely as possible the data that you will require, and determine for certain whether the required data is in fact available in the form and of the quality necessary for the PIP to succeed.

To assess this, you may wish to conduct a small but representative pilot test focused on the feasibility of the data. Even when the design of the indicators is based on detailed information about data sources, hands-on inspection of the actual data often brings unforeseen complications. Characteristic of the iterative process that typifies PIP projects generally, it is often necessary at this point to return to the indicator design phase and make the modifications necessary to have adequate data.

Potential data sources for PIPs include an expanded list of those used to identify study questions:

- Medical records
- Tracking logs
- Encounter and claims systems
- Provider interviews
- Consumer interviews
- Surveys
- Telephone logs
- Complaints or grievances
- Appointment data
- Pharmacy data
- Hybrid data (e.g., medical records and access data)
- Utilization management, authorization, notice of action and denials

For example, if your study question and indicators require identification of groups with potentially greater health risks or problems with utilization of services you may consider enrollment files, encounters and claims data, and clinical records to the extent they provide information on characteristics such as age, sex, race/ethnicity/language, diagnosis and functional status.

If the question and indicators call for information on general utilization you may wish to use encounter databases, medical records, tracking logs and surveys.

For information on appropriateness, you will probably need to review medical records, perhaps combined with automated data systems such as pharmacy or encounter records, and utilization management data.

For information on satisfaction, you would use interviews, surveys and complaint records.

For study topics related to access you may use encounter systems, interviews, surveys, and telephone and appointment records (for waiting times).

Step 2 *What are baseline data?*

Baseline data are simply information about current performance against which improvement is to be measured. Collection of baseline data may be the primary focus in the early phases of a PIP program or of a new PIP. The protocols for collection of baseline data should be at least as rigorous as those for post-intervention data, as the measure of the improvement strategy's success will depend completely on the reliability and validity of baseline information.

Step 3 *What should I consider for my data collection instrument?*

The form of the data collection instrument will of course vary depending on the type of information to be collected and the purpose for which it will be used. Some general principles should be considered however:

Simplicity: The instrument should be designed to collect only the information that is specifically designated in the data collection plan. A common error is to assume that additional data would be worthwhile since it would add little to the burden of collection, even if the purpose is not fully determined at the outset. In reality, however, this is usually a waste of resources, as the information thus collected often turns out to be missing some component necessary for any specific purpose that might be identified subsequently.

Ease of data entry: The instrument should be designed with an eye toward the task of data entry. This requires a prior determination about the structure of the database which will contain the information, with variables and values clearly identified and conveniently located in the instrument.

Clear instructions: The instrument should be accompanied by clear instructions, especially when persons collecting the data may not have been involved in the planning of the study. Instructions

should include information about how to handle situations or responses not explicitly included as components of the study.

Field testing (pilot studies): This is a critical step that is almost certain to result in modifications of the instrument and method of data collection. Field testing should be conducted with a sample of the unit of analysis, whether that is medical records, consumers, or automated data files. To ensure that the test will address the full range of possibilities, the sample should be chosen randomly. Members of the PIP team involved in developing the instrument should participate in the test, and carefully note all unexpected problems and complications, for the team to review and make appropriate modifications.

Data analysis: the data analysis plan should be developed prior to designing the data collection instrument to ensure that all information required for specific statistical procedures will be available. It is important to have the instrument reviewed by the project data analyst or a statistician.

Step 4 *When should baseline and follow-up data be collected?*

The frequency and timing of PIP data collection should be guided by the goal of demonstrating significant improvement sustained over time. This implies at least three data points: 1) baseline (prior to the intervention); 2) post-intervention (measurement of improvement); and 3) follow-up (sustainability of improvement).

The timing will depend on the nature of the study topic, but in general, the collection of baseline data should occur relatively shortly before the period of intervention to insure that it reflects current practice. At the same time it should be collected over a long enough period that it represents more than isolated or atypical events.

Follow-up data likewise should be collected after a period sufficient for the that the improvement strategy to have been incorporated into regular practice, but not so long that the initial gains would have dissipated before a strategy for sustaining improvement, based on the results, can be implemented.

As with most elements of the PIP process, data collection is cyclical and may require revision as the project proceeds. For example, the post-intervention measurement may indicate initial intervention has been ineffective, thus requiring implementation of a revised intervention and re-measuring the effect.

The timing and frequency of data collection are judgment calls to be made by the PIP team but should be informed by the input of stakeholders knowledgeable about the processes under consideration. The written PIP plan should include clear specifications of these time frames.

Step 5 *Who should collect the data?*

The nature of the data to be collected will determine the required qualifications of data collection staff. Medical record abstraction to identify and assess complex procedures will require highly trained clinicians, whereas clerical staff may conduct a review of records to identify the presence of a specific entry in a structured chart. A chart review for less specific indicators that require informed judgment such as cultural competence, for example, will likely require a person with special expertise in that area.

For review of medical records, the PIP team should develop data abstraction tools that limit as much as possible the extent of individual judgment required and thereby enhance uniformity (inter-rater reliability) among reviewers.

The PIP data collection plan should include a glossary of terms for the project to reinforce consistent interpretation among the project staff.

Collection of information from an automated data system may require a person with minimal clinical background, but solid skills in information technology.

In turn, for practical reasons of the usual resource constraints, the PIP team should give consideration to the skills of available staff when designing the data collection instruments, developing instructions and planning training.

The field test should include a careful assessment and documentation of needs for further training or different skill sets, and the data collection plan should be adjusted accordingly.

Step 6 *What should be included in the written PIP plan?*

A number of the preceding steps concluded with a recommendation that the particular task be included in the written PIP plan. At this point, these should be organized into a clear and succinct written plan to serve as a guide for everyone involved in the project.

The plan should include at least:

- A rationale for why this topic was selected, including a description of the RCA or analogous process.
- A careful definition of the study population, with explicit criteria for inclusion and exclusion.
- Definitions of indicators.
- Definitions of indicator criteria, including specification of numerator and denominator.
- A concrete and detailed description of the proposed intervention.
- A description of the data collection and analysis plans.
  A timeline for each task, identifying the individual responsible.

You may wish to develop a template that would serve to establish consistent and comprehensive plans. An example from the National Committee for Quality and Accountability (NCQA) is available at: http://www.qualityprofiles.org/qia_Form/index.asp.

It should be emphasized that this plan is subject to review and revision at any time based on new information acquired in the course of the project. For example, the proposed intervention may well be modified based on results of the baseline measurement.

Phase III

To sample or not to sample?

Steps in this phase

1. What is a sampling plan?
2. How do I define the study population?
3. How do I define the sampling frame?
4. How do I decide on the type of sampling?
5. How do I determine the sample size?
6. What should I include in the sampling strategy?

Once you have identified the information necessary to answer your study question, you are faced with a choice between two options: whether to collect every instance of the required information (a census) or some representative subset (a sample). Sampling is most frequently applied in the case of enrollee surveys, but it is a consideration whether the unit of information is MHP enrollees, individual providers, processes of care, treatment episodes, medical records, or any other type of information.

Choosing between a census or a sample involves certain pragmatic tradeoffs. When the information required for the indicator is from an electronic database, generally sampling is unnecessary. Otherwise sampling offers the advantages of:

- Costing less
- Taking less time
- Providing information that is more current

The key issue in sampling is to insure that the units selected (individuals, clinics, records, etc.) are *representative* of the total population of interest. A flawed selection process produces a *biased sample*—one that differs in some systematic way from the population as a whole. Inferences about the population based on information gained from the sample are therefore likely to be inaccurate and misleading.

It is not the case, however, that census data necessarily avoids this risk, as the census enumeration process also may be flawed, resulting in a failure to capture some groups that differ systematically from the larger population as a whole. The classic example is the case of the homeless population, which is likely to be omitted in a population census that does not employ special outreach techniques.

Census data may be more credible for a lay audience, though this confidence is not always justified, for the reasons cited above. A carefully drawn sample will be more representative than a census that omits important subgroups.

Sampling can require considerable technical expertise in order to avoid bias. A number of excellent texts and web-based resources are available, some of which are listed in the Appendix E. The fundamental purpose of this overview, therefore, is to enable you to decide whether to design the sample selection process yourself or to seek expert consultation from a sampling statistician and, if the latter, to enable you to identify more precisely the nature of the consultation required.

Step 1 *What is a sampling plan?*

If you decide on a sampling approach, you should begin by developing a sampling plan. The sampling plan consists of the following elements:

- Definition of the study population
- Definition of the sampling frame
- Decision of type of sampling
- Determination of the sample size
- Specification of sample strategy

These steps are described in detail below. The specific decisions at each step should be recorded in the written PIP plan.

Step 2 *How do I define the study population?*

The study population, sometimes referred to as the "at risk" population, or simply the "universe" consists of all possible units of analysis, e.g., all of the MHP's enrollees, all elderly enrollees, all clinical staff, etc. This is the target population for the PIP intervention.

The written PIP plan should include a description of this population, including its size, and its socio-demographic and clinical characteristics.

Step 3 *How do I define the sampling frame?*

The sampling frame consists of members of the study who qualify for the study according to certain specified *inclusion and exclusion criteria.* An example of an inclusion criterion would be enrollment for a minimum period, e.g., three months, on the grounds that more recent enrollees would yet to have encountered the quality issue under study. An example of an exclusion criterion would be, for a study topic involving access for minority immigrants, if they had been resident in the U.S. for longer than say, ten years, on the grounds that they might be expected to be sufficiently acculturated that their participation in the study would mask the quality issues involving more recent immigrants.

The definition of the sampling frame, including specific descriptions of exclusion and inclusion criteria, and the numbers of both the included and excluded, should be included in the written PIP plan.

Step 4 *How do I decide on the type of sampling?*

The following offers a general overview of sampling methodology. More detailed and technical explanations are available from a number of texts on survey and sampling methodology (see the Resources). This discussion may serve as a guide to determine when you might wish to consult a *sampling statistician* and what questions you may want to pose.

The two general types of sampling used in health services research and performance measurement are *nonprobability* and *probability* sampling. Each of these has several subtypes which are briefly described.

What is probability sampling?

There are multiple types of sampling methods (convenience sampling, quota sampling, etc.) but by far the most effective method for insuring that your sample is representative is *probability* (or random) sampling. Probability sampling is simply any method that leaves selection of units from the population totally to chance. In simple probability sampling therefore, each individual unit of analysis (consumer, medical record, etc) in the entire group of interest has an equal probability of being selected.

Stratified random sampling is used when the target population consists of non-overlapping sub groups or strata, such as gender or race. The purpose of this method is to gain greater precision in population estimates when there is reason to believe that sub-groups differ systematically in how they respond. Thus, stratified sampling will provide a more precise estimate for each of the groups than it would for the population as a whole where these group differences are pooled. Stratified sampling also provides information about subgroups of interest. Stratified random sampling requires more information about the population and requires a larger overall sample than simple random sampling. It also requires somewhat more complex data analytic techniques and specialized software. For these reasons, both planning and analysis are likely to require consultation with a survey statistician.

Cluster sampling is used when a comprehensive sampling frame is unavailable. Units in the population are gathered or classified into groups, similar to stratified sampling. This method requires prior knowledge about the population. Once clusters are identified, a random sample of clusters is selected.

Step 5 *How do I determine the sample size?*

The accuracy of a survey, i.e., the extent to which results are consistent with the "true" values in the population as a whole, depends in large part (though not exclusively) on the size of the sample.

The decision about appropriate sample size depends on two considerations. The first is the desired *level of certainty* for the estimate of the population value, and the second is the acceptable *margin of error*, or range of values within which the population value may differ. In fact, the two are interrelated: sample size calculations identify the probability (e.g., 95 percent) that the true population mean will

fall within a specified range around the sample mean. In other words, determination of sample size rests on a trade-off between confidence and precision: one may be reasonably confident, say with a 95 percent probability, that the difference between the sample and population means are within 10 percent. Likewise one might say with less confidence, e.g., a 90 percent probability, that the means of the same sample and population fall within 5 percent of one another.

This decision is based on pragmatic or policy grounds. A larger sample (and consequently more expensive survey) offers more precision at a higher level of confidence. The organization must decide whether the added certainty and precision is worth the additional cost—that is, whether the range of possible values is to uncertain for decision making purposes, and this of course will depend on the nature of the decision.

Most introductory texts on sampling methodology explain how to calculate sample sizes, and a number of interactive sample size calculators, offering various levels of explanation, are available online for public use. A popular one with clear explanatory text and relevant formulas is on the website of Creative Research Systems, Inc. at http://www.surveysystem.com/sscalc.htm.

Sample size and response rates

An important consideration about sample sizes that you should keep in mind is that although sample size is often emphasized in discussions of survey methodology, it will be irrelevant if your response rate is inadequate. If your sample is biased due to non-response (systematic differences between respondents and non-respondents), increasing the sample size will not reduce the bias. A better of use of resources than, for example, sending out a single mailing to a large number of people, would be to target a smaller sample and devote more resources to methods for enhancing response rates. These methods are described in more detail below.

What about sampling from diverse population densities?

Many geographically extended mental health systems serve areas that include both densely populated urban centers and sparsely populated rural areas. This difference in denominators raises issues for sample size, in that the proportion sampled in an urban area may be perfectly adequate for purposes of statistical analysis, but the same proportion in a rural area will capture only a small number of people, insufficient for analysis.

The same problem may arise with subgroups in the population who may be small but nonetheless a focus of concern or interest, such as minority groups. A simple random sample may not pick up enough of this group to draw inferences about the sub-population as a whole.

The solution to this problem is to stratify the sample according to the variable of concern e.g., region, race, etc. and then to "over-sample" those in the smaller group, that is to draw a specified greater proportion from the sampling frame, thus obtaining an adequately precise estimate of the population mean for that group.

This approach, as noted above, raises some complications for data analysis, however, and therefore probably requires consultation with a survey statistician.

Step 6 *What should I include in the sampling strategy?*

Once you have identified the sampling frame and determined the sample size you will need to specify, carry out and document the procedures for extracting the sample from the sampling frame and managing the resulting dataset.

For a random sample the simplest method of drawing the sample is to assign a random number to each individual in the sampling frame, order the list according to the assigned number and select from this list in order until reaching the specified sample size. A number of random number generators, including programs for entering a data list (e.g., members of the sampling frame) to be assigned numbers, are available online, for example at http://graphpad.com/quickcalcs/randomselect2.cfm. Other methods are described in survey research texts and online sites listed in the Resources section.

In consultation with your IT staff and data analyst, you should determine the format and mode of delivery to the predetermined data repository.

Phase IV

When and how do I conduct a survey?

Steps in this phase:

1. How do I define the purpose of the survey?
2. Who is the audience for the survey results
3. How do I select a survey instrument
4. How do I maximize response rates?
5. What should I include in the survey implementation plan?

The situation in which you may consider a survey is when the information required to answer your study question is not available from existing data sources, such as administrative information systems or clinical records. Surveys are most useful for accessing consumers' and family members' *experience of care* (i.e., satisfaction), but may be used to obtain many other types of information, such as utilization (when administrative data is unavailable or unreliable), the perceptions of other stakeholders such as individual providers, or other systems with which the MHP interacts (for example, primary care providers).

However, to serve this purpose effectively, that is to produce results that are reliable and valid, the survey must be designed and conducted according to established methods. These are described in an extensive body of literature, and the following is only an overview to convey what you must consider if you decide to conduct a survey. Sources for additional information are identified in Appendix E.

Step 1 How do I define the purpose of the survey?

The purpose of the survey should, of course, reflect the study question, and should be defined with the same rigor as discussed above. That is, the PIP team should develop a written statement that addresses two questions:

1. What is it that we want to learn from the survey?
2. How will we use the results?

The general response to the first question is: we wish to learn something about some aspect of the quality of care provided. This should be clearly expressed in specific detail, however, again by asking a series of questions, such as "What aspect of care?" for example, access; "What aspect of access?"—access to new generation anti-psychotic medications; "Access to medications by whom?"—minority subpopulations; "Which minority subpopulations?" etc.

Similarly, the answer to question #2 is: to improve the way in which care is delivered, but the statement should spell out the nature of the improvement, i.e., the intervention. This will guide the design of the survey instrument to be sure that it captures the information necessary to target the intervention.

As with the study question, the survey statement of purpose should be revised and refined in response to input from stakeholders and additional information until the team is confident that it is clear and unambiguous, at which point it should be added to the written PIP plan.

Step 2 *Who is the audience for the survey results?*

The question of who will use the information obtained by the survey will influence the design, protocol for administering the survey, and format for reporting results. Common audiences for consumer surveys in health care are:

- Consumers themselves, to be informed about their MHP;
- MHP management to make decisions about quality initiatives and other operational functions;
- Policy makers, to assess and plan broader policy initiatives.

Surveys of individual providers or other stakeholders are most likely to be used for management purposes, and therefore intended for MHP managers, but, even with this, it would be beneficial to specify the level and category of management.

Being clear about the audience for the survey results, as well as the purposes of the survey, the PIP team may be more certain of obtaining the appropriate information. The written description of the survey plan should therefore explicitly define the intended audience and the nature of the information they require.

Step 3 *How do I select a survey instrument?*

In choosing a survey instrument, the PIP team is faced with three options: 1) Using an existing instrument; 2) Developing a new instrument; or 3) Modifying an existing instrument. There are tradeoffs among these choices.

Developing a survey instrument is a resource intensive activity requiring considerable investments of time, labor and technical expertise. An existing survey that has been adequately tested is therefore an efficient choice. However, the particular topic for the PIP may call for information that no existing instrument can supply, in which case the PIP team, wishing to avoid the costs of developing a new one, may choose to modify one that is close to what required. Depending on the nature and extent of the modification, however, this may cancel out the benefits of a well-tested survey if the modifications significantly change the properties of the instruments.

In any case, the two principle considerations in selecting an instrument are its *reliability* and its *validity*. As described above, these are technical terms with complex definitions that are presented only in overview here. Should these issues arise in the course of planning a PIP, the team is advised to explore the extensive literature on the subject and/or consult with a survey expert, such as may be found in an EQRO, or an academic health services research organization.

The survey instrument should ask questions about socio-demographic and clinical characteristics to the extent feasible, as these are important for analysis, as well as for estimating non-response bias as described below.

Considerations in designing survey instruments:

The following, adapted from SURFSTAT.Australia at http://www.anu.edu.au/nceph/surfstat/surfstat-home/surfstat.html summarizes some key points to keep in mind when designing the survey form:

- Be clear about the purpose of the survey, the information you will need and how it will contribute to the research question
- Consider the pros and cons of using a survey and be sure that it is the most suitable method
- Decide what type of information your questionnaire will be seeking and whether open (where respondents can create their own answers) or closed (respondents have a choice of a limited number of responses) questions, or perhaps a combination of the two, are most appropriate
- List the broad question areas and then write specific questions, keeping in mind the respondent population and make sure that questions are simple enough to be easily understood
- A question to measure someone's attitude or behavior relating to a certain topic such as cigarette smoking, may be more sensitive than a question about an attribute, such as the person's gender or nationality
- The questions should follow a logical order. The questionnaire must look and read simply and clearly
- Pilot-testing is essential, both informally among colleagues and formally among a sample of respondents
- Establish coding categories for each question to facilitate data entry
- Make decisions concerning the analysis to be conducted while the questionnaire is being constructed.

Step 4 *How do I maximize response rates?*

The difficulties of achieving satisfactory response rates for surveys of both Medicaid enrollees and persons with serious mental illness are well known. The strategy for obtaining adequate response rates, therefore, should be considered carefully in advance and include contingency strategies to address potential problems.

A very common problem with these populations (and poverty-level groups in general) is the lack of reliable contact information. Medicaid enrollment files are frequently out of date, inaccurate or incomplete. Members of these groups tend to move frequently, often among fairly anonymous settings, and a disproportionate number lack telephones entirely. The MHP, therefore, should make it a priority to obtain reliable contact information for all enrollees—this might be a PIP in itself.

What is the best method of contact?

There are three primary methods of contacting potential respondents: mail, telephone and personal interviewing. There is no one method that is preferable or most effective in all circumstances and the choice will depend upon familiarity with the circumstances of the MHP's enrollees as well as pragmatic considerations of cost and time constraints.

Frequently the most effective approach is a combination of these: an initial mailing, a follow-up post card to non-responders.

As discussed above, your response rate is, within limits, more important than sample size, in that you may draw no valid inference from a sample biased by non-response, whereas the risk from having an undersized sample small sample size is the lesser one of possibly failing to detect smaller differences.

Step 5 *What should I include in the survey implementation plan?*

Detailed procedures for implementing the survey should be developed and recorded in the written PIP plan. These should include at least the following:

- Key staff and their responsibilities
- Deliverables with time lines
- Plans for addressing any potential problems that may be identified
- Decision rules for handling missing responses and incomplete forms in the analysis

The implementation plan should include in addition procedures for management and analysis of data and reporting results. These apply to PIP data generally and are described in more detail below.

Phase V

What should I include in the data management and analysis plan?

Steps in this phase:

1. How do I develop and disseminate a data archive?
2. What should I consider in the data analysis?
3. How should I judge performance?
4. What is case mix adjustment?

Step 1 *How do I develop and disseminate a data archive?*

You should develop a detailed description of the data file known as the data archive and a plan for making it available as appropriate. The following are the necessary steps:

1. Obtain the original data collection form or survey so users of the archive can see how a question was asked and coded in order to formulate their own questions and hypotheses.

2. Create data sets in the preferred statistical software (e.g., SAS, SPSS) from the raw data.

3. Create a data dictionary for each data set in the archive, including description of the data set, a codebook with variable descriptions and values, frequency runs on all variables and selected crosstabs.

4. Run and review frequencies for accuracy. For data sets that are updated on a regular basis, it is important to insure consistency from one period to the next.

If the data file is to be made available for ongoing management or QI purposes, you will also need to:

5. Establish a policy for data requests and for subsequent analyses.

6. Establish the availability of an analyst who can explain how the data set can answer questions of interest.

7. Establish a process for disseminating findings of subsequent analyses.

Step 2 *What should I consider in the data analysis?*

This step consists of carrying out the procedures described in the data analysis plan developed prior to data collection (Phase II). The key consideration for both analysis and reporting of results is that they should be as simple and clear as possible within the bounds of methodological soundness. Thus, you should bear in mind that the purpose is not that of a research project, to discover new knowledge, but simply to provide a measure of the MHP's performance on the study indicators, as

compared to some pre-determined standard, goal, previous level of performance, or benchmarks as described in Step 5 below.

Step 3 *How should I judge performance?*

Once your analysis of the data has established the level of performance with regard to the study question, the sometimes-difficult task is to determine whether this is good, bad or indifferent. Several types of data can provide the basis for making this determination: standards, means, norms and benchmarks (Hermann and Provost 2003).

Standards (targets) are performance expectations established by individuals or groups, such as accreditation agencies, purchasers, such as state agencies, or the MHP itself. They may be based on scientific evidence, expert opinion or arbitrary decisions. The limitation of standards is that it may be difficult to ascertain the evidence for whether they represent an appropriate level of performance.

Means are derived from research or QI initiatives. They are more evidence-based than standards, but they have two limitations: first, they represent only the average, and not the highest possible level of performance, and second it may be difficult to ascertain how much the population from which they were derived is similar or different that of the MHP conducting a PIP.

Norms are like means except instead of originating from studies, they are average results for large, representative population-based samples. The limitation of these, are similar to that of means, is that they may have the effect of reinforcing the status quo rather than establishing goals for improvement.

Benchmarks are goals established through statistical processes developed in industrial QI models. Typically benchmarks represent the achievements of the highest performing organizations of a similar type. Their advantage over other types of comparison is that they are a means of establishing quantitative levels of excellent but achievable performance. Thus a benchmark goal for improvement may be set at the level of, for example, the top 10 percent of Medicaid behavioral MHPs.

Step 4 *What is case mix adjustment?*

Anytime your assessment of performance involves comparisons with some other group, you must consider the need for case mix adjustment. Case-mix adjustment is the process of statistically controlling for group differences when comparing non-equivalent groups on outcomes of interest. The groups may be treatment agencies, consumers, providers, programs, regions, or states. It is done on a post-hoc basis, when performance measures are collected for already-existing groups, i.e., when there is no procedure such as random assignment to minimize differences between groups prior to measurement. Typically case-mix adjustment uses regression analysis techniques to adjust for differences between groups in different system that might affect outcomes apart from the quality of the care provided by each system.

For a more technical explanation, including exercises with simulated datasets to demonstrate case-mix adjustment techniques, see (Hendryx 2004), available at www.tecathsri.org

Phase VI

How do I develop the improvement strategy?

Steps in this phase:

1. What should I consider in reviewing the data?
2. What should I consider in reviewing the data?
3. How do I Identify barriers to quality?
4. How do I develop the improvement intervention(s)?

Having measured baseline performance in the area of the study topic and, assuming a judgment that improvement is both desirable and feasible, you are now prepared to design the improvement strategy.

Basic Principles: What is an improvement strategy?

An improvement strategy is an intervention designed to change behavior at an institutional, practitioner or beneficiary level. This should be a system intervention that might be expected to have a lasting effect as opposed to a one time intervention, such as a staff training, that may produce short-term improvements but not be sustained (Protocol).

Step 1 *How do I Identify barriers to quality?*

In most cases, the quality gap will be readily apparent, but identifying the most appropriate and effective strategy requires more effort. This step represents a continuation of the root cause analysis described above (Phase I, Step 3), sometimes referred to as "barrier analysis." The review needn't be particularly formal, but may include methods such as focus groups and simple surveys. Ideally it would include representatives of all stakeholder groups involved in, or affected by, the process under study.

For example, the study question may be how to reduce time spent in the waiting room, an issue identified by the complaint system and enrollee satisfaction survey. Clinicians may suggest that the schedule backs up because routine brief medication management visits are extended due to the needs of a certain subgroup of enrollees. Front desk staff may see that the method of scheduling appointments creates delays. Enrollees may identify difficulties with access to the building as a factor causing them to be late.

The PIP team should record detailed descriptions of these barriers in the written PIP plan.

Step 2 *How do I develop the improvement intervention(s)?*

With barriers clearly and comprehensively described, the PIP team and participating stakeholders are in position to devise the intervention, which typically consists of a number of actions corresponding to the number of barriers identified. Those who have identified barriers will very

often have ideas for how they may be eliminated, while others may be able to assess the feasibility of these recommendations or identify ways of modifying them to be more effective.

Other sources of ideas for interventions and alternative best practices are available as well, including:

- Facilities that are already achieving excellent results
- Research studies
- Disease management models
- Other reports of promising practices

You should allow adequate time for reaching consensus, as proposed interventions that are not fully understood often create anxiety and resistance. You might wish to prepare and circulate materials describing potential interventions for discussion among the team, and to employ some of the formal consensus-building strategies described in the materials cited in the Resources section.

What are some types of quality improvement interventions?

In Table 1 Shojania (2004) provides taxonomy of QI strategies derived from a review of studies of QI practices.

Table 1. Taxonomy of QI strategies with examples of substrategies

QI strategy	Examples
Provider reminder systems	• Reminders in charts for providers • Computer-based reminders for providers • Computer-based decision support
Facilitated relay of clinical data to providers	• Transmission of clinical data from outpatient specialty clinic to primary care provider by means other than medical record, e.g., phone call or fax
Audit and feedback	• Feedback of performance to individual providers • Quality indicators and reports • National/State quality report cards • Publicly released performance data • Benchmarking – provision of outcomes data from top performers for comparison with provider's own data
Provider education	• Workshops and conferences • Educational outreach visits (e.g., academic detailing) • Distributed educational materials
Patient education	• Classes • Parent and family education • Patient pamphlets • Intensive education strategies promoting self-management of chronic conditions
Promotion of self-management	• Materials and devices promoting self-management
Patient reminder systems	• Postcards or calls to patients
Organizational change	• Case Management, Disease Management • TQM, CQI techniques • Multidisciplinary teams • Change from paper to computer-based records • Increased staffing • Skill mix changes
Financial incentives, regulation, and policy	Provider-Directed: • Financial incentives based on achievement of performance goals • Alternative reimbursement systems (e.g., fee-for-service, capitated payments) • Licensure requirements Patient-Directed: • Co-payments for certain visit types • Health insurance premiums, user fees Health System-Directed: • Initiatives by accreditation bodies (e.g., residency work hour limits) • Changes in reimbursement schemes (e.g., capitation, prospective payment, salaried providers)

Phase VII

How do I implement the improvement intervention?

Steps in this phase:

1. Review the Plan Do Check Act cycle
2. Assess the requirements for leadership
3. Assess the role of providers/practitioners
4. Identify and adjust for potential confounding factors
5. Assess the human factors in organizational change
6. Package your data
7. Pilot test the PIP
8. Assess the requirements for sustaining improvement

Basic Principles: Continue to Learn

Organizational change intended to improve quality is always a complex and difficult process. At the conclusion of every step, therefore, you should ask yourself the question, "What did we learn?" At times, it may be that this question can be answered only by reactivating the Root Cause Analysis process in its entirety. Then, based on that information, you should ask the second question, "Given what we know now, how should we modify our plan?"

Much of the activity in the implementation phase is pragmatic, detailed oriented, and focused on organizational factors related to the capacity for change. The type of organizational change required for improvement in the processes and outcomes of care is complex and difficult requiring significant investments of effort, time and other resources. Given these "close to the ground" features, the implementation phase may not always receive the same attention as more exploratory and technical steps preceding it. It must be emphasized, therefore, that real, sustained improvements in quality require *a continuous cycle of measuring and analyzing performance, and developing and implementing system-wide improvements.*

The key words in the above sentence are *continuous cycle.* Lasting quality improvement of any importance seldom results from a one-time intervention, after which the team can move on to other concerns. Likewise it is seldom linear, but rather, it typically requires reassessment and revision, based on information gathered along the way. Therefore, though the following activities are identified as steps, they should not be considered as sequential but rather as considerations to be revisited repeatedly through the implementation phase.

Step 1 Review the Plan Do Check Act cycle

It is through the cyclical and iterative nature of the process that a PIP succeeds in uncovering complex and multifactorial quality gaps and overcoming the inherent inertia of organizations to achieve sustained improvement. Each stage in the Plan Do Check Act cycle involves in itself a

process of review and revision. In the Do (implementation) stage, for example, the pilot testing may very likely reveal that the planned intervention proves to be unfeasible, inadequate or off target. This is not grounds to consider the Plan phase a failure, however, but rather to revisit the process, review the data, perhaps conduct an entirely new RCA, and then to revise, or even totally redesign, the intervention. The same applies to each subsequent stage in the cycle. The more carefully you conduct these within-step cycles the more likely you are to be rewarded for your efforts by a project that effects true change.

Step 2 *Assess the requirements for leadership*

The literature of organizational change and technological innovation and diffusion emphasizes the key role of leadership, both in the form of authorities able to integrate the goals of quality improvement with the overall strategic mission of the organization, and also leadership by "champions" who are not necessarily in positions of high authority but promote adoption of innovation through enthusiasm and example.

The process of implementation should include maximum efforts to identify, recruit and foster the development of both kinds of leaders.

Step 3 *Assess the role of providers/practitioners*

Enlisting the support of individual staff (clinicians, administrative staff, etc.) is probably the most important ingredient for the PIP's success, as the goal of most intervention is achieve changes in the behavior or practice patterns of these groups. A considerable amount of research has been invested to identify the most effective ways to achieve changes in practice patterns and provider behavior, and though findings are promising, no completely effective strategy has yet to be found. It is certain, however, that meaningful change is much more difficult to achieve without the active involvement and support of those affected by the intervention.

Step 4 *Identify and adjust for potential confounding factors*

Another challenging but important requirement for the success of a PIP project is to clearly see the primary study question, and to maintain focus upon it throughout the project cycle. Behavioral health care organizations are complex, and the task of preventing intertwining processes from obscuring the intended target requires discipline and clear thinking.

Taking care earlier to define a clear and focused study question (Phase 1, Step 2) will do much to avoid the possibility of the PIP's impact being diffused by issues that may be important but lie outside the scope of the project. To maintain focus as you proceed with implementation, you may wish to regularly revisit the written statement of the study question, and you may find that the revisions or further clarifications are called for in the light of new information.

Step 5 *Assess the human factors in organizational change*

Inquiry into the psychosocial aspects of change has produced an enormous, multi-disciplinary literature, and though your PIP team should not expect to assimilate this knowledge, it would be beneficial to include someone with some background in this area, and in any case, to keep these issues in the forefront of your attention as your proceed with your PIP.

Two activities in particular can serve to prepare the way for introduction of the PIP: assessing organization readiness and social marketing of the PIP. These are only briefly mentioned here, but there is an extensive literature describing both of these activities, some of which is cited in the Appendix E.

What is organizational readiness?

Organizations vary extensively in their responsiveness to innovation, their capacity to adapt to changing circumstances and their willingness to adopt new approaches. Experts in organizational behavior have developed a variety of methods for assessing and addressing these characteristics. The various branches of the quality improvement field (total quality management, continuous quality improvement, etc.) have been particularly attentive to these issues

What is social marketing?

Social marketing is the process of using methods developed in commercial public relations to promote social programs. This approach was developed first in the public health field, where it was used to disseminate programs such as anti-smoking campaigns, and it has since been adopted elsewhere in various non-profit enterprises. With respect to quality improvement, it offers tools to educate and prepare the workforce for changes that otherwise might be threatening or confusing.

Step 6 *Package your data*

Reporting quality information occurs at two stages of the PIP process. The first, described here, consists of presenting information about the need and potential for improvement in preparation for the proposed intervention. The second stage, presenting results of the completed PIP, is described below in Phase VII.

As performance measurement and quality improvement in health care (as well as other areas such as government and industry) have advanced over the past 10+ years, experience has shown that organizational change that produces real and sustained improvement requires more than simply producing information about quality gaps: it depends as well on effective strategies for presenting and disseminating the information. Relevant to the complexity and difficulty of achieving organization change for purposes of quality improvement, two fundamental principles stand out:

Information about quality gaps should be presented to all relevant stakeholders, and should be tailored to their respective needs, interests and capabilities. It may be worth the additional effort to conduct focus groups to assess stakeholders' requirements and preferences for quality improvement information.

2) Information should be presented through multiple formats or media such as technical reports, brief summaries, presentations at meetings, graphic and narrative representations, etc.

Step 7 *Pilot test the PIP*

As suggested in Step 1, pilot testing is a critical task exemplifying the cyclical nature of the PIP and should be conducted thoroughly. The pilot test may be on a relatively small scale but should be designed to represent the entire range of anticipated circumstances. For example, if your study requires record abstraction at several sites, you need not review many records in your pilot test, but you should review a few at every site.

You should anticipate the likelihood that the pilot test will indicate the need for extensive modifications in nearly any aspect of the PIP, including the data collection procedure and the intervention plan. You should not enter into the pilot test with an expectation of confirming your initial approach, therefore, but rather to identify any modifications that might be required.

Step 8 *Assess the requirements for sustaining improvement*

Your PIP group should periodically revisit the target area project to ensure that improvements have been fully integrated into the ongoing procedures of the organization. This process may take the form of a less intensive replication of the data collection and analysis process (Phases II-IV). These procedures need not be as rigorous as in the initial cycle (for example a convenience sample may substitute for a random sample, or only one of a set of original indicators may be examined), but they should be adequate to identify evidence of diminished results. Where such evidence is found, you should conduct a more thorough analysis and, if necessary, re-activate the project to identify what barrier remain or have re-occurred and devise new interventions as appropriate.

Phase VIII

How do I report the results of the PIP?

Steps in this phase:

1. What should I include in the PIP report?
2. What should I consider in designing my report
3. How do I disseminate the report?

Basic Principles: Reach your audiences

Every PIP should culminate with a single, comprehensive report presenting the information described below. In addition, however, as with the initial quality information discussed above, you will likely wish to provide individual groups of stakeholders with more focused reports and presentations, with the format and content to be determined by the needs of the particular audience.

Step 1 *What should I include in the PIP report?*

The two fundamental considerations for the report are *credibility* and *utility*. The primary report should be a comprehensive description of all phases of the PIP; at the same time, it should be focused and succinct. It should include a balance of narrative text and graphics, with graphics capable of standing alone.

At a minimum the report should provide the following information:

- The study question.
- The indicators used to address the question, and the indicator criteria.
- Quantitative findings in response to the study question.
- The meaning of these findings.
- Interventions implemented to address them.
- The outcomes of these interventions, including quantification of the effect on quality of care, expressed as the number and proportion of members directly affected.
- Any secondary or expected results such as cost savings and other benefits of the project.
- Any unexpected negative effects, and actions taken to address them.
- Evaluation of the PIP process, including description of did and did not work, barriers encountered and methods of addressing them, and lessons learned.
- Proposed follow-up activities.
- The effect of the project on the capacity to affect improvement, including factors that led to capacity building and what obstacles that were overcome or proved insurmountable.

Consider a template

You might wish to develop a form or template for reporting in condensed form the basic elements of a PIP that are consistent from one project to another. An example of such a form is NCQA's

Quality Improvement Activity Form, available at
http://www.qualityprofiles.org/qia_Form/index.asp

You should remember that this is not research report, and therefore requires little discussion about contributions to new knowledge or technical issues of methodology. Discussion of methodology should serve the purpose of making results understandable and credible, and should be presented without jargon. The area where this balance is particularly important—and challenging—is the discussion of statistical procedures. The Agency for Healthcare Research and Quality (AHRQ) , on its website dedicated to designing quality report cards for consumers recommends that the two statistical concepts that report authors should understand thoroughly and describe clearly are *confidence intervals* and *statistical significance*. Their discussion of these concepts is presented below.

Basic Principles: Two Statistical Concepts You Should Understand (From AHRQ http://www.talkingquality.gov)

Confidence intervals

A confidence interval refers to the statistical likelihood that a score falls within a given range around an estimate. This is important because nearly all health care quality scores are developed using a statistical sampling method, which means that there is some uncertainty about whether the sample reflects the population.

The confidence interval tells you how confident you can be that the score for the sample represents the score for the entire membership or population. For instance, a 99 percent confidence interval means that, if you drew numerous samples, 99 percent of the estimates would fall within the given range. A narrower range (e.g., a 90 percent confidence interval) would give you less certainty about the estimate. A 95 percent confidence interval is a standard used in scientific research, because it creates a high degree of certainty that the sample reflects the population. It also tends to be the compromise position between the high degree of certainty desired by providers and the lower confidence interval desired by purchasers, who often want to see a more even distribution of scores across the tiers.

Statistical significance

Statistical significance tells you whether two scores are really different from each other. If one health plan has a mammography rate of 77 percent and another has a rate of 80 percent, is one truly better than the other? We determine statistical significance by checking to see if the confidence intervals around the scores overlap; confidence intervals reflect the uncertainty of the estimated rate, which is based on a sample drawn from the larger population. If they do overlap, the scores are basically comparable. If they don't, you can say they are statistically different.

However, statistical significance is not the same as practical significance. If you have enough data points (i.e., if the sample size is large enough), every difference may be statistically significant, but that doesn't necessarily mean

that they are worth drawing attention to. For example, satisfaction scores often have a small range of variation, maybe 10 points. If you were to try to compare plans based on these scores, you would need to consider whether a very small difference (e.g., 91 percent versus a 92 percent) really matters. Will consumers get noticeably better care? Similarly, sponsors have to question whether differences are clinically meaningful. For instance, are consumers likely to experience a substantial difference in outcomes because of minor differences in rates? Maybe yes, maybe no, but you should know that before you present the information to consumers.

Step 2 *What should I consider in designing my report?*

As discussed in Step 6 of Phase VI above, the past few years have seen the development of an entire science of reporting quality information, supported primarily by government agencies and private foundations. Much of this has focused on "report cards," i.e., publicly available health plan performance information intended to promote consumer choice based upon quality. Recommendations for designing effective quality reports are provided on a variety of websites, identified in Appendix E.

To some extent the findings of studies in this area and the principles developed are applicable as well to the purposes of quality improvement involving a more diverse assortment of users of quality information, including providers, managers, policy makers and other stakeholder groups. Whereas the design of quality report cards is shaped by the primary purpose of facilitating consumer comparison of organizational performance, PIP reports will be directed toward stakeholders within the organization who are concerned about quality issues.

The design therefore should support this function by first *placing the study question in clear focus* and then presenting the information relevant to it (as described in Step 2, above) in a manner that serves the purpose of 1) identifying the need for improvement; 2) describing the chosen strategy for accomplishing improvement; 3) presenting the results of that strategy.

What about charts and graphs?

You may wish to present a pictorial representation to supplement your narrative explanation of the intervention and the results of measurement.

NCQA recommends attaching a graphic representation for any project involving more than two measurement periods in order to show the relationship between the timing of the interventions (the cause) and the result of the re-measurement (the effect) (http://www.qualityprofiles.org/qia_Form/QIA_Form_Instructions.pdf)

There is no single best type of chart or graph. You should use whatever type of chart (line, bar, mixed) that most clearly represents both the interventions and the measures. A simple line graph might be appropriate for service activities with multiple data points, while a bar chart might be

more appropriate to show changes in measures with annual measurement points. Control charts that display upper and lower confidence limits, may be used to demonstrate the stability of the measure over time.

There is no one correct way to represent these elements and the particular approach will vary depending on the intended audience. The same principles apply here as in most other steps of the PIP process: test your method with representatives of your target audience(s) and modify it accordingly.

Step 3 *How do I disseminate the report?*

The dissemination strategy should be developed early in the planning process, and should follow the same general principles as those informing the design of the report; that is, dissemination should be targeted at those who are interested and affected, and it should be conducted through multiple media as appropriate for maximum effectiveness.

Dissemination is likely to involve more than just the report itself but also various forms of summary information packaged as appropriate to the purpose and the audience. Moreover in the interests of sustaining the improvement gains achieved by the PIP, dissemination will be an ongoing process, with follow-up materials such as wall charts, reminders, prompts, etc.

Phase IX

How do I evaluate the PIP process?

Steps in this phase:

1. How should I define the purpose of the process evaluation?
2. How should I define the evaluation questions?
3. How do I assess the PIP implementation process?
4. How do I assess the quality of the data collected?
5. How do I assess the cost-effectiveness and organizational burden of the PIP?
6. How should I define the purpose of the impact evaluation
7. How should I define impact evaluation questions

Basic principles: Evaluation for improvement

If performance improvement is based on the tenet that evaluative information can improve service delivery, then the performance improvement system itself should use evaluative information to improve its own processes:

You should begin the evaluation process with a clear and shared understanding of the purpose of the evaluation

You should consider both process and impact evaluations, as described below

You should include, as a major goal of the evaluation process, assessment of stakeholder satisfaction because of its importance to the success of the project.

NOTE: The following steps are adapted from The MHSIP Quality Report Toolkit (the Evaluation Center@HSRI, 2004). That source applies to system-wide performance measurement projects such as quality report cards. *It is unlikely that internal evaluation of a PIP project would require all of these steps in this level of detail; they are described here however to serve as a checklist for ensuring that all aspects of a comprehensive evaluation, such as those conducted by an external reviewer, have been considered.*

PIPs are to be evaluated by EQROs, in accordance with the guidelines specified by the Protocol cited throughout the Technical Assistance Manual. The following recommendations describe considerations for *internal evaluation* of the process of developing the PIP and of the results of this process. These two overarching goals lead to two distinct types of evaluations of a performance improvement project: 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 PIP, in terms of its appropriateness, efficiency, resource expenditure, etc. The impact evaluation has as its overall goal to assess the degree to which the PIP achieved its goals of real and lasting improvement.

As with any type of evaluation, the planning of an evaluation of a PIP 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.

How do I conduct the PIP process evaluation?

Step 1 *How should I define the purpose of the process evaluation?*

You should 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. Some potential purposes are:

- Obtaining feedback to improve the process for developing and implementing PIPs
- Assessing stakeholder satisfaction and buy-in with the process.
- Modeling the use of information for improvement.
- Analyzing cost-benefit. The process evaluation may include an analysis of the cost of developing and implementing the PIP in relation to benefits expected
- Comparing the project with national models

Step 2 *How do I 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 are some questions that might be addressed in a PIP process evaluation.

- Were stakeholders satisfied with their level of involvement in the process?
- Was the development process efficient?
- Were there clear lines of responsibility for tasks?
- How much time did the process consume (for both project and other stakeholder groups)?

Step 3 *How do I assess the implementation of the PIP?*

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

Technical assistance: Was sufficient technical assistance provided to persons and organizations required to participate in the project? Was it provided to the right individuals/organizations? Is it appropriate for its intended audiences?

Training: Was sufficient training provided? Was it provided to the right individuals/organizations? Was the level of training appropriate for its intended audiences? Were there appropriate mechanisms to ensure that training diffuses through the system the extent required by the intervention?

Barriers to implementation: What were the barriers encountered in attempting to implement the system? How were these addressed?

Problem resolution: How effectively does the process resolve problems related to implementation? Were problems addressed at the most appropriate system level?

Protocol Adherence: Were the data collection protocols followed rigorously?

Sampling: Were samples drawn according to the sampling plan?

Data collection timeline: Were data collected during the specified time period?

Data collection completion: Were data collected for all of the performance indicators included in the project?

Protection: Were protocols related to privacy, confidentiality and informed consent followed?

Step 4 *How do I 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 *How do I assess the cost-effectiveness and organizational burden of the PIP?*

This is a highly technical area and it is unlikely that many PIP projects will involve a true cost effectiveness analysis. There is little published information about the cost of performance improvement activities, and that which exists focuses on certain components such as conducting surveys and medical record reviews. Nonetheless as part of your own evaluation and improvement process, you should monitor costs as closely as possible and seek to identify and record cost-benefits whenever possible. Additionally you should assess the extent of the burden imposed by the project by means of feedback mechanisms targeting stakeholders.

You should also be prepared for the possibility that the PIP may result in *increased costs*. For example, a project focusing on improving access for some subpopulation will by definition result in increased service utilization and, therefore, expenditures. Nonetheless in such cases you should seek to identify and report any potential cost-offsets such as the use of inappropriate services (such as emergency rooms or inpatient care) or the reduction in risk factors that would otherwise produce long-term costs.

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

- Training and technical assistance
- Quality monitoring
- Data collection
- Data entry/management
- Data analysis/reporting
- Costs associated with the process of planning the project
- Burden placed on various stakeholder groups
- Burden placed on staff/consumers

How do I conduct the PIP impact evaluation?

Step 6 *How should I define the purpose of the impact evaluation?*

The purposes of an impact evaluation should follow from the specific goals of the PIP as articulated in the study question early in the process of planning and development. The impact evaluation will serve to supplement the assessment conducted by the EQRO and therefore should therefore focus on internal concerns and management issues important to the MHP, for example, workforce deployment, service planning, etc.

Step 7 *How should I define impact evaluation questions?*

Again, the specific evaluation questions will follow from the intended impact of the PIP. Some more general questions that might apply to many impact evaluations are:

- Are measurable changes in the system consistent with the PIP goals?
- Do diverse stakeholders find the information generated by the process meaningful and relevant?
- Is the format of reports appropriate for the content and the level of expertise of various audiences?
- Was the PIP process as conducted generally acceptable to all stakeholder groups?

Methods for evaluating a PIP

Table 2 presents a matrix of suggested methods for addressing the questions that might be the focus of an evaluation of a PIP. Two other strategies, not included in the matrix of methods, are case studies based on specific types or aspects of PIPs 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 PIP process. Comparative qualitative or quantitative studies of systems, one with and the other without a PIP process in place, may highlight advantages of implementing a process that would be undetectable without an external comparison.

Table 2. Methods for Evaluating a PIP, by Type of Evaluation Question

Evaluation Question	Focus Groups	Inter-views	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				

Appendix A:
Glossary of Terms

The following glossary is adapted primarily from the Quality Assessment and Performance Improvement (QAPI) Project Completion Report Instructional Guide.

Accuracy: The closeness of a computed, estimated, or measured value to the exact or "true" value. For example, if an individual is weighed three different times, the closeness of the three weights to the individual's real weight is the degree of accuracy in the weight measurements.

Barrier analysis: The identification of the underlying factors, issues, conditions, or situations that contribute to a resistance to change for a specified topic.

Baseline data measurement: The initial data gathering and assessment process that takes place before interventions are instituted. Results are compared with data collected after full intervention implementation to determine whether the interventions have been effective.

Bias: A flaw in a study design such that it systematically favors certain outcomes.

Collaboration: MHPs may combine efforts to achieve common goals of the PIPs. Examples of potential areas of collaboration are:

- Developing quality indicators
- Collecting the baseline data
- Designing and implementing interventions
- Re-measuring for demonstrable improvement
- Implementing subsequent or modified interventions
- Re-measuring 1 year for sustained improvement

Confounding: Factors that differ between comparison groups in observational studies.

Consultation/technical assistance: The process of seeking assistance regarding a specific component of the PIP, distinct from a standard type of comprehensive collaborative project. Technical assistance might include medical record abstraction, abstractor training, data collection and analysis, data and process validation, statewide comparisons/benchmarking, design and development of interventions (graphics and printing), dissemination of intervention materials (mailings), study design development, and training.

Continuity and coordination of care: The manner in which care is provided when a patient receives care from multiple providers and across multiple episodes of care.

Demonstrable improvement: A significant improvement in a quality indicator over time—the second data point for a PIP. Significant in this context does not imply statistical significance.

Denominator: Mathematically, the number below the horizontal line in a fraction, denoting into how many equal parts the unit is divided. In a quality improvement activity, the **denominator represents the set from which the numerator, a sub-set, is derived.**

Descriptive statistics: Statistics that are used to describe and summarize numerical data that do not involve generalizations to a larger set of data. Quantitative data assessments generally require descriptive statistics, which include measures of central tendency (i.e., mean, median, or mode) and measures of variability (i.e., range or standard deviation).

Experimental study: A study that involves assignment of subjects into experimental and control (comparison) groups.

Median: The middle score of a set of scores.

Mean: The arithmetic average of a set of scores.

Mode: The most frequently occurring score in a set of score: Mathematically, this is the number above the horizontal line in a fraction, denoting the number of fractional parts taken. In a quality improvement activity the numerator represents the subset of the denominator for which results being studied were achieved.

Observational Study: A study design that involves post-hoc comparisons of group formed for purposes other than the study itself.

Power analysis: A statistical procedure for: 1) Determining the required sample size needed for adequate sensitivity or precision; and 2) Estimating the likelihood of committing a Type II error. One can increase statistical power (precision, sensitivity) by increasing the size of the effect (a stronger treatment), reducing variability (moving from the field to the laboratory or using more reliable tests), or reducing the standard error (increasing sample size).

Project: An initiative by an organization to measure its own performance in one or more focus areas, undertake system interventions to improve its performance, and follow up on the effectiveness of those interventions.

Proxy measure: A substitute or surrogate measure for an outcome which may be resistant to direct or timely measure (e.g., availability of evidence based practices).

p-value: A statistic that expresses the degree to which a finding is probably related to chance. Its calculation is dependent on the number of subjects.

Performance target: The desired level of achievement that the organization sets for itself as its targeted standard of care.

Qualitative analysis: The systematic summation and interpretation of non-numeric data to reveal meaning and patterns of relationships.

Quality indicator: A clearly defined and objectively measurable aspect of care or service that is expressed as a rate or ratio built from a numerator and denominator. Quality indicator measures compared over time indicate whether changes are leading to improvements.

Quantitative analysis: The systematic statistical manipulation and analysis of numerical data to describe phenomena or the numerical relationships among them.

Random: Unbiased, that is, having no pattern or differing likelihood structure.

Random sample: A sample obtained by chance through random assignment procedures (randomization). Simple random sampling requires that each unit of the population has an equal chance of being selected.

Reliability: The extent of a measure's consistency, repeatability, and reproducibility.

Root cause analysis: A retrospective approach to error analysis, widely applied to investigate major industrial accidents, and mandated by the Joint Commission on the Accreditation of Healthcare Organizations (JCAHO) for investigation of sentinel events in accredited hospitals.

Standard deviation: A measure of the variability that indicates, approximately, the average variation from the mean value for all values in the data set; technically, the square root of the variance.

Statistical inference: The use of data to draw conclusions about some wider population. The reasoning of inference relies on the laws of probability.

Statistical significance: The probabilistic determination that the results obtained in the analysis of the data from the sample are unlikely to be due to chance according to the preset level of risk.

Stratified random sample: A procedure through which a random sample is selected in proportion to each independent stratum or subgroup in the population. The subgroups represent characteristics important to the study.

Sustained improvement: A significant improvement in a quality indicator maintained over time—the third data point for a PIP. Significant in this context does not imply statistical significance.

Target audience: The group for whom a behavioral change is desired (e.g., a physician or physician group targeted to change the manner of prescribing therapeutic doses of anti-depressant medication).

Target population: The clearly defined and objectively measurable population of the eligible Medicare membership to be affected by the quality indicator. For example, anti-depressant medication prescription, the target population consists of all eligible MHP members continuously enrolled and having a diagnosis of major depression during the dates of the study).

Type I error: The error of indicating that a relationship exists between variables when it does not.

Type II error: The error of indicating that a relationship does not exist between variables when it does.

Validity: The extent to which something measures what it is intended to measure. The validity of a study is a composite of: 1) the extent to which the data collected for the study accurately measure the characteristic or property they are intended to measure (internal validity), and 2) how sound and justifiable are the inferences made from study data (external validity).

Variable: A property or quantity that is recorded for each observational unit, i.e., measured characteristic of participants in a study, such as age, sex, diagnosis.

Other Glossaries:

Other more specialized but nonetheless useful glossaries for statistics and health care research may be found at the following websites:

Joint Commission on Accreditation of Healthcare Organizations: Glossary of Terms for Performance Measurement

http://www.jcaho.org/pms/reference+materials/glossary.htm

UK National Screening Programmes (for projects involving health care screening)

http://www.nsc.nhs.uk/glossary/glossary_main.htm#glossary

Center for the Advancement of Clinical Research,

Clinical Research and Good Clinical Practice Dictionary (especially for Federal regulations of clinical trials)

http://www.med.umich.edu/cacr/dictionary/A-B.htm#a

Informedhealth online: Dictionary of Research Terms for Consumers

http://www.informedhealthonline.org//item.aspx/item.aspx?tabid=15

Statistics:

Valerie J. Eastman and John H McColl at
http://www.stats.gla.ac.uk/steps/glossary/

Appendix B:

Crossing the Quality Chasm and the New Freedom Commission: New concepts of quality measurement?

The Donabedian Taxonomy: Structure Process and Outcomes

The classic framework for conceptualizing health care quality was developed in the early 1960's by Avedis Donabedian, M.D. Donabedian's seminal contribution to the field was to formulate the dimensions of health care quality as consisting of *structure, process and outcomes.*

Some have suggested that Donabedian's classification of quality measures as addressing these three domains be supplemented, or even supplanted entirely, by two potential paradigms for quality, one in health care generally and the other specific to behavioral health care. These are, respectively, the report from the Institute of Medicine in 2002 entitled *Crossing the Quality Chasm* (available at http://www.nap.edu/openbook/0309072808/html/) and the report of the President's New Freedom Commission on Mental Health in the following year, entitled *Achieving the Promise: Transforming Mental Health Care in Americ*a (available at www.mentalhealthcommission.gov).

Both reports lay out a vision for quality defined in terms of a set of goals and principles, and it is these, some have suggested that should provide the conceptual basis of quality measurement in place of the Donabedian taxonomy.

The Concept of Quality in the Technical Assistance Manual

Despite the intriguing possibilities of a new approach to conceptualizing quality indicators, this Technical Assistance Manual adheres to the conventional Donabedian taxonomy. This decision is based on two primary considerations.

The first is that is still too early to know what this new formulation will be, if in fact there is to be one. The IOM and the New Freedom Commission reports are both intended as the initial stage of laying the ground work for an extensive process of defining the means to the aims and goals identified. This work is already underway in both cases, especially in the case of the earlier IOM report, which has resulted in a number of subsequent reports and other products, and it is not evident at this stage that these efforts aim to establish an alternative conceptual framework from that of Donnabedian's for quality measurement per se. In fact, the documentation of quality indicators being developed for the "National Health Report" to provide baseline measures of national health care quality, refers explicitly and extensively to the structure, process and outcome framework (see http://www.ahrq.gov/qual/nhqr04/himeasures.htm).

In short, it appears that the work developing from the IOM report is directed not at reformulating the essential character of health care quality indicators but rather the goals and infrastructure of measurement, and it is reasonable to assume that activities following from the New Freedom Commission will follow a comparable course. (For a report of the relationship between the two initiatives see: Neal Adams, M.D. and Allen S. Daniels, Ed.D. "From Policy to Service: A Quality

Vision for Behavioral Health Care" American College of Mental Health Administration, available at http://www.acmha.org/).

The second reason why the Technical Assistance Manual retains the conventional framework is that this is the context in which the quality of the PIP indicators will be evaluated by the EQRO. The Protocol for the EQRO evaluation, which the Manual follows closely, discusses the nature and relationship of process and outcomes indicators in considerable detail. We would be doing a disservice to users of the Manual, therefore, to propose approaches that are inconsistent with the standards by which they will be evaluated.

It is expected that the Technical Assistance Manual will be revised and supplemented as needed to reflect policy changes and new developments in the field, including approaches to quality measurement and the nature of quality indicators.

Appendix C:
Desirable Attributes of Performance Measures

Attribute	Definition
1. Importance of Topic Area Addressed by the Measure	
1A. High priority for maximizing the health of persons or populations	The measure addresses a process or outcome that is strategically important in maximizing the health of persons or populations. It addresses an important medical condition as defined by high prevalence, incidence, mortality, or morbidity.
1B. Financially important	The measure addresses a clinical condition or area of health care that requires high expenditures on in-patient or outpatient care. A condition may be financially important if it either has high per-person costs, or if it affects a large number of people.
1C. Demonstrated variation in care and/or potential for improvement	The measure addresses an aspect of health care for which there is a wide variation in care and/or potential for improvement. If the purpose of the measure is internal quality improvement and professional accountability, then wide variation in care across physicians or hospitals is not necessary. However, if the purpose of the measure is to make comparisons of health care systems, then potentially wide variation in care should exist.
2. Usefulness in Improving Patient Outcomes	
2A. Based on established clinical recommendations	For process measures, there is good evidence that the process improves health outcomes. For outcome measures, there is good evidence that there are processes or actions that providers can take to improve the outcome.
2B. Potentially actionable by user	The measure addresses an area of health care that potentially is under the control of the physician health care organization or health care system that it assesses.
2C. Meaningful and interpretable to user	The results of the measure are reportable in a manner interpretable and meaningful to the intended user. For example, physicians must be able to use the information generated by the measure to improve patient care. Health care organizations must find the information useful for decisionmaking purposes. When measures are used to compare health care systems, users should be able to understand the clinical and economic significance of differences in how well systems perform on the measure.
3. Measure design	
3A. Well defined specifications	The following aspects of the measure are to be well defined: numerator, denominator, sampling methodology, data sources, allowable values and method of reporting.
3B. Documented reliability	The measure will produce the same results when repeated in the same population and setting (low random error). Tests of reliability include (a) test-retest reproducibility: test-retest reliability is evaluated by repeating administration of the measure and calculating agreement among the repetitions; (b) inter-rater: agreement between raters is measured and reported using the kappa statistic; (c) data accuracy: data are audited for accuracy; and (d) internal consistency for multi-item measures: analyses are performed to ensure that items are internally consistent.
3C. Documented validity	The measure has face validity–it should appear to a reasonable observer to measure what is intended. The measure also should correlate well with other measures or the same aspects of care (construct validity) and capture meaningful aspects of this care (content validity).
3D. Allowance for risk	The degree to which data collected on the measure is risk adjusted or risk stratified depends on the purpose of the measure. If the purpose of the measure is for internal continuous quality improvement and professional accountability, then requirements for risk adjustment or risk stratification are not stringent. If the purpose of the measure is comparison and accountability, then either the measure should not be appreciably affected by any variables that are beyond the user's control (covariates), or any extraneous factors should be known and measurable. If case-mix and/or risk adjustment is required, there should be well-described methods for either controlling through risk stratification or for using validated models for calculating an adjusted result that corrects for the effects of covariates. (In some cases, risk stratification may be preferable to risk adjustment because it will identify quality issues of importance to different subgroups).

SOURCE: The Performance Measurement Coordinating Council (PMCC) was established in 1998 as a collaborative venture among the American Medical Association (AMA), the Joint Commission on the Accreditation of Healthcare Organizations (JCAHO), and the National Committee for Quality Assurance (NCQA). The PMCC Desirable Attributes of Performance Measure is a consensus document developed jointly by the three sponsor organizations, AMA, JCAHO, and NCQA, in April 1999.

Appendix D:

What kind of consultant do I need?

Not all PIP projects will require the input of technical expertise. The manual identifies several areas of more complex methodology where such consultation might be needed, however. The following table identifies a range of these areas and the type consultation appropriate to each.

Illustrative List of Areas of Consultation and Experts for PIPs

(from (McGlynn, Damberg et al. 1998))

Potential Consultation Area	Type of Expert
Defining stakeholder questions	Focus group facilitator
Deciding on data elements for issues related to:	
Costs of health care	Economist
Access to health care	Health services researcher
Quality of care	Physician
Organization of services	Organizational behaviorist
Patient satisfaction	Psychometrician
Insurance coverage	Economist
Developing clinical data bases	Physician Biostatistician Health services researcher Medical records technician Nurse
Conducting surveys	
Survey design	Survey research specialist Psychometrician
Sampling	Statistician
Fielding surveys	Survey research specialist
Analysis	Health services researcher Statistician Economist Clinician Computer programmer

Appendix E:

Miscellaneous Resources for Conducting PIPs.

Published Quality Improvement Case Studies:

General Health Care:

NCQA Quality Profiles

http://www.qualityprofiles.org/quality_profiles/case_studies/

Behavioral Health Care

Duckworth K, Hanson A: Managed Care: Using a Clinical and Evidence-Based Strategy to Preserve Access to Psychiatric Medications. Psychiatric Services 53:1231-1232, 2002.

Describes a state-wide initiative conducted jointly by Massachusetts Department of Mental Health and Division of Medical Assistance (Medicaid) to control costs and preserve access to psychiatric medications for persons with serious mental illness.

Dickey B, Normand S-L, Hermann R, et al.: Guideline recommendations for treatment of schizophrenia: The impact of managed care. Archives of General Psychiatry 60:340-348, 2003

A research study that demonstrates how the Schizophrenia Port guideline may be used to measure quality of care for persons with this diagnosis. This case also demonstrates the similarities and differences between a research project and a PIP. The purpose of the study was to develop new knowledge about the quality of managed care vs. fee for service systems. A PIP might follow the same protocol, including comparisons of provider groups, to assess the quality of pharmacotherapy for schizophrenia.

Selecting and Developing Indicators (Measures)

Selecting Process Measures for Quality Improvement in Mental Healthcare.
Richard C. Hermann, MD, MS, H. Stephen Leff, Ph.D, and Greta Lagodmos, BA.

This toolkit is designed to help healthcare organizations to identify and select measures meeting their needs for use in quality assessment and improvement activities. It builds on the National Inventory of Mental Health Quality Measures, an interactive database found at http://www.cqaimh.org/quality.html. Both the inventory and the toolkit focus on process measures, a type of quality measure that evaluates components of the interaction between patients and the healthcare system. This toolkit will describe characteristics of process measures, how to select measures meeting an organization's needs, and their use in conjunction with other types of quality measurement.

Case-Mix Adjustment:

A Toolkit for Conducting Case Mix Adjustment of Mental Health Performance Indicators. Michael Hendryx, PhD. The Evaluation Center@HSRI May 2004

The purposes of the tool-kit are, first, to provide a description of the definition, rationale, limitations, required tasks, and analytic methods of mental health performance indicator risk adjustment; and second, to provide the reader with computer exercises using a hypothetical database to practice conducting a risk-adjustment using either SAS or SPSS. A training module appropriate for distance learning adaptability will be released to accompany the toolkit..

For information on ordering hard copy and electronic versions go to www.tecathsri.org.

Quality Improvement Tools

NCQA Quality Profiles. The NCQA website provides examples of numerous templates for QI tools used by member organizations such as patient education brochures, focus group questions and provider information letters. These are oriented toward the HEDIS measures used for NCQA reporting, and include only a few examples specific to behavioral health, but many are easily adapted to that purpose. The templates may be accessed at http://www.qualityprofiles.org/qia_Tools/index.asp.

NCQA Quality Improvement Activity form, along with detailed instructions containing numerous examples relating to general health care is available at http://www.qualityprofiles.org/qia_Form/index.asp.

AHRQ Child Health Care Quality Toolbox is an online, interactive source of widely applicable, hands-on information about performance measurement and quality improvement, available at: http://www.ahcpr.gov/chtoolbx

Substance Abuse and Mental Health Services Administration. Keys to Quality: Conducting a Performance Improvement Project for Behavioral Health in Managed Care Based on the Principals of QISMC. October 2000, Produced by The Evaluation Center@HSRI, www.tecathsri.org.

Statistics online

Textbooks:

Hyperstat Online Textbook http://davidmlane.com/hyperstat/index.html

Surfstat.*australia*: an online text in introductory statistics,University of Newcastle, Australia. http://www.anu.edu.au/nceph/surfstat/surfstat-home/surfstat.html (also research methods

Statsoft Electronic Statistics Textbook http://www.statsoft.com/textbook/stathome.html

Rice Virtual Lab in Statistics: http://www.ruf.rice.edu/%7Elane/rvls.html

Calculators:

Sample size calculator: Creative Research Systems, Inc. at http://www.surveysystem.com/sscalc.htm.

Assign subjects to research groups
http://graphpad.com/quickcalcs/index.cfm

Randomly select a subset of subjects
http://graphpad.com/quickcalcs/randomselect2.cfm

Calculate a confidence interval, knowing mean, sd and n
http://graphpad.com/quickcalcs/CImean2.cfm

Research methods:

Web Center for Social Research Methods (Cornell)

http://www.socialresearchmethods.net/

Cochrane Effective Practice and Organisation of Care Group (a component of the Cochrane Collaboration) Database of systematic reviews of health care quality improvement strategies
http://www.epoc.uottawa.ca/reviews.htm

Implications of HIPAA for research:
http://privacyruleandresearch.nih.gov/

References

Buchanan RW, K. J., Zito JM, Lehman A (2002). "The schizophrenia PORT pharmacological treatment recommendations: conformance and implications for symptoms and functional outcome." Schizophrenia Bulletin **28**(1): 63-73.

Hendryx, M. (2004). A Toolkit for Conducting Case Mix Adjustment of Mental Health Performance Indicators. Cambridge, MA, The Evaluation Center@HSRI.

Hermann, R. C. and S. Provost (2003). "Best Practices: Interpreting Measurement Data for Quality Improvement: Standards, Means, Norms, and Benchmarks." Psychiatric Services **54**(5): 655-657.

Institute of Medicine (2001). Crossing the Quality Chasm: A New Health System for the 21st Century. Washington DC, National Academy Press.

McGlynn, E., C. Damberg, et al. (1998). Health Information Systems: Design Issues and Analytic Applications, RAND Health.

Perneger, T. V. (2004). "Why we need ethical oversight of quality improvement projects." Int J Qual Health Care **16**(5): 343-344.

Shojania, K., K. McDonald, et al. (2004). Chapter 2. Evidence-based Review Methodology for the Closing the Quality Gap Series. Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies Volume 1—Series Overview and Methodology. Rockville, MD, Prepared by: Stanford–UCSF Evidence-based Practice Center for Agency for Healthcare Research and Quality U.S. Department of Health and Human Services.