

**Recommendations for design, development and implementation of a clinical information system to measure and manage performance in Washington's public mental health system.**

**Prepared for the Joint Legislative Audit and Review Committee by the Center for Clinical Informatics, LLC.**

**June 23, 2000.**

# Contents

---

<b>Overview</b>	<b>2-3</b>
<b>Ten Guiding Principles</b>	<b>4-10</b>
<b>Access to Care</b>	<b>11-17</b>
<b>Appropriateness of Care</b>	<b>18-21</b>
<b>Outcomes</b>	<b>22-39</b>
<b>Structure/Plan Management</b>	<b>40-41</b>
<b>Implementation</b>	<b>42-47</b>
<b>Cost Estimate</b>	<b>48-50</b>

## Overview

---

The purpose of this document is to propose a conceptual framework for developing a practical performance management system designed to improve outcomes while leading to a more rational and ethical allocation of care, focusing limited resources on those most in need. The primary indicators for the system are proposed in the Summary of Recommended Performance Measures for Washington's Public Mental Health System prepared by Clegg & Associates, University of Washington Health Policy and Analysis Program and the Center for Clinical Informatics.

The framework consists of three levels of criteria, moving from the abstract to the specific.

1. **Guiding Principles:** These criteria consists of rules that could be applied to the design any performance management system intended for a large, complex enterprise with operations at multiple geographically dispersed locations. These principles are not necessarily specific to behavioral healthcare systems.
2. **Requirements:** These are the criteria for the specific indicators that the performance management system must include. These criteria are specific to behavioral healthcare systems, are structured around the NASMHPD domains of indicators.
3. **Recommended Specifications:** The specifications will suggest how the requirements can be fulfilled. The specifications cover both the data elements and the reporting requirements for effective management of the indicator.

The recommended core data set presented in the specifications is for statewide use. Additional data elements may be added at the program or regional level to address specific needs, but these should not supplant the core data set.

Reporting specifications are broken out by the following organizational levels.

1. Individual clinician
2. Facility or clinic
3. Regional Support Network
4. MHD
5. State Legislature

For each level, the recommended report specifications will include the content and frequency of the report. A Rationale and Discussion section will provide the reasoning behind the specifications.

This document is intended as a starting point for system design. By providing a logical framework with specificity in the data set and reporting requirements, the authors hope to provide a reference point for further informed dialog among stakeholders in the design of the performance management system.

The concluding section of this document proposes a plan to implement the system. An estimate for the development and deployment of the necessary information system is included.

Total cost for design, development and implementation are estimated to be less than \$900,000 spread over a two-year period.

## Ten Guiding Principles

---

A core set of premises guide the rationale for the system criteria. The principles are distilled from the survey of best practices and the authors' own experience designing such systems. Each element of the system will reflect these principles.

1. Measure to manage.
2. Management requires frequent feedback over time.
3. Keep it simple, make it matter.
4. Keep it brief, measure often.
5. Create benchmarks, compare results.
6. Minimize opportunity for feedback induced bias.
7. Provide the right information at the right time to the right person to make a difference.
8. Build in the flexibility so that the system evolves with the experience of the users.
9. Maintain central control of data and reporting.
10. Establish and protect a core data set.

### Principle 1: Measure to manage

If it's worth measuring it's worth managing. While this may seem self evident, in practice it's easier to design a system to measure than to manage. Measurement can consist of a snap shot at a moment in time, management requires tracking changes over time.

The following hypothetical scenario will illustrate this point. Let's assume that a state legislature is interested in consumer's satisfaction with mental health services provided through publicly funded clinics, and asks the state's Mental Health Division to provide a report on consumer satisfaction. The Division is pleased that the legislators are interested, and informs them that they are conducting a state of the art study on consumer satisfaction using the full length (40 item) MHSIP consumer survey utilized by many other states.

In order to assure that the data is as unbiased and objective as possible, the state contracts with an external vendor to administer the survey in a scientifically rigorous manner. To assure candor, neither the identity of the consumer or the provider is a part of the data. The survey is administered to a representative random sample of consumers, and every effort is made to assure return of the completed survey. After all the data is collected, the Division provides the legislature a report summarizing the results of the survey. The legislature now has a scientifically valid snapshot of consumer satisfaction at one point in time.

This is good measurement, but it's not management. Let's assume that the legislature decides that there are areas that need improvement, and so ask for a

report in one year's time showing that satisfaction has increased. Now what does the Division do? There is no data on the source of variation in satisfaction, no way to identify success stories or problem providers. The Division has no way to target an intervention for the problem provider, or to monitor if the intervention successful. In this scenario, even if the state conducts a second consumer survey a year later, it would be difficult to demonstrate that any change in satisfaction that might occur was the result of the Divisions management.

Let's look at an alternative scenario, this time designing into the study the capability to manage results. The Division realizes if it wishes to improve satisfaction, it needs to understand the sources of variation in satisfaction. Which providers are doing well. What can be learned from them? How can this knowledge be transferred and reinforced for providers that need improvement?

The Division decides to get a sense of the base rate of satisfaction by asking all consumers seen in a single week to complete the satisfaction survey. Since the consumer is being asked to complete the form at the time of service, the Division decides to utilize the briefer 20 item version of the survey. Clinic reception staff are asked to give the survey to all consumers seen during a given week, asking the consumer to cooperate because it will help the clinic to know how to better serve them.

The state realizes that asking staff to give out the survey may introduce bias into the results, but decides to go with it because the data can be collected quickly and economically. Also, by involving the clinic staff in the data collection, it sends a clear message to all that the MHD is interested in knowing that the services are meeting the needs of the consumers.

After the first wave of data collection, the data is analyzed to determine base rates of satisfaction and the variability across sites. Not surprisingly, the state finds that some providers are in fact much better at satisfying consumers than others.

In order to simplify the report and increase the statistical power of the measure, the Division combines the items to form two scales, one measuring global satisfaction and the other the perception of benefit from treatment. Now rather than measuring the percentage of consumers indicating satisfaction on a specific item, the Division has scales that are a continuous variable with known psychometric properties, including mean, standard deviation and internal consistency.

The Division prepares a report showing the average scores across the entire system. Each individual site is also given their own results and their percentile rank compared to all other sites. Clinics that are at the 85<sup>th</sup> percentile or high are congratulated, and a representative from the Division visits to try and understand what they are doing right. Clinics in the bottom 15% are reassured that they will

not be penalized for these results, but that the Division is interested in helping them improve.

In order to provide performance feedback and monitor improvement, the Division requests that the clinics administer the survey during the last week of each month for the next three months. Results are fed back to the clinic each month, and a regular teleconference is scheduled to go over the results. Clinic directors are urged to share results with their staff.

Gratifyingly, results begin to improve, particularly for at those sites that had the poorest results. They of course have the most room for improvement. Over a period of a few months, the variability in satisfaction decreases and the average satisfaction scores for the entire system has increased. As a reward for their good work, the Division permits the clinics to drop back to a quarterly rather than monthly administration of the survey.

At the next funding cycle the Division is able to demonstrate to the Legislature that not only did they measured satisfaction, but they succeeded in managing for better results. Granted, this is a simplistic example, but it serves to illustrate the differences between simple measurement of performance and active management to achieve improved performance.

## **Principle 2: Management requires frequent feedback over time**

The scenario illustrates this principle. It isn't sufficient to provide feedback annually. In order to be effective the feedback must be frequent enough that those who can make a difference can act on the information in a timely manner.

## **Principle 3: Keep it simple, make it matter.**

The point is to choose carefully what to measure but then make it important to the people who can make a difference. Too much information distracts the use from acting on the important information.

In the scenario, satisfaction is measured as global construct, using a single scale. Certainly the individual items could be analyzed to look at the components of satisfaction, but asking clinic staff to pay attention to results on multiple separate items rather than a single global measure makes the feedback needlessly complex. A researcher or clinic director may be interested in the results at the item level, but for the purpose of measuring and improving satisfaction, the total score is more important than any single item. A single satisfaction score is comprehensible and change over time can be plotted on a graph.

#### **Principle 4: Keep it brief, measure often.**

The more frequent the measurement, the greater the opportunity to manage. But frequent measurement is time consuming. If the collection of the data becomes too burdensome, the result will be poor compliance and missing data. A shorter measure with complete data is often more valuable than a longer measure with missing data. Brief measures administered frequently are better management tools than long measures administered infrequently. Consider the “ten minute rule”: Consumers and clinicians should be able to complete outcome assessment questionnaires in the time between routine outpatient appointments.

#### **Principle 5: Create benchmarks, compare results**

Measurement of an indicator tells us little, unless there is something to compare it too. Of course, the comparison has to be fair as possible. This is where the concepts of case mix adjustment come into play. However, measures do not have to be perfectly fair to be useful. In the satisfaction scenario, there may well have been unknown variables other than the quality of care contributing to low satisfaction. Nevertheless, the fact that results from each clinic were compared to the results of the other clinics at least provides some context to evaluate the results while providing a natural motivation (competition) to improve. In this case, the fact that satisfaction improved relative to the baseline is more important than knowing at all variables that affect satisfaction are accounted for.

#### **Principle 6: Minimize opportunity for feedback induced bias.**

Whenever possible, an indicator should be independent of the individuals whose performance may be evaluated. This is particularly problematic for clinician rated measures of outcome. Knowledge that clinical improvement is a performance indicator introduces incentives to bias the ratings, however unconsciously. This is particularly true if comparative results are fed back to the providers.

Consumer self-report measures are much less susceptible to this effect. While a provider’s bias in completing a measure contaminates all cases rated by the provider, patients’ idiosyncratic response sets are likely to be randomly distributed across providers.

#### **Principle 7: Provide the right information at the right time to the right person to make a difference.**

The system should provide information to help people do their job more effectively. This means getting them the right information at the time they need it for the task at hand.

In the satisfaction scenario, the Legislature is the highest level of authority. In order to do their job, legislators need to know that the system is effective and improving. They probably only need to know this once every year or two years, depending on the funding cycle. They want to see evidence of good management and quality improvement. In the case of satisfaction, it may be sufficient to know that the global measure of satisfaction rose 20% since its measurement began, that it remains stable, and that variability in satisfaction across the system is being managed effectively.

The hypothetical Director of the MHD needs at a minimum quarterly reports on global satisfaction, and enough information to know which clinics are performing well and which need improvement. The Clinic Director needs more detail about satisfaction at his or her site to manage effectively. What is the source of satisfaction and dissatisfaction? Do consumers feel like they are being treated courteously and with respect by the office staff? Do they feel like the clinical staff understands their concerns? If satisfaction is below average, she or he needs the information monthly in order to act in timely manner if results are to improve on the next quarter's report.

And the office manager may only need to pay attention to the single item about courtesy of the office staff. In each instance, the individual that is in the best position to make a difference gets just the information they need to do their job effectively and they get it in time to do something about it.

### **Principle 8: Create flexibility so that the system evolves with the experience of the users.**

This may be the most important design principle of all. If the information provided by the system matters to the users, then the users will ask for changes in how that information is provided. If the information is not available or not being collected, then the system will need to adapt accordingly. Ideally, the system should evolve over time through a process of trial, feedback and enhancement.

Suppose, for example, that the Division decided to invest in software to help automate the process of capturing and reporting data from the satisfaction survey. As it happens, a vendor has developed software just for that particular survey and is willing to license it a very attractive price.

What might appear to be an easy decision may have a hidden cost. What if a year later the Division wishes to add additional items to the basic survey? Or the users request a change in the reports to make them easier to understand. Or what if an alternative survey with much better benchmarking data becomes available. Can the system be modified to accommodate these changes, or does the system itself begin to dictate how things are measured. A system that can be easily enhanced over time in response to the experience of the users stands a much better chance of providing long-term value, even if the initial investment may appear greater.

## **Principle 9. Maintain central control of data and reporting.**

One of the greatest challenges to operating a viable performance management system is to insure the integrity and comparability of data. This is best achieved by maintaining the data warehouse and reporting functions centrally under the highest level of administrative authority responsible for performance of the system.

Centralizing data management and reporting functions also has the advantage of creating economies of scale and expertise. From a data processing and staff productivity point of view, analysis and reporting for an entire state system need not be substantially more work than for a single region.

## **Principle 10. Establish and protect a core data set.**

The choice of the data set is critical to the usefulness of the performance management system. Data elements and measures should be consistent across the system of care. There is a natural tendency for programs or regions to want to choose their own measures. This creates enormous challenges in looking at comparability of performance across the system. Once a data set is established, any changes should be the result of systematic analysis and evaluation, not simply in response to the wishes of the users.

The following sections will provide the requirements and recommended specifications create a clinical information system for performance and outcomes management. The term clinical information system (CIS) is utilized to convey the integration of clinical and quality data with other financial and administrative indicators of system performance.

Of course these specifications will in all likelihood be modified the course of actually developing a system, but our hope is that the concrete nature of these recommendations will provide a starting point for thinking about the ultimate design of the system.

## Access to Care

---

Access to care is assessed from three perspectives:

1. Utilization
2. Consumer perception and satisfaction with access.
3. Responsiveness of system to new requests for service

The following sections provide specifications for the data elements and reports for each of these perspectives.

### Utilization

#### Requirements

1. Provide accurate and timely information on number of persons served on who receives services (penetration)
2. Provide information on number and type of services provided (utilization)

#### Recommended Specifications

##### *Data set*

- In order to simplify reporting and comparisons across sites or regions, the myriad of services should be classified into a limited number of levels of care such as:
  1. Routine outpatient.
  2. Intensive outpatient (more than 2 sessions per week on a regular basis).
  3. Day treatment/Partial hospital
  4. Residential
  5. Inpatient
- Services provided must be coded according to a standardized set of definitions.
- The coding system must include designations for admission or intake assessments as opposed to ongoing and recurring services.
- Data set must include demographics of race/ethnicity, age and gender
- State wide census data

## ***Reporting***

### *Individual clinician; outpatient sites only (monthly, quarterly, and annually)*

- Number of intakes during reporting period
- Average number of sessions per case
- Break out results by age group.
- Compare results to averages for other clinicians.

### *Program, facility or clinic*

- Number of admissions and average length of treatment , broken out by age group and race/ethnicity. *(monthly, quarterly, and annually)*
- Provide comparison to similar sites *(quarterly and annually)*

### *RSN*

- Copies of utilization reports for sites. *(monthly, quarterly, and annually)*
- Provide comparisons of similar sites within RSN *(quarterly and annually)*
- Provide break out of Medicaid/non Medicaid consumers. For Medicaid population, report admissions and encounters or inpatient bed days per thousand members annually for each level of care. For non medicaid consumers, report admissions encounters, bed days, etc per 100,000 lives in the general population, as determined by census data. Break out results by age group and race/ethnicity. *(quarterly and annually)*

### *MHD (quarterly and annually)*

- Report admissions, length of treatment, and days or visits/1000 members or 100,000 general population for each level of care, aggregated at the RSN and state level. Break out results by age group and race/ethnicity.
- Provide break out of Medicaid/non Medicaid consumers. For Medicaid population, report admissions and encounters or inpatient bed days per thousand members annually for each level of care. For non-Medicaid consumers, report admissions encounters, bed days, etc per 100,000 lives in the general population.
- Provide graphical comparisons of admissions and length of treatment across RSNs.
- Provide detailed sub reports breaking results out by age group, race/ethnicity, and level of care.

### *Legislature (annually)*

- Report admissions, length of treatment, (days or visits) for each level of care, aggregated at the state level.
- Provide detailed sub reports to break out results by age group and race/ethnicity.
- Compare race/ethnic utilization to state wide census data.
- Provide comparisons to utilization to other state systems

## **Rationale and Discussion**

Consistent and reliable utilization data is absolutely essential to effective financial management of the delivery system. It also provides the necessary information to manage the allocation of services. Who receives the services? How much do they receive? Does this reflect the priorities and mission of the of the publicly funded mental health system?

Reducing the variety of services down to four or five “levels of care” simplifies reporting at a system level. It also makes it easier to compare utilization across sites or RSNs.

More detailed information on services provided may be useful for management at the local level. For example, a clinic administrator will probably want to know how many sessions are provided by the medical staff versus non medical staff because of the difference in cost. This level of detail is not necessary for managing utilization at the macro level.

Whenever feasible, comparisons across sites or RSNs should be presented in a simple graphic manner rather than as tables of numbers. While tables may provide supporting documentation and detail, they are difficult for the non-statistically inclined to digest quickly. Graphics communicate the information most clearly and efficiently.

In order to permit comparisons across RSNs, the measure of utilization must be expressed as the number of units over a uniform denominator. In the case of a defined health plan population such as the Medicaid population, utilization is reported as units of service per 1000 members annually. When plan membership is not available, an alternative is to use the total state population as the denominator, expressed as units of service per 100,000 population annually. When combining Medicaid/non-Medicaid population for summary purposes, use 100,000 population as the denominator.

The number of consumers served and the average number of sessions provided measure productivity at the individual clinician level. The individual clinician should receive this feedback on a monthly basis. Monthly reports for individual staff are important to provide ongoing feedback, but this is not an adequate period of time for a meaningful comparison to other clinicians. For this reason, the quarterly reports provide combined results for the three-month period. The average admissions and length of treatment for all clinicians at similar sites in the system provides a benchmark for comparison purposes. Clinic directors should also receive reports on individual staff as a tool to manage productivity and allocation of staff resources.

Similarly, reporting utilization on a monthly basis to individual facilities and clinics provides information necessary to manage the allocation of services, but comparisons to similar sites occurs only on a quarterly basis. Quarterly reports to the MHD enables the Division to monitor the overall performance of the system and to make comparisons across RSNs. An annual report summarizing 12 months worth of data permits tracking of trends from year to year.

This cycle of monthly, quarterly and annual reports is utilized throughout this document for many of the indicators at the clinic/facility or RSN level. A quarterly and annual reporting cycle is specified for the other indicators at the RSN level and all indicators at the MHD level. Legislative reports are always annual.

As part of an annual report, the MHD should seek out results from others states to serve as a basis for benchmarking the utilization within Washington. This may prove challenging at first, but with time and effort collaborative arrangements between states should facilitate this exchange of information.

## Consumer Perception of Access

### Requirements

1. Assess consumer perception of access to care using a consumer survey.
2. Administer confidentially; no personal identifiers on form.
3. Administer the surveys during a one week period of every month to all consumers receiving ongoing services at the treatment site.
4. Data on satisfaction should be linked to specific facilities, clinics or community providers.
5. The satisfaction measure should be uniform across all RSNs, and comparable to satisfaction measures utilized by other states.

### Recommended Specifications

#### Data set

- Utilize a form of the MHSIP consumer survey. Four items are related to access:
  1. The location of the services was convenient.
  2. My caregivers were willing to see me as often as I felt it was necessary.
  3. My calls were returned within 24 hours.
  4. Services were available at times that were good for me.
- Capture site ID with MHSIP data
- Combine items to create an Access Satisfaction Scale, expressed as the average response across the four items.

#### Reporting

*Individual clinician*

NA

*Program, facility or clinic*

- Report average Access Satisfaction Scale scores. (*monthly, quarterly and annually*)

*RSN*

- Copies of reports for individual sites. (*monthly, quarterly and annually*)
- Provide graphical comparisons of similar sites within RSN (*quarterly and annually*)

*MHD (quarterly and annually)*

- Copies of reports for individual RSNs
- Provide graphical comparison of RSNs.

*Legislature (annually)*

- Report average Access Satisfaction Scale scores and where possible compare to results from other states using MHSIP survey.

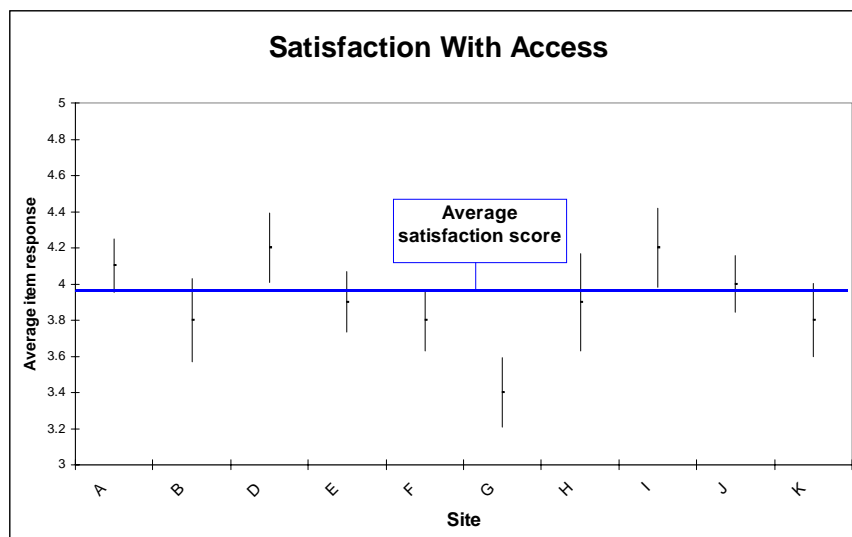
## Rationale and Discussion

The MHSIP survey is recommended because its wide spread use offers the best hope of benchmarking results against those of other states. A short 20-25 item version is adequate to sample the various components of satisfaction, while brief enough for frequent administration. The data set specification calls for capturing the ID of site where the survey is administered so that site specific results can be reported.

The specifications call for monthly administration of the survey. Monthly administration is an aid to local management. Problems with access can be responded to proactively. Monthly administration of the satisfaction survey also underscores the commitment to the consumer to high quality care. The simplest method for administration is to ask all consumers served during a specified period (such as the last week of every month) to complete the short survey. As data cumulates, sites with consistently good results can drop back to a quarterly schedule for assessing satisfaction.

Combining items to form subscales like the Access Satisfaction Scale simplifies reporting and comparison across sites. Using a multi item scale to measure the construct provides a more reliable measure any single item. By calculating the scale score as the average of the items, the results are easily compared to that of other states using the MHSIP survey without the need for identical item content on the scales.

Results should be portrayed graphically. Following is an example of how this might be accomplished. The plots show the average item response with a confidence interval above and below this mean. The confidence intervals convey the degree of uncertainty due to sampling and measurement error. In this case, there is a 90% confidence that the actual results fall within the range of scores covered by the band. One can quickly see that all but one a very close the benchmark, as defined by the average score across all sites.



## Responsiveness

### Requirements

1. Provide information on the speed with which the system responds to consumers seeking care.

### Recommended Specifications

#### *Data set*

- Record data of first contact (usually by telephone) with clinic.
- Capture date of services provided.
- In order to differentiate between the urgency of the need for services, each request for service should be classified as emergency, urgent or routine.

#### *Reporting*

##### *Individual clinician*

NA

##### *Program, facility or clinic (quarterly and annually)*

- Report average days from first contact to first kept session, broken out by three levels of urgency.

##### *RSN (quarterly and annually)*

- Copies of reports for individual sites.
- Provide graphical comparisons of similar sites within RSN

##### *MHD (quarterly and annually)*

- Copies of reports for individual RSNs
- Provide graphical comparison of RSNs

##### *Legislature*

- none

### Rationale and Discussion

The simplest and most universal indicator of responsive to new consumers is “wait time to first appointment”; that is the amount of time between when the consumer first requests services and when they are seen by a provider. Measuring the time to the first kept appointment, as opposed to the first offered appointment provides a more accurate indicator of actual responsiveness. Measuring time to first offered appointment may reward offering appointments at inconvenient times rather than responsive to the consumers needs.

Of course, the time to first appointment should vary with the urgency of care. Therefore, it’s necessary to break out wait time by levels of urgency, the most common being emergency, urgent and routine. Each level needs to be defined with specific criteria in order to assure uniformity in coding urgency.

## Quality/Appropriateness of Care

---

The quality and appropriateness domain is assessed from two perspectives:

1. The consumers perception of the services
2. Indicators of linkages and care coordination within the mental health system between clinics and facilities and linkages outside of the system to other health care providers and the criminal justice system.

### Consumer Perception of Appropriateness

#### Requirements

1. Assess consumer perception of appropriateness using a consumer survey.

#### Recommended Specifications

##### *Data set*

- Utilize short form of the MHSIP consumer survey. Six items are related to appropriateness:
  1. My caregivers believed that I could grow, change and recover.
  2. I felt free to complain.
  3. I was told what side effects to watch for.
  4. My wishes about who is and is not to be given information about my treatment were respected.
  5. My caregivers were sensitive to my cultural/ethnic background.
  6. My caregivers helped me obtain the information needed so I could take charge of managing my illness.
- Include at least one item inquiring about the consumers' perception of active participation in treatment plan.
- Combine items to create an Appropriateness of Care Scale expressed as the average of responses across all items.

## ***Reporting***

### *Individual clinician*

None

### *Program, facility or clinic (monthly, quarterly and annually)*

- Report average Appropriateness of Care Scale scores.
- Report percentage of consumers who actively participated in treatment planning.

### *RSN (quarterly and annually)*

- Provide graphical comparisons of Consumer Perception of Appropriateness Scale and percentage of consumers participating in treatment planning for similar sites within RSN

### *MHD (quarterly and annually)*

- Provide graphical comparison of results across RSNs.

### *Legislature (annually)*

- Report average Appropriateness of Care Scale scores and where possible compare to results from other states using MHSIP survey.

## **Rationale and Discussion**

The rationale for creating a subscale of MHSIP items and the frequency of administration is covered in the previous section. The MHSIP survey does not currently include an item regarding involvement in treatment planning. Such an item might be worded: “I felt like I was actively involved in helping to plan my treatment.”

## Linkages and Coordination

### Requirements

1. Monitor the percentage of consumers in contact with the outpatient provider within 7 days of hospital discharge.
2. Evaluate linkage to physical health services
3. Track the percentage of cases readmitted within 30 days of hospital discharge.
4. Assess responsiveness to the mentally ill in jail/detention.

### Recommended Specifications

#### *Data Set*

- Dates of admissions and discharges from inpatient care
- Dates of outpatient encounters
- Include item regarding linkages to health services on a quarterly clinician completed assessment. (Focus study needed to validate this indicator).
- Data from jails and detention centers, including social security numbers of those admitted

#### *Reporting*

##### *Individual clinician*

None

##### *Program, facility or clinic (quarterly and annually)*

- Report average days between inpatient discharge and outpatient contact.
- Report percentage of cases with linkages to physical health care providers
- Report percentage of mental health consumers arrested or detained that receive services while incarcerated.

##### *RSN (quarterly and annually)*

- Provide graphical comparisons of all indicators for similar sites within RSN

##### *MHD (quarterly and annually)*

- Provide graphical comparison of indicators across RSNs.

##### *Legislature (annually)*

- Report average Appropriateness of Care Scale scores and where possible compare to results from other states using MHSIP survey.

## **Rational and Discussion**

1

The time between hospital discharge and kept outpatient appointment is a critical indicator of the systems ability to coordinate care for the most at risk consumers. In order to achieve a smooth transfer of care without having the patient “fall between the cracks” the inpatient facilities and outpatient sites most coordinate care in a proactive manner.

The system needs to track the discharge dates of individual consumers from facilities and link this to the date they are seen by the outpatient provider. It may prove useful to create specific reports to alert clinicians and care managers of cases being discharged, and to target for review those cases not seen with within 7 days.

The specifications suggest asking the clinician to provide information on linkages to other health care providers as part of the quarter clinical assessment. However, the validity of information obtained in this way is unknown. A focus study can be used to validate a measure. In this case, chart audits in a selected sample of consumers might be used to determine if there is evidence on linkage. This information can then be cross-referenced with the clinician’s response on the quarterly clinical assessment to determine the reliability of the clinician-generated information. The knowledge that a focus study may be used to periodically validate the information also serves to encourage clinician accuracy in reporting.

One of the most challenging indicators will be to determine the percentage of mental health consumers who a jailed or detained that receive mental health services while incarcerated. Ideally, this will be achieved by obtained a file of the social security numbers of everyone added to a jail or detention facility and cross-referencing this with known consumers. If this proves unworkable, the alternative would be to ask the clinician to include this information as part of the quarterly clinical assessment.

## Outcomes of Care

---

The primary function of the mental health delivery system is to assist consumers with mental illness to achieve the highest level of functioning and well being possible given the limitations of resources and current treatment methods. Process and satisfaction indicators owe their significance to the fact that they are thought to contribute to optimal outcomes.

For the vast majority of consumers the expectation is that after some period of treatment their level of well being and functioning is superior to when treatment began. The focus of the outcomes model proposed in this section is to assess the *difference* that mental health services made in the consumer's life. Difference implies change from one point in time to another.

For many patients, the course of their illness is characterized by periods of decompensation requiring intensification of treatment followed by periods of relative stability. Even for these cases characterized by chronicity and instability, there is still a presumption that the intensification of treatment results in near term improvement in the patient's condition.

There is no single measure of treatment outcome that can adequately assess the multifaceted nature the illness and to response to treatment. There is also the question of perspective. Who judges the outcome? The consumer, the clinician, a parent or other societal agents all may have a legitimate interest in and perspective on the outcome.

Then there is the problem of evaluating the outcome. What is reasonable to expect? What is a good outcome? How can results from different locations with arguably different patient characteristics be compared in a manner that is meaningful?

It is little wonder that so much effort has been given to the various process indicators while the actual measurements of outcomes remains largely the province of research studies rather than an integral part of the delivery of care. The complexity of the measurement problem and perceived burden of implementation seem to be overwhelming obstacles.

The following model represents an attempt to overcome these obstacles. A central premise of the model is that the consumers' own experience of benefit as communicated directly by the consumer, is the gold standard by which to judge results.

The proposed framework evaluates outcomes from four perspectives:

1. Consumer retrospective perception of benefit using a standardized survey.
2. Consumer self reported improvement using repeated measures of a standardized outcomes questionnaire administered concurrently with treatment.
3. Clinician assessment of improvement using a standardized rating scale.
4. Indicators of collateral outcomes such as improved employment, maintaining a stable and safe living situation, reduction in involvement in the criminal justice system.

The following sections present the requirements and recommended specifications for each of these four perspectives.

## Consumer retrospective perception of benefit

### Requirements

1. Utilize consumer survey administered at place of service.
2. Exclude from analyses those consumers being seen for intake only.

### Recommended Specifications

#### *Data set*

- Utilize a short version of the MHSIP Consumer Survey.
- Item content should include 7-10 items assessing the direct impact of services, including statements about symptoms, functioning at work/school or other daily activities, quality of familial and social relationships and housing and independent living. At least one item should tap perception of hope for the future.
- Combine items to form an Impact of Services scale.

#### *Reporting*

##### *Individual clinician*

- Reporting at the individual clinician level not possible due to fact that there is no patient or clinician identifier on the questionnaire. Even if there were, consumers often see more than one clinician.

##### *Program, facility or clinic (monthly, quarterly and annually)*

- Report average scores on Impact of Service scale.
- Break out results by age and racial/ethnic groups.

##### *RSN (quarterly and annually)*

- Provide copies of quarterly reports for facilities and clinics.
- Aggregate results at the RSN level.
- Provide graphical comparison of results for similar sites within the RSN.
- Breakout results by level of care, age and racial/ethnic groups.

##### *MHD (quarterly and annually)*

- Report Impact of Service scale aggregated at the state level.
- Break out results by level of care, age and racial/ethnic groups.
- Provide graphical comparison of results on Impact of Services scale across RSNs.

##### *Legislature (annually)*

- Report Impact of Services scale aggregated the state level.
- Break out results by type of program and age and racial/ethnic groups.
- Provide comparisons to results from other states.

## Rationale and Discussion

The most common method for assessing the consumer perception of outcome is the retrospective anonymous survey method. The MHSIP survey (long form) contains 15 items inquiring directly about the impact of services on symptoms in a variety of areas including symptoms, functioning at work and school, living situation and family and social relationships. Using seven to ten of these items will provide a broad indicator of consumers' perspective on the impact of services. Examples:

As a direct result of the services I received:

- I deal more effectively with daily problems.
- I am better able to control my life.
- I am better able to deal with crisis.
- I deal better with people and situations that use to be a problem for me.
- My housing situation has improved.
- I do better in work school.

The MHSIP survey does not contain an item about hope and sense of empowerment for the future. Such an item might read: "As a direct result of the services I received, I feel more optimistic about my future."

There are serious limitations to the survey method of judging outcomes. Because it is anonymous and retrospective, so it provides no basis for monitoring and improving care for the individual patient as the treatment unfolds. Also, the authors own research suggests that retrospective consumer reports of benefit correlate poorly with change as measured using a consumer self report questionnaire at the time of service. It appears that the perception of benefit has more to do with the general well being of the patient at the time of the survey than the actual change that occurred during the treatment period.

Despite the limitations of retrospective surveys as outcome measures, the fact remains that the MHSIP survey is needed to measure other aspects of quality and satisfaction. There is minimal additional cost to capture the perception of outcome items at the same time.

## Concurrent consumer self report

### Requirements

1. Measure change in symptoms, functioning and wellbeing using a reliable and valid consumer self-report questionnaire. The indicator of outcome is the change score, calculated as the difference in test scores from intake to end of treatment episode.
2. Assess change in young children using parent/guardian completed questionnaire.
3. For routine outpatient care, questionnaire is to be administered at intake, at the third session and again at the fifth session, followed by quarterly assessments thereafter.
4. For more intensive levels of care, questionnaires are to be administered minimally at admission and discharge for short stay, crisis stabilization care. For longer term care, administration should occur at regularly spaced intervals (at least monthly), depending upon the expected length of treatment. *Note: Self report questionnaires may not be valid for involuntary patients.*
5. A case mix adjustment model is necessary to benchmark outcomes.
6. Report change scores using the standardized effect size statistic (raw score change divided by the standard deviation of the measure).
7. Monitor ongoing response to treatment, proactively identifying cases at high risk for poor outcome.

### Recommended Specifications

#### *Data set*

- Utilize self report measures of 20-45 items; item content should include statements about symptoms, interpersonal relationships, work school functioning, suicidal ideation, assaultive ideation, and drug/alcohol use.
- Measures should be backed by an adequate data repository to develop valid models for case mix adjustment and projecting trajectory of improvement.
- Utilize a child/adolescent measure that can be completed by either the adolescent or significant adult in the case of younger children.
- Utilize the same core set of items across entire system of care. Additional items can be added to create versions targeted for special populations such as SPMI.

## **Reporting**

### *Individual clinician*

- Provide clinical decision support reports to alert clinician to risk indicators such as suicidal ideation and self reported drug/alcohol abuse (*as needed*)
- Alert clinician and care managers to cases failing to respond to treatment and at risk for premature termination (*as needed*)

### *Program, facility or clinic*

- Provide to outpatient clinics reports listing high risk cases for that site (*weekly*)
- For outpatient care, report outcomes on cases in treatment longer than 90 days. Utilize a moving 12 months window, i.e. cases entering care between 15 and 3 months previously. (*monthly*)
- For facilities, report on discharged cases. (*quarterly and annually*)
- Compare actual change to expected outcome using case mix model at the individual patient level.
- Break out results by age group and severity range. Include average number of sessions for each group.
- Provide graphical comparison to results from similar sites. (*monthly*)
- Provide outpatient clinics with reports on projected trajectory of change for active cases. (*monthly*)

### *RSN (quarterly and annually)*

- Provide copies of monthly outcomes report for facilities and clinics.
- Aggregate outcomes for each level of care. Break out results by age group and severity.
- Provide graphical comparison of outcomes of similar sites within the RSN.

### *MHD (quarterly and annually)*

- Report outcomes aggregated at the state and at the RSN level. Break out results by level of care, severity and age group.
- Provide graphical comparison results across RSNs.

### *Legislature (annually)*

- Report percentage of cases meeting the Reliable change Index criteria for improvement. Break out results by age group and race/ethnicity.

## Rationale and Discussion

Repeated administration of a consumer self-report concurrent with the treatment episode provides one of the most accurate and sensitive methods to measure directly the impact of care. The information obtained from the measures can also be used in treatment decision making, such as determining the optimal intensity and duration of care and assessing response to treatment. Proactive identification of cases at risk for poor response and premature termination provides a powerful tool to improve outcomes for the most at risk consumers. The repeated administration at fixed intervals also permits the modeling of dose benefit ratios. How much treatment is needed? When is it enough?

However, the use of repeated measures is seen as the most burdensome method for assessing outcomes. From another perspective however, the problem comes down to a question of whether or not it is reasonable to expect clinicians to routinely assess their patients response to treatment in a manner that is reliable and valid. The simple fact is that traditionally mental health systems have not felt the need to do this, relying rather on the clinical judgement of the professional. At the risk of over simplification, this is analogous to a physician taking temperature and pulse without the benefit of a thermometer, or stopwatch. Much of the resistance to measuring outcomes is quite simply the reluctance of mental health professionals to relinquish the centrality of clinical judgement to how patients improvement is assessed.

In the proposed outcomes management model, the consumers' self report of improvement is at the heart of the system. The *output* of the mental health system is the *difference* the treatment makes in the consumer's level of wellbeing and functioning. The consumer's own experience is the "gold standard" by which the system is evaluated.

Improvement is measured by the "change score". This is simply the difference between the intake score on the measure and the score at the end of treatment. The change score can be converted to the effect size statistic by dividing the raw score change by the standard deviation of the instrument. In this manner a standardized metric is created that is independent of the instrument being used and permits pooling of data for different instruments. For example, results for children/adolescents using one measure can be combined with results for adults using a different measure, allowing a single outcome score for the combined population.

There are a number of strong reasons for using a consumer self report measure of change:

*High face validity:* A self report measure with items reflecting the most common reasons individual seeks treatment is readily seen as relevant by consumers, clinicians and other stakeholders.

*Cost effectiveness:* The use of consumer self-report measures minimizes the need for clinician completed measures.

*Relatively immune to systematic bias as a result of measurement:* Clinician rated measures are susceptible to rater bias, particularly when used to evaluate performance. While consumers also may have idiosyncratic ways of responding to the questionnaires, these individual differences are presumed to be randomly distributed so as not to systematically bias the results or any particular site.

*Sensitive to change:* Self-report measures tend to be sensitive to change over time, with many consumers reporting rapid improvement in early phases of treatment.

*Sensitive to negative effects:* A relatively small percentage of cases become significantly more distressed during treatment and are at higher risk for premature termination. The authors' analyses of large outpatient samples indicates that clinician completed measures are not as likely to show negative effects for these treatment failures.

*Risk indicators:* Specific items about suicide or drug/alcohol use can be used to identify high-risk patients.

*Case mix adjustment:* The intake score on self-report measures is predictive of the change score. The authors' analyses of large data sets consistently yield a linear relationship between intake score and amount of change. The higher the level of distress, the greater the probability of improvement with treatment. This characteristic of self-report measures enables the use of regression formulas to predict the improvement for each individual case. Other variables can be factored into the model, but the severity at intake accounts for the greatest percentage of variance in the change score.

Of course, there are also significant limitations to use of self-report measures. Patients in treatment on an involuntary basis are more likely to minimize problems, thus invalidating the measure. Also, patients with severely impaired functioning may not be able to complete the measures in a reliable manner. In these cases, outcomes must be assessed using clinician measures.

The model presumes frequent assessment early in treatment for all cases including routine outpatient care. On the surface this may appear burdensome. Applying the "ten minute rule" can minimize this; i.e. routine clinical assessment and outcomes measurement instruments should be brief enough to complete in the time between outpatient appointments. Much longer than that and the process becomes disruptive to the service setting, requiring changes in scheduling of patients and allocation of clinician time. The more burdensome the instrument, the greater will be the resistance to implementation.

The proposed model is intended to provide meaningful information on outcomes at the system level and at the level of the individual patient. The model does not presume that the measures utilized represent a comprehensive clinical assessment adequate for

diagnostic or treatment planning purposes. Rather, it provides a good enough measure to systematically evaluate the effectiveness of a system of care while providing the clinician with information that may augment, but not replace, clinical judgement.

A case can be made that the primary mission of the mental health system is to serve the seriously mentally ill and that outcome measurement effort should focus on this group alone. Furthermore, the average cost of an outpatient episode of treatment is relatively modest compared to the ongoing cost of treatment the SPMI population. One could argue that it is simply not worth the cost and effort to measure outcomes for none SPMI consumers.

We do believe that a strong case can be made to the contrary, that measuring outcomes on routine outpatient care is cost effective and adds value to the services. Our rationale includes both financial and organizational considerations.

From the point of view of the organization, incorporation of outcome assessments into the delivery of care requires a substantial shift in how most clinicians have done their jobs. Understandably, resistance is inevitable. It's difficult to be against assessing outcomes per se, so the resistance will manifest itself in other ways such as poor compliance, objections to the measures used, questions about the validity of the findings, etc. While valid points will be raised, there is always the risk that the objections will paralyze attempts to measure outcomes. The decision to measure outcomes on all consumers underscores the commitment of the leadership to see this change occur. Measuring outcomes only some of the time on some of the patients may tend to undermine this message.

There are also financial reasons to measure outcomes. The authors' analyses of very large outpatient samples have consistently found that patients higher levels of distress show greater improvement in treatment than those with less serious problems. The rate of improvement is also greater for more symptomatic patients. By focusing on change as the outcome criteria, the system encourages clinicians to direct limited resources to the most symptomatic consumers because these are the cases most likely to show substantial improvement.

Despite this pattern of greater benefit to more distressed patients, the length of treatment for the least symptomatic varies little from that of the most distressed consumers. There is also tendency to keep the highest functioning patients in treatment beyond the point at which there is demonstrable added benefit, while more distressed patients are leaving care prematurely despite the fact that they are likely to experience continued improvement with additional care. By monitoring the severity of symptoms on all cases and directing services to those most in need at any point in time, the system of care can accomplish a more rational and ethical allocation of resources, resulting in greater overall consumer benefit for the dollars expended.

This is best achieved by monitoring the progress of individual patients and providing information on high-risk cases to the treating clinician. The trajectory of improvement in

the early phase of treatment is highly predictive of the eventual outcome. Cases worsening or failing to respond to treatment are at risk for premature termination. Identifying these at risk cases can aid the clinician in proactively reviewing the treatment plan and working to keep the patient engaged in treatment. High-risk case reports can also target cases with severe suicidal ideation or where drug and alcohol abuse may be a factor.

Case mix adjustment is accomplished by comparing the actual change score for each patient to an expected change score based on outcomes for similar cases in the data repository for the instrument. This method provides a means to benchmark outcomes against a normative sample and to estimate the statistical and clinical significance of results above or below this benchmark.

Comparing the actual to expected outcome on a case by case basis is a more powerful statistical technique than simply reporting the percentage of cases that meet the criteria for improvement using the Reliable Change Index (RCI) method. The RCI method involves using a formula to estimate the number of points change on an instrument necessary to state with statistical confidence that an individual consumer has experienced reliable and meaningful improvement. This method does produce easy to understand statistics such as “65% of patients improved”. However, the RCI method does not account well for differences in the severity of the case mix, and makes it very difficult to make meaningful comparisons across sites. For this reason, it’s recommended as the indicator only at the Legislative level.

For outpatient care, it is often difficult to determine when care ends. A formal discharge session is rare. The specifications suggest reporting on cases that have been in treatment at least 90 days. The authors research indicates that most of the change in treatment occurs during this period. Basing the reports on a 12 month moving window of cases admitted between 3 and 15 months previously provides a fair estimate of the clinics outcomes. The 12-month window allows for adequate sample size for meaningful statistical comparisons with other sites.

Following is an example of an actual outcomes report using the model. Note that results are broken out by severity range and that the average number of sessions is also reported.

Provider ID:  
Date of report: 5/8/00

1

Cases included in this report began treatment between 3 months and 15 months prior to the date of the report.

Age Group Severity at intake	Total Cases	> 1 data point		Change (effect size)		Change Index (actual-expected)
		Number cases	Sessions/ Case	actual	expected	
Normal range	68	35	7.29	-0.15	-0.16	0.01
Mildly distressed	78	39	5.87	0.27	0.18	0.10
Moderately distressed	70	41	6.44	0.51	0.43	0.08
Severely distressed	62	39	6.69	0.63	0.77	-0.14
<b>Combined Adult</b>	<b>278</b>	<b>154</b>	<b>6.55</b>	<b>0.33</b>	<b>0.32</b>	<b>0.01</b>

Children & Adolescents						
Normal range	26	11	7.18	-0.27	-0.09	-0.18
Mildly distressed	12	9	6.67	0.03	0.15	-0.12
Moderately distressed	6	5	5.00	0.88	0.36	0.52
Severely distressed	10	5	6.20	1.45	0.45	0.99
<b>Combined Child/Adolescent</b>	<b>54</b>	<b>30</b>	<b>6.50</b>	<b>0.30</b>	<b>0.15</b>	<b>0.15</b>

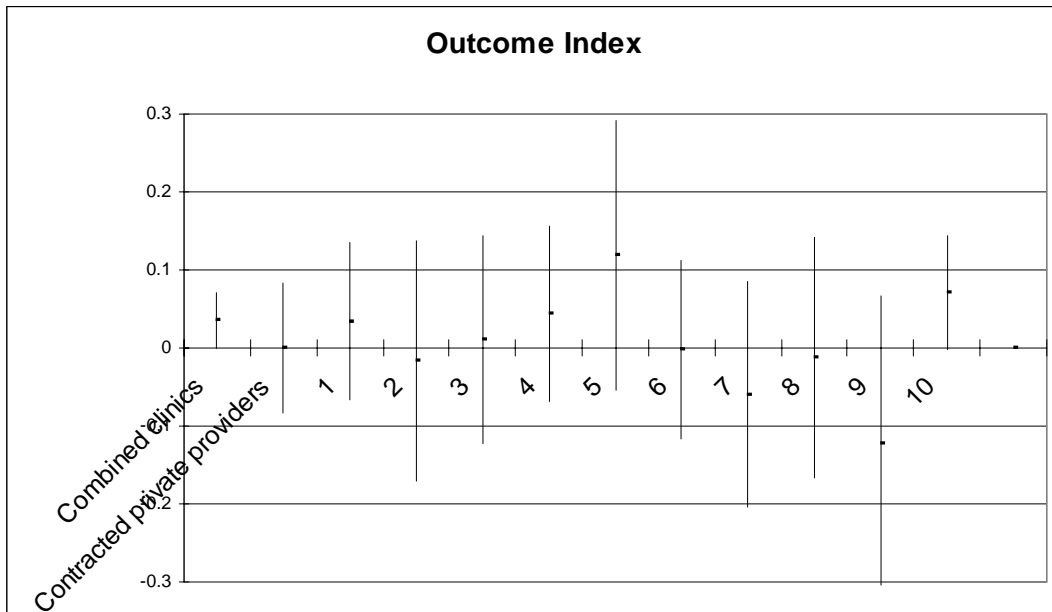
**Aggregate Results for All Age Groups**

Total number of cases:	332			<b>Change</b>	<b>Index</b>
Number of cases with > one data point:	184			<i>actual</i>	<i>expected</i>
% of cases with > one data point:	55%			<b>0.32</b>	<b>0.29</b>
Sessions Per Case:	6.54			<b>0.03</b>	
<i>(actual-expected)</i>					
Average					

The sample illustrates the relationship between severity and the rate of improvement as well as the fact that there is little differentiation in the length of treatment based on severity. Note that cases in the “normal range” actually show slightly negative results. This is not necessarily cause for concern due to the fact that even with a slight worsening of symptoms, the cases are still largely in the normal range at the end of treatment. But it suggest that some of these clinical resources might have been better to allocated to patients in greater need.

The difference between the expected results and the actual results is displayed in the column labeled Change Index. In this case, the actual results slightly exceeded the expected results, resulting in a Change Index Score of .03. This difference is not statistically significant, so the outcomes are categorized as “average”.

Using this method of reporting aggregated results, it’s possible to graphically compare sites. Results are plotted along with the confidence interval. In the following example the results for the combined clinics and for site number 10 show results above average at the 95% confidence level.



This method of reporting outcomes requires instruments with adequate normative samples to yield the case mix formulas that serve as the benchmark. There are several instruments available that would meet this criterion. Where benchmarks aren't available, consistent use of the measures across the state will permit the establishment of statewide benchmarks within the first year of implementation. The benchmark can be adjusted as the data accumulates, assuring the case mix model is always current.

While the previous examples are based on outpatient samples, the same method can be applied to any treatment that is episodic in nature, including crisis stabilization units, inpatient facilities and intensive outpatient programs.

Using an Outcome Index type indicator, results from multiple levels of care can be combined. While different levels of care may result in very different rates of improvement, the Outcome Index always reflects the variance from the average improvement multiple sites with the same level of care. The Outcome Index should average 0 for each level of care across the entire system of care, while individual sites at each level of care will likely vary from this average. By expressing the variation as effect size, a standard metric of outcome is created that is largely independent of the level of care or instrument employed.

The specifications do not envision reporting of outcomes for individual clinicians. Many consumers are seen by more than one staff member, making it difficult to assign "ownership" of the case. Also, the sample sizes for individual clinicians may not be large enough in any given reporting period to permit meaningful comparisons. Lastly, reporting outcomes for individual clinicians can raise anxieties about performance evaluation and tend to undermine acceptance and compliance with the outcomes program.

The specifications also call for a monthly report showing current and projected trajectory of change on all active cases. This report provides an indication of the most recent outcomes and also serves as a case management tool. Following is an example of this kind of report.

Provider ID: Example

Date of Report: 09/21/99

Summary Statistics			
Average Baseline Expected Change	3.9	Average Current Change Index (assumes all cases stop now)	-1.0
Average Current Change Score	2.9	Average Projected Change Index (assumes all cases continue)	1.0
Average Projected Change Score (assumes all cases continue in treatment)	4.9	Change Index - most likely outcome (assumes 40% of cases continue)	-0.2

Average results

**Italicized** indicates case has >75% probability of below average outcome.

Name	Intake Date	Baseline LSQ Score	Baseline Expected Change	Most Recent Session Date	Most Recent Session Number	Most Recent Score	Current Change Score	Current Change Index	Projected Change Index (assumes continued treatment)
<b>Patient A</b>	<b>4/23/99</b>	<b>60</b>	<b>8.12</b>	<b>7/16/99</b>	<b>4</b>	<b>99</b>	<b>-39</b>	<b>-47.1</b>	<b>-20.7</b>
	5/7/99	55	7.34	8/6/99	3	55	0	-7.28	-1.68
	4/29/99	16	-6.52	6/3/99	3	21	-5	1.46	-1.04
	8/19/99	61	9.11	9/1/99	3	58	3	-6.11	-0.38
	4/10/99	22	-0.55	5/14/99	2	17	5	5.55	0.58
	5/13/99	38	1.12	7/13/99	5	35	3	1.88	0.99
	8/20/99	16	-10	9/14/99	9	18	-2	8.02	1.8
	4/14/99	90	19.35	5/12/99	3	78	12	-7.53	2.19
	7/29/99	49	4.94	8/26/99	5	38	11	5.66	4.19
	7/14/99	65	5.64	8/5/99	3	61	4	-1.64	4.71
Patient B	7/17/99	33	-0.62	8/7/99	2	21	12	12.62	7.67
Patient C	5/25/99	58	8.39	7/12/99	5	27	31	22.61	13.61

The report sorts cases from those most at risk for premature termination to those doing exceptionally well according to the projected change index. Cases at the top of the list are responding poorly to treatment but are likely to experience significant benefit if treatment continues, while those towards the bottom of the list could probably be terminated safely from treatment without compromising outcomes. Cases with an 80% or high probability of a poor outcome are highlight in **bold italics**.

In this representative sample of outpatient cases, we have one case (Patient A) that has deteriorated 36 points since intake. If the patient stops at this point, the change index is – 47 (raw score), a very poor outcome indeed. Even if the patient continues, the prognosis for a superior outcome may be poor, but the patient stands to gain 27 points improvement (projected change index of –20). At the other end of the continuum, patients C and D have responded well to treatment, and have scores similar to individuals not in mental health treatment. These patients are not likely to show significant additional improvement and unless their diagnosis was such that ongoing care and medication management was required, could be discharged from treatment in order to shift resources for those cases most likely to benefit further from treatment.

Because the use of repeated measures does require the greatest investment in staff time and behavior change, it may be advisable to pilot the use of these measures on a limited basis during the first year of implementation. This provides an opportunity to validate the model and perform a cost benefit analysis before proceeding to statewide implementation. The pilot project would provide the opportunity to further investigate how outcomes of care items on the retrospective MHSIP consumer survey correlate with real time measures of improvement.

The primary limitation to using consumer self-report measures is the capacity of the individual consumer to comprehend and complete the questionnaire in a reliable manner. If this is not the case, then a clinician rated measure must be used as the primary indicator of improvement.

## **Clinician rated measures**

### **Requirements**

1. Assess functional status for severely mentally ill using standardized clinician rating scales at intake and quarterly thereafter.
2. Assess functional status at intake in all cases. If functioning is above a predetermined level, repeat measures are not required for routine outpatient care lasting less than 90 days.
3. Assess functional status at 90 day intervals for all cases in continuing care.
4. Assess functional status at time of any change in level of care, such as admission to and discharge from an inpatient facility.

### **Recommended Specifications**

#### ***Data set***

- Utilize 10-20 item scales. Different scales may be used for adults or child/adolescent populations.
- Item content should include level cognitive functioning, capacity to carry out activities of daily living, risk of harm to self or others, social role functioning, work or school functioning, living situation, drug/alcohol use and abnormalities in mood and thought processes.

#### ***Reporting***

##### *Individual clinician*

- None

##### *Program, facility or clinic (monthly)*

- Report change in functional impairment scores from intake to current assessment. Utilize a moving 12 months window, i.e. cases entering care between 15 and 3 months previously
- Use case mix model similar to one used for self-report measures
- Break out results by level of care, severity level at admission and age group.

##### *RSN (quarterly and annually)*

- Provide copies of monthly report for facilities and clinics.
- Aggregate results across multiple provider sites. Breakout results by level of care, severity and age group.
- Provide graphical comparison to results from similar sites within the RSN

##### *MHD (quarterly and annually)*

- Report outcomes aggregated at the RSN level. Break out results by level of care, severity and age group.
- Provide graphical comparison of RSNs' results.

*Legislature (annually)*

- Report outcomes aggregated the state level using the RCI method.

### **Rationale and Discussion**

The outcomes framework relies on clinician rated measures when self-report measures are likely to be problematic. As a rule, the more disturbed the patient, the greater the need to rely on clinician measures.

The specification for the item content focuses on functionality rather than symptoms per se. Functional impairment is a primary determinant of level of care. Use of standardized functional rating scales is useful in the more restrictive levels of care because item the content assesses problems most likely to result in hospitalization or need for more structured living environment.

The requirements envision the completion of the rating scale at intake for all cases as part of the assessment and treatment-planning phase. Those in treatment for an extended period will receive repeated assessments.

The change score on clinician rated measures is calculated just as with self-report measures. Likewise, case mix adjustment models will be similar.

## Collateral outcomes

### Requirements

1. Assess success of program in helping consumers to be gainfully employed. For children, use school attendance as an indicator.
2. Assess success at assisting adult consumers maintaining safe and stable housing. For children/adolescents, the indicator is success at maintaining the child in the home environment.
3. Track success in preventing hospitalizations and jail/detention stays.

### Recommended Specifications

#### *Data set*

- Use clinician completed quarterly assessment to capture data the number of hospitalizations and jail/detention stays in the past 90 days.
- Use quarterly clinician assessment to capture data on percentage of work days employed (adults) or school day attended (children) during the past 30 days.
- Include items on consumer survey regarding current housing, length of time at current residence, and perceived safety at current residence.
- Consider adding item asking number of days worked in the past 30 days to self-report outcome measure.

#### *Reporting*

##### *Individual clinician*

- NA

##### *Program, facility or clinic (quarterly and annually)*

- Report percentage of adults employed one or more days in the last 30 days.
- Report change in days worked from intake to current assessment.
- Report percentage school days attended during the past 30 days.
- Report percentage of consumers in the state's priority population without a jail or detention stay in the last 90 days.
- Report of consumers in the state's priority population without a hospital stay in last 90 days.

##### *RSN (quarterly and annually)*

- Copies of quarterly reports for facilities and clinics.
- Compare results of similar sites within the RSN.

##### *MHD (quarterly and annually)*

- Report results aggregated at the RSN level.
- Provide graphical comparison across RSNs.

##### *Legislature (annually)*

- Report combined results on each indicator for a 12 month period aggregated at the state level.
- Where possible, benchmark results against those of other states.

### **Rationale and Discussion**

This group of outcome indicators is appropriate for the sub group of consumers that is the state's target populations. These consumers are most likely to have significant functional impairment. To attempt to use these indicators for an overly broad sample would dilute their utility. By making this part of the quarterly clinician assessment, the data is captured only for those cases receiving longer term care and presumably exhibiting higher levels of functional impairment.

Ideally, one would want to be able to link utilization data and data from the criminal justice system to get the most accurate estimate of days spent in mental health or criminal justice institutions. However, linking this data to specific patients treated in the system may prove difficult.

As an alternative, the clinician is asked to provide the information as part of each 90 day review. The clinician is simply asked to provide information on number of days during the past 90 days that the consumer has been in restrictive settings. Specifying a 90 period provides the denominator when calculating the percentage of days.

Estimating percentage of work/school days attended is more problematic because of the difficulty in determining how many work/school days there actually were in the reporting period. For this reason, the clinician is asked to give an estimate of the percentage.

An intriguing alternative to relying on the clinician estimate on number of days worked is to ask the consumer directly. A single item asking for days worked in the past 30 days could be added to the repeated measure questionnaire. This would open the possibility of measuring change in days worked over time, from intake to the most recent assessment.

Housing is a major concern for many consumers from the point of view of availability, stability and safety. The consumers perspective of the their housing can be readily obtained by adding items to the MHSIP survey.

The case mix adjustment for these indicators is achieved by comparing results for individual programs to those of similar programs. As data accumulates, the case mix model can be refined.

## Structure/Plan Management

---

### Requirements

1. Provide reliable information on the average cost per consumer and cost per unit of service.
2. Provide reliable information on percentage of revenues allocated to administrative costs.
3. Evaluate satisfaction of other community agencies and referral sources, with results tied to individual sites.
4. Evaluate degree that the staff reflects the racial and ethnic mix of the population served.
5. Evaluate the responsive of the system to patients with dual diagnoses.

### Recommended Specifications

#### *Data set*

- Accurate accounting data.
- Standardized service unit definitions
- Average cost per unit of service at the individual site level.
- Accurate count of units of service at both the patient and site level.
- Community Partner Satisfaction Survey (content to be developed). Each clinic or facility should designate who should receive the surveys, with survey responses tied to individual sites.
- Racial and ethnic make up of staff and contracted private providers.
- Use quarterly clinician assessment to evaluate percentage of consumers with dual diagnoses whom have service plans coordinated with other systems.

*Note: A focus study is recommended to validate this measure. See discussion below.*

#### *Reporting*

##### *Individual clinician*

NA

##### *Program, facility or clinic (quarterly and annually)*

- Report average cost per unit of service
- Report average cost per consumer. Break out SPMI/SED and non SPMI/SED if appropriate.
- Report percentage revenues allocated to administrative overhead.
- Report average Community Partner Satisfaction Survey score.
- Report percentage of consumers with dual diagnoses that have service plans coordinated with appropriate state systems.

*RSN (quarterly and annually)*

- Report average cost per unit of service across entire RSN.
- Provide graphical comparisons between similar sites within RSN.
- Report average cost per consumer across RSN for SPMI/SED and non SPMI/SED consumers.
- Report average Community Partner Satisfaction Survey score for entire RSN. Provide graphical comparison of results from similar sites within the RSN.
- Report percentage of consumers with dual diagnoses that have service plans coordinated with appropriate state systems.
- Report percentage of racial/ethnic make up of staff compared to that of consumers

*MHD (quarterly and annually)*

- Copies of all indicator reports for individual RSNs
- Provide graphical comparison of RSNs for all indicators.

*Legislature (annually)*

- Report average cost per unit of service, broken out by level of care.
- Report average cost per consumer.
- Report average Community Partner Satisfaction Survey score.

### **Rationale and Discussion**

The financial indicators of plan management depend upon an accurate and uniform accounting system and method for coding units of service. While this may prove to be one of the greatest challenges to implement, an accurate and consistent method of accounting for cost and units of service is essential for responsible management of limited resources.

The Community Partner Survey provides an indication of how other community professionals who interact with the mental health system on a regular basis view the system performance. By tying survey to individual sites, the information can best be used for quality improvement at the site level.

It is critical to coordinate mental health treatment with alcohol and substance abuse services or services for developmental disabilities when dual diagnoses are present. Tracking this as in indicator focus the therapists' attention of the importance of coordination. However, this is a difficult indicator to capture. Simply asking the therapist to provide the information made not be reliable, and labor intensive methods such as chart reviews are not viable. As with the similar indicator of linkage to physical health services, a small scale focus study using chart reviews can be used to validate the clinician report.

## Implementation

---

Following is a suggested plan to development and implement the performance and outcomes management system. There are of course many pathways to achieving this result. The following model is drawn from the authors' experience in successfully implementing clinical information systems. The term clinical information system (CIS) is utilized to convey the integration of clinical and quality data with other financial and administrative indicators of system performance.

The challenges of implementing a CIS project are both technical and organizational. From the point of view of information technology, the task is to utilize data already being collected within existing databases and integrate this with new data generated through various clinical instruments and consumer surveys. The system must also provide a cost effect and flexible means to capture data and generate regular reports on a myriad of indicators.

The organizational challenge is to overcome the institutional resistance to wide spread use of performance indicators, especially indicators of clinical outcomes. The complexity of the problem of how to measure and compare performance fairly and the natural fear of being evaluated in this manner combine to contribute to this resistance. While all of these concerns deserve attention, they should not be become the basis to derail the process. To Strong leadership is required at the state level to achieve this result.

The plan envisions that the design, development and ongoing operation of the clinical information system will be to a large extent centralized at the state level under the direction of the MHD. This permits economies of scale and contributes to a coherent statewide strategy for performance measurement.

The following sections outline the plan for developing and implementing the CIS. They will cover project management, system design, programming, rollout, and cost estimates.

## **CIS Project Team**

A small group empowered to make the necessary decisions necessary to move the project best guides the task of system design and implementation. The specific issues at a design level are too complex to be managed through committees or other broadly inclusive processes.

At least two of the team should be assigned to the project full time. One is a Project Director. This is a senior position requiring an M.A. or higher, and knowledge of clinical operations as well as research and measurement theory. The Director will have responsibility for oversight and project management. The Director will also need strong communications and writing skills, as this individual will be responsible for training materials and other information disseminated.

As data begins to accumulate the Director will need to guide analyses and reporting of the data. Once the system is fully implemented, the Director's role increasingly focuses on research, data analysis and further system enhancements.

The second full time position is an Analyst/Programmer. This individual should have a BA or higher in statistics and experience using the software applications employed to develop the system. This person will do much of the work to actually program the system.

The other project team can serve on a part time basis and should be broadly representative of the perspective of primary stakeholders including clinicians, administrators, and consumers. Team members may be expected to devote 10-20 hours per week during the first few months of the project, with time decreasing as implementation proceeds. The team needs to be small enough that decisions can be made quickly with the support of all team members. In the authors' experience, four to six total members is a workable size group.

The use of outside consultants familiar with design and implementation of similar systems can facilitate the team's work. In this way the team can avoid the mistakes and build upon the successes of other similar projects.

The CIS team needs to be sensitive and responsive to the needs of the larger organization. The system design must necessarily reflect the mission, values and practical needs of the mental health system. While responsibility for the specifics of the system design reside with the team, one of the functions of the team members will be to elicit input from various stakeholders. Team members are also responsible for articulating to the other stakeholders the rationale for the decisions of the team.

Once the system is implemented, the time commitment of the part time team members can decrease significantly. However, there is an ongoing role for the project team in providing continued guidance on how the system should evolve.

## **System Logic**

The system logic includes definitions of the data set and specifications for how the data elements are to be processed in creation of reports and decision support tools. The system logic can be designed by the project team and does not require extensive knowledge of database structure, programming techniques, etc. The Programmer/Analyst will be responsible for these details. An outside consultant will be useful to guide the team through this process, providing examples of reports and decision support tools from other clinical information systems.

Tasks for the team include:

- Precisely define the data set and how the data will be captured
- Design of new forms for data capture.
- Determine the outputs of the system
  - format of performance indicator reports
  - logical steps to create reports
  - logic of clinical algorithms for care management
  - other decision support tools

Prototype reports can be created to elicit feedback from users in the field. However, the flexibility of the system design will permit easy modification of reports in the future as experience grows.

Time requirement for system design is estimated to be two months.

## **Programming**

The programming methods described here are based on working examples of clinical information systems in the public and private sector. The most relevant comparison is the Illinois mental health data warehouse developed by Weishin Wang, M.A., Bureau Chief of Data Program Analysis, Illinois Department of Human Services. Estimate of time and cost requirements to develop the system are based in part on his experience and advice.

The key to rapid development and flexibility in the system is to use a higher order programming language designed for data management, analysis and reporting on a large scale basis. The SAS system offers this solution. SAS is widely used for statistical analysis and is often required learning for graduates of statistics programs.

While SAS cannot be used to develop stand alone programs that can be distributed without a SAS license, the scripting language used by the SAS software offers a powerful tool for rapid development of complex data management and reporting applications. It also offers state of the art tools for advanced statistical analysis. For these reasons the software is widely used throughout private industry and government. The Illinois project provides a good example of SAS's capabilities. SAS functions as the "backbone" of the system, organizing and storing data for reporting and viewing by end users.

SAS can be interfaced to any ODBC compliant database. This means that existing pathways for data capture and storage in legacy databases can be utilized by the new system. Likewise SAS can accept data from other sources such as data files created through scanning forms, spreadsheets or other database programs. SAS is able to merge the data from different sources to create SAS datasets. These datasets serve as the database for the system. SAS can output data from these datasets for use by other Windows applications. The end user is able to access relevant data or reports using familiar Windows applications such as an Access or Excel.

One of the great advantages of using SAS and other Windows applications in this way is that the entire system can be programmed by the same staff that is responsible for statistical analysis and reporting. Time demands on existing traditional Information Services staff are minimal. Because of the rapid development time offered by this approach, it is also easy to modify the system based on feedback from the users.

Much of the data can be collected using pencil and paper forms. This is one of the most simple, efficient and user friendly human interfaces for data capture available. The forms are converted to computer image files using fax machines or scanners so that data is captured using optical character recognition software.

Teleform is the most widely used application for this purpose. This Windows program permits rapid design of paper forms. In fact, the forms can be saved and distributed as Word documents, to be printed out and copied at the site utilizing the forms.

When a Teleform created form is faxed or scanned to create an electronic image, Teleform can recognize which form it is, read the data, and then output the results to a file with name and format specified by the user. The application can recognize any number of different forms of its own creation, and output the data to the correct file associated with that form, ready to be accessed by the SAS system.

Teleform does not require the presence of a human “verifier” to assure maximum data accuracy. However, the verification process is extremely efficient and a single operator can process hundreds of forms per day. For this reason Teleform lends itself to a centralized operation. Because the program can process such high volume, very few staff are needed to process the forms for the entire state. There are also third party vendors who provide high volume Teleform services for a fixed fee per form.

Combining Teleform and SAS capabilities enables rapid development and deployment of new data sets while simultaneously creating the logic necessary to capture, store, analyze and report on the data. Consultants with SAS experience can accelerate development time. They can assist with the actual programming while providing training and guidance to the Project Director and analyst/programmer. Time requirement for system programming and testing, assuming two FTE SAS programmers, is estimated at three to four months.

## Rollout

The project team will be responsible for planning and directing the implementation. This is best accomplished in two phases, a six-month multi site pilot project followed by full implementation. Tasks include:

- Develop user friendly documentation and training materials.
- Conduct on site training as necessarily.
- Elicit feedback from pilot sites
- Modify forms, reports and decision tools as needed
- Provide consultation and support to local administrators responsible for implementation at specific sites
- Provide consultation in understanding, interpreting and utilizing reports

The pilot phase can begin as soon as the system is complete, approximately six months from start of the project. A sufficient number of sites are needed to fully test and refine the system. Training materials may need to be improved based on user input. The pilot phase should last at least 6 months to provide sufficient experience to understand the steps needed for full implementation. Estimated time requirement for the pilot phase is four to six months. The total time requirement from start of the project to completion of the pilot phase is estimated at 12 months.

The team will need to take time to analyze the data collected during the pilot phase validate the various statistical models and case mix adjustment formulas embedded in the reporting system. Adjustments to the models can be applied as needed. This is also the opportunity to digest lessons from the pilot implementation before proceeding to a full rollout of the performance system. Particular attention need to be paid to sites performing well from a compliance and outcomes perspective. These sites serve as models of best practices to guide future implementation. The project team should issue a report highlighting key findings from the pilot and providing a clear rationale for proceeding to the next phase.

The second year of the project will be dominated by the training, implementation and compliance issues. Resistance to the process is natural and inevitable. Certain sites will do a much better job of collecting data on a consistent basis while other sites can be counted on to test what happens if there is minimal compliance. The task of the project team in the second year is to provide performance feedback on compliance with the measurement requirements. Assuming strong administrative support at senior levels, providing feedback on compliance should shape the system towards more uniform data collection throughout.

In the third year, focus shifts from monitoring compliance to managing results. At this point sufficient data has been collected that differences in results across sites and regions become apparent. Sites with significantly below average results on specific indicators can learn from sites achieving superior results. The role of the Project Director and team becomes much more consultative. The constant feedback and comparisons across sites

and regions should tend to reduce variability and drive performance results upward across the entire system. By the end of the third year, improvement trends should become apparent, particularly for sites with below average results.

## Cost Estimate

---

The cost estimate for the project is based on allocation of salaried staff time and the cost of outside consultants. The estimate does not include the cost of computer hardware, software etc., office space, communications, etc. The cost is broken out by year.

### Year One

#### *Internal staff*

The implementation plan envisions at least two full time employees assigned to the project in the first year: the Project Director and an Analyst/Programmer. The plan envisions two to four additional professional staff, each devoting 10-20 hours a week to the project during the first year. For purposes of this estimate, this is treated as the equivalent of one FTE.

In the second half of the year, the project will need clerical support for processing the data forms that are being generated from the pilot. This individual will primarily utilize Teleform for this purpose. Estimated requirement is .5 FTE for six months. This requirement will vary of course with the number of sites involved and forms to process. However, the cost of support staff for data processing is not a primary driver of the cost for the system.

Total internal staff requirement is projected at 3.25 annual FTEs. Estimated cost of this staff (including benefits) is \$250,000 to \$300,000 in the first year.

#### *Consultants*

The external consultants fulfill two roles. One is to provide technical assistance and guidance throughout all phases of the project, from system design to implementation. The other is to assist in developing the SAS programs and other Windows based applications to view, analyze and report the data.

The hourly rate for high level consultants is estimated at \$100 per hour, while SAS programming time is projected at \$50 an hour.

A budget of \$150,000 to \$250,000 for external consultants in the first year should be adequate to assure that the project can be completed within the time line projected.

**Total costs in the first year is projected to be \$400,000 to \$550,000.**

## Year Two

### *Internal Staff*

The second year of the plan assumes continued full time involvement of the Director and Analyst/Programmer. In addition, support staff will need to be increased to handle the volume of data and reports being generated as a result of full implementation. Support staff requirement is estimated at 2 FTEs.

The time requirements from other members of the Project Team can be expected to decrease during the second year. However, the team should remain in tact to provide continued guidance through the implementation phase. Time requirement is estimated at .5 annual FTE.

Total estimated internal staff requirement is 4.5 FTEs in the second year. The additional FTEs are lower cost support staff, which is partially offset by a reduction in the involvement of high cost administrative and clinical staff.

Total cost for this staff is projected at \$300,000 to \$350,000 in the second year.

### *Consultants*

The need for external consultants should be greatly reduced by the second year of the project, though some involvement may be helpful during the implementation phase. The system is still learning, and as data accumulates the case mix models also need to be tested and revised.

The use of outside “experts” may also help to lend credence to the analyses and reporting being performed by Project Team. Inevitably as results are fed back to the field, there will be challenges regarding the integrity of the data and appropriateness of the analyses.

Estimated consultant costs in the second year is \$30,000 to \$50,000.

**Total personnel and consultant cost in the second year is projected at \$330,000 to \$400,000.**

## **Year Three**

### *Internal staff*

The third year of the project is characterized by routine reporting and continued maintenance and enhancement of the system. The focus of the internal staff is increasingly to support the field in quality improvement efforts, focus studies, etc. The full time project staff remains the same, at two professional FTEs and 2 support staff FTE.

Involvement of other members of the project team falls off to a total of .25 FTE or less. Their continued involvement is recommended to help guide further developments in the system and to provide continued input from users in the field.

### *Consultants*

The need for external consultants should be eliminated by the third year.

**Total staff is projected at 4.25 FTE, at a cost of approximately \$250,000 per year on an ongoing basis.**

## **Summary**

The plan envisions that the system can be designed, programmed, piloted, tested and fully implemented within a two-year period at a cost of between \$770,000 and \$900,000 dollars.

Ongoing operational costs after the second year are projected at \$250,000 per year.