

Behavioral healthcare performance and outcomes management: A survey of best practices

Prepared as part of work performed for the [Joint Legislative Audit and Review Committee](#) of the Washington State Legislature

Jeb Brown, Ph.D.
[Center for Clinical Informatics, LLC](#)

Gary Burlingame, Ph.D.
Michael Lambert, Ph.D.
Brigham Young University

OVERVIEW

Introduction

The purpose of this report is to summarize findings from a survey of performance measurement methods for publicly and privately funded behavioral healthcare delivery systems. The report and supporting documentation is intended to serve as reference material that may inform the design of a performance and outcomes management system for the state of Washington. The term management is used to reflect the authors believe that in order to justify the cost of implementation, performance measurement systems should result in actionable information that assists clinicians and managers at all level to maximize the performance of the system.

Two related documents are particularly relevant to understanding this report. While the length of the document prevents inclusion in the hard copy of this report, the documents can be readily obtained via the Internet.

1. The MHSIP Consumer Oriented Mental Health Report Card: The Final Report of the Mental Health Statistics Improvement Program (MHSIP) Task Force on a Consumer Oriented Mental Health Report Card.
<http://www.mentalhealth.org/consumersurvivor/rptcrd.htm>

This report summarizes the recommendations of the Task Force working under the auspices of the Mental Health Statistics Improvement Program of the Center for Mental Health Services within the U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration.

2. Performance Measures for Mental Health Systems
<http://www.nasmhpd.org/nri/document.htm>.

This document presents a “framework” of performance indicators to monitor and improve mental health services endorsed by the National Association of State Mental Health Program Directors (NASMHPD).

While there are other efforts at creating a framework for performance measurement, the NASMHPD framework of indicators, which builds on the work of the MHSIP Task Force, has achieved broad acceptance among publicly funded mental health systems and is the product of the collaboration of a large number of system administrators, researchers and consumer advocates from across the country. The indicators chosen have been piloted in a number of states, and are currently being adopted in one form or another by a majority of

the state mental health systems. For these reasons the authors have chosen to organize the report around the NASMHPD domains.

The NASMHPD framework includes five domains of performance indicators:

Domain	Examples of indicators
1. Access to Care	<ul style="list-style-type: none"> □ Utilization rates □ Consumer perception of access
2. Quality/Appropriateness	<ul style="list-style-type: none"> □ Consumer participation in treatment planning □ Contact within seven days following hospital discharge □ Consumer perception of quality/appropriateness
3. Outcomes	<ul style="list-style-type: none"> □ Consumer perception of outcome □ Functioning and symptom relief □ Employment (adults)/School improvement (children)
4. Structure/Plan management	<ul style="list-style-type: none"> □ Per member per month average resources spent for mental health □ Proportion of expenditures on administration □ Consumer family involvement in policy development, quality assurance and planning
5. Early Intervention/Prevention	<ul style="list-style-type: none"> □ Use of self-help/self management □ Substance abuse screening

This framework is the basis for the ORYX approved system operated by the NASMHPD Research Institute (NRI). The ORYX initiative is a set of performance and measurement requirements developed by the Joint Commission on Accreditation of Healthcare Organization (JCAHO). JCAHO accredited organizations must submit performance measurement data to a JCAHO-ORYX approved performance measurement system, which in turns submits data electronically to JCAHO. The NRI system is tailored to the individual states' needs, and over 30 states are presently utilizing the system.

The NASMHP performance measures provide a broad framework for evaluating the performance of mental health delivery system but leave room for considerable variability in how the measures are implemented and utilized. The sites surveyed fall on a continuum from being at the initial stages of measurement for certain indicators to having relatively sophisticated methods in place for managing performance targets. A few of sites are employing decision support tools of various kinds. Decision support tools are reports or algorithms that are designed to aid decision-making while treatment is occurring. The intent is to improve outcomes of the system through shaping the process

of treatment planning and delivery. For this reason, information on decision support tools is included in the report.

Site Selection

A representative sample of national managed care companies and state mental health authorities was selected for inclusion in the survey based on a review of scientific and professional literature and recommendations from knowledgeable individuals within academic and governmental institutions. While the survey is not exhaustive, the authors believe that the sites selected are representative of current innovations and best practices of performance indicators systems in use today.

The relative importance of various indicators varies from site to site for both philosophical and practical reasons. The result is that one site may have a performance management system that places great emphasis on a particular indicator, such as clinical improvement, while another may choose to emphasize an indicator such as access to care and consumer satisfaction. Sites were selected for inclusion not so much because of their choice of indicators, but because the site provided an example of how to successfully implement the measurement and management of the indicator(s) selected.

Following is a list of states and organizations surveyed with the abbreviation (if any) that will appear in tables throughout this report.

Arizona	Ariz.
California	Calif.
Florida	Fla.
Georgia	Ga.
Magellan Health Services	MHS
Massachusetts	Mass.
Minnesota	Minn.
New York	N.Y.
Ohio	Ohio
PacifiCare Behavioral Health; Accountable Behavioral Healthcare Alliance (Oregon)	PBH/ABHA
Texas	Texas
Utah	Utah
United Behavioral Health (Washington)	UBH
ValueOptions	VO
Virginia	Va.

Organizing the Data

The site data was gathered via searches on published literature and Internet sites, as well as interviews with one or more informants from each site. A standardized template was utilized to shape the collection and summarization of information from each site. The template helped assure that the same information was being collected from each site while organizing the information for ease of comparison across sites. The templates for each site along with supporting documentation are included as appendices.

The templates do not include information on reporting of utilization rates, cost of care, or percentage of costs allocated to administration. Such financial indicators are based on accounting data necessary to conduct business. Since most states and virtually all managed care companies track and report this information in a relatively consistent manner, to reflect this information in the templates would simply be redundant. However, reporting and interpretation of utilization data will be covered in the body of this report.

The domain of Structure/Plan Management largely reflects the fiscal management of the plan, and is also not included in the templates for the reasons stated above. The issue of consumer and family involvement in policy development is discussed in the body of this report, but not included on the templates. Indicators for Early Intervention/Prevention are in early stages of development, and are not yet widely employed by states or managed care companies. Likewise, these are not included in the templates, but will be addressed in this report.

The following sections provide a summary of findings and recommendations organized by domain. Each section will begin with a very brief synopsis of the main points and recommendations, followed by a table displaying which sites are utilizing specific indicators. For more detail on the sites, refer to the templates in the Appendices.

ACCESS TO CARE

Main Points

- **Accurate utilization and cost data are a prerequisite to effective management of a system of care.**
- **Wait time from request for service to first appointment is a useful indicator of overall responsiveness of the system of care.**
- **>80% consumer satisfaction rate for access to care is a reasonable target.**

Key Indicators

*** indicator utilized**

! indicates that the site is an example of best practices in use of the indicator

	Utilization data	Wait time to first appointment	Telephone statistics	Consumer satisfaction with access
Ariz.	*	*		*
Calif.	*			*
Fla.	*	*		*
Ga.	*			*
MHS	!	!	!	*
Mass.	*	*	*	*
Minn.	*			*
N.Y.	*	*		*
Ohio	*			
PBH/ABHA	*	*	*	*
Texas	*			!
Utah	*	*		*
UBH	!	*	!	*
VO	!	!	!	!
Va.	*			*

Utilization

Access to care is assessed by examining patterns of utilization of services and survey member's perception and/or satisfaction with access. Good utilization data is necessary from the point of view of cost accounting. As a consequence most states, and all managed care organizations (MCOs) collect and are able to report detailed data on the use and cost of services. The MCOs included in this survey have the information infrastructure (collection, analysis and reporting) and many years of institutional experience interpreting utilization data and as such are examples of best practices for this indicator.

Following are the most commonly reported measures of utilization:

Penetration is the percentage of members using the specified service(s) in the course of a year. It is generally expressed as a percentage of the total membership. A related measure is expressed as *admits per 1000*. This is defined as the number of admissions to a specified level of care per 1000 enrollees annually. Admission and penetration are closer related, however, total admissions per 1000 will always be higher than penetration due to the possibility of multiple admissions for a single member.

Duration is a measure of the number of units of service per episode of care. This is generally expressed as an arithmetic mean, such as *average length of stay* for inpatient services or *mean sessions per case* for outpatient services.

Utilization denotes the quantity of services for a specified population during a specified time period. The industry norm is to express utilization as *units of service per 1000 covered lives annually*. Covered lives may also be expressed as *member months*. A *member month* is defined as one individual enrolled for one month. An individual enrolled for 1 year then is the equivalent of 12 member months. *Units of service per 1000 covered lives annually* and *units of service per 12,000 member months* are equivalent expressions of utilization, and may be expressed as $u/1000$, where u denotes the quantity of units of services and *1000* represents *12,000 member months*.

Utilization data is generally broken out by age and sex. However, it is important to be able to break out the information by other categorizations such as ethnicity, diagnosis, SPMI status, etc. in order to assure that target populations are adequately served.

Detailed utilization reports are necessary for effective management, but can easily overwhelm the non-statistically inclined reader with the sheer volume of numbers. The appendices contain three report cards with utilization data that serve to illustrate this point.

The Clark County Report Card (UBH/Washington) states its intention to serve three primary purposes:

- inform key policy makers and stakeholders,
- to communicate with providers, consumers and interested persons
- to provide the system with public accountability

It accomplishes this by reporting key utilization indicators simply and graphically, including graphs of trends over time. Compare this with the Utah Annual Statistical Report or the King County RSN Report Card (UBH/Washington). Both contain pages of table with detailed numbers. While this detailed information is valuable to managers within the system, it probably distracts from the ability of the report card to communicate the most important information clearly and simply.

While it is relatively easy to report utilization using the statistics cited above, determining what is the appropriate level of utilization for a given population is much more problematic. For instance, what represents excessive utilization; and, when is utilization so low as to threaten quality of care? Answers show serious variability. For example, Massachusetts' Partnership Medicaid carve-out reported penetration rate of 23.8% in 1998, while Ohio reported 2.4% in 1997. Similarly, New York reported 449 days/1000 enrollees for its Medicaid enrollees in 1997, while Colorado reported 108 days/1000 in its Mental Health Capitation Project (1997).

This kind of variability was more common in commercially insured populations before the advent of wide spread managed care. Now, however, virtually all commercially insured or employer self-funded mental health care has some degree of managed care methodology incorporated into the plan. The variability in utilization has been greatly decreased, and is more a function of plan design than of the population served. As a result, industry standards for utilization by plan type are available. For example, the Milliman and Robertson, Inc.¹ Healthcare Management Guidelines consist of several volumes of clinical guidelines that represent best-practice benchmarks and utilization data. This is a widely used source that provides detailed information on hospital admissions, length of stay and outpatient utilization for behavioral health care plans. The guidelines specify utilization based on whether the plan is loosely managed, moderately managed or tightly managed.

The relative newness of using managed care methods in the public domain of behavioral health care, and the previously cited wide variations in utilization from state to state make it premature to expect generally recognized benchmarks for appropriate utilization in publicly funded care. The issue of a methodology for arriving at expectations for appropriate utilization will be discussed further under the section on Structure and Plan Management.

¹ Milliman and Robertson Inc. is a large firm of actuaries and consultants serving the full spectrum of business, governmental and financial organizations. For more information go to: <http://www.milliman.com>.

Wait Time to First Appointment

A very direct indicator of the ease of access to care is the time between the initial contact requesting services and the date of the first appoint. This is a particularly important indicator when services are provided through designated clinics so that the consumer does not have the freedom to search out the provider with the shortest wait.

Most MCOs have a standard that varies with urgency of need, i.e., 24 hours for emergencies, 48 hours for urgent care and 5 days for routine outpatient appointment. In practice it is often difficult to compare data across sites because of differences in how data is collected. For example, should the indicator be based on first offered appointment or first kept appoint?

However the measure in calculated, is a highly useful indicator of the capacity of the system to respond readily to the needs of consumers, and as such should be monitored on a continuing basis.

Telephone Statistics

Most MCOs maintain some kind of a call center through which the consumer is asked to access. In these cases, the responsiveness of the call center is an important indicator of access because this is often the point of first contact with the system of care. The two most common indicators are rings to first pickup and the call abandonment rate.

Consumer Satisfaction with Access

The consumers' perception of access is generally assessed as part of a satisfaction survey, such as that utilized in the MHSIP Consumer Survey, (see Texas appendix for example). NASMHPD reports that as of December 1998, 35 states were using or planned to use survey items related to perception of access. While specific wording on items varies from site to site, examples of items tapping consumer perception of access are presented below.

From New York's Mental Health Services Survey

- How would you rate the following? (poor, fair, good, excellent)
 - Getting help in an emergency.
 - Ease of getting services when you need them.

From Value/Options 2000 Member Survey:

- How would you rate the therapist on offering convenient appointment times? (1 excellent, 5 poor)
- Were you able to get your first appointment within 7 days of your call? (y/n)

From Utah's Client Self-Assessment questionnaire: (Responses are rated on a 5-point scale from "strongly disagree" to "strongly agree")

- Services were available at times that were good for me.
- Staff was willing to see me as often as I felt necessary.

Consumer perception of access is almost always reported as a percentage expressing satisfaction. For example, the Value/Options Corporate Quality Indicators report sets a standard of at least 85% expressing satisfaction with availability and accessibility of providers.

Consumer satisfaction with access to care is difficult to benchmark. This is in part due to the tendency of all satisfaction measures to reflect a general high level of satisfaction. Problems with access to care are likely to manifest themselves on consumer surveys as "condemnation with faint praise." ValueOptions sets a standard of 85% satisfaction with access. Utah reports satisfaction with access in the 80% range. The Clark County Report Card (UBH/Washington) shows an improving trend in satisfaction, from below 75% in 1996 to well over 80% satisfied in 1999. These examples are consistent with the authors' experience reviewing satisfaction data. With this in mind, it is probably fair to assume that anything less than an 80% satisfaction rate for access to care is cause for concern.

QUALITY AND APPROPRIATENESS

Main Points

- **Days from hospital discharge to outpatient appointment is a useful indicator of ability of the system to manage high risk consumers through appropriate outpatient services.**
- **Use of guidelines to improve quality and outcome of care is largely in the investigation stage.**
- **Monitoring adherence to treatment guidelines will probably have little impact on outcomes for most outpatient cases.**
- **Consumer participation in treatment planning is easily assessed through use of a consumer survey.**

Key Indicators

* *indicator utilized*

! *site is an example of best practices in use of the indicator*

	Adherence to best practices or treatment guidelines	Appropriate use of medications	Contact with outpatient provider within 7 days of hospital discharge	Consumer participation in treatment planning
Ariz.		*		*
Calif.				*
Fla.			*	
Ga.				*
MHS	!	!	!	*
Mass.		*	*	*
Minn.				*
N.Y.		*	*	*
Ohio				
PBH/ABHA			*	*
Texas	!	!	*	*
Utah			*	*
UBH			*	*
VO			*	*
Va.		*	*	*

This domain is perhaps best conceptualized as measuring compliance to a recommended “best practice.” This may relate to the process of care, such as consumer participation in treatment planning, or have to do with the coordination of care, such as contact with an outpatient provider within 7 days of hospital discharge, or linkage to primary healthcare services. Still others tap use of appropriate modalities, such as use of “Assertive Community Treatment” or supported employment for patients with “Severe and Persistent Mental Illness” (SPMI).

Adherence to various treatment guidelines for specific disorders also falls within this domain. Guidelines may include indicators like number of patients receiving new generation “atypical medications” or appropriate matching of diagnosis, symptoms and treatments.

Because of the complexity of tracking appropriateness of care at the individual patient level, managed care plans will sometimes engage in focus studies that attempt to look at the care for a sample of patients that meet specific criteria. This may be achieved through retrospective chart audits or looking at pharmacy data to determine that the medication usage is appropriate for a particular diagnosis.

Decision support tools are sometimes used to improve treatment appropriateness. Treatment guidelines are one type of decision support tool. For example, the Texas Medication Algorithm Project is an ambitious collaboration between the University of Texas and the Texas Department of Mental Health to develop and implement medication guidelines in order to drive improvements in quality of care and clinical outcomes while increasing predictability of costs for treating certain disorders.

Two of the managed care companies surveyed employ automated decision support tools for care managers that incorporate aspects of treatment guidelines. Magellan Health Services asks providers to submit information on diagnosis, symptoms and treatment plan using the Treatment Request Form to request authorization for services. The form is faxed to a central location where the data is captured electronically using Teleform, a widely used PC based program for this purpose. Once the data is in electronic form, computerized “filters” scan for treatment plans that do not meet the criteria established for diagnosis and symptoms. These plans are targeted for further review by care management, while plans passing all of the “filters” can be certified without further review.

PacifiCare Behavioral Health (PBH) uses complex clinical algorithms in the ALERT system to target at risk cases in both their commercially insured and public sector populations (see case study for Accountable Behavioral Healthcare Alliance in Oregon). Among other things, the system tracks the patient diagnosis, severity of symptoms using a standardized outcome measure, trajectory of improvement as measured by repeat administrations of the outcomes measure, and whether or not the patient has been prescribed a medication.

The appendices contain several examples of report cards with indicators of appropriateness. For example, the ValueOptions Corporate Quality Indicators include information on percentage of provider treatment records that document involvement of the consumer and/or family members as well as the primary care physician in the treatment planning process. The King County Regional Support Network Report Card documents a decrease in the number of days between hospital discharge and face to face contact with an outpatient provider. The Annual Statistical Report from the Utah Division of Mental Health reports the percentage of consumers completing a consumer survey who agreed with the statements that they were included in decisions about treatment services and that questions or concerns were promptly resolved.

More complex indicators such as adherence to treatment guidelines are difficult to report at the system level, but are ideal candidates for focused research studies. United Behavioral Health for example is investigating the impact of disseminating treatment guidelines for specific disorders on subsequent treatment decisions by the provider and consumer. Similarly, the Texas Algorithm Project is investigating the impact of medication algorithms on the treatment of patients enrolled in the project.

It is probably premature to expect external benchmarks for appropriateness indicators. One problem is that some variance from treatment guidelines is to be anticipated as normal. For example, a guideline for depression might recommend that the patient remain on an antidepressant for at least six months. In practice, however, it is not reasonable to expect this to be true for all cases, and it is also unclear when variance from this guideline reflects a proper clinical decision on the part of the provider versus substandard care.

The authors own analysis of large datasets of cases in outpatient manage care populations suggests that it may be difficult to prove that adherence to treatment guidelines will measurably improve outcomes. The analyses have failed to find reliable differences in results based upon which treatment was provided, including the use of antidepressants for depression.² For example, patients with severe depression appeared to experience similar rates of improvement whether placed on a medication or not. Since this data does not come from a randomized trial, the results cannot be interpreted to mean that all treatments are equal, but rather that the providers and patients are making pretty good decisions about what and how much treatment is appropriate. So while over two-thirds of patients with severe depression are on a medication, the minority that are not on medication experience similar improvement because as a group they were judged to be most likely to benefit from psychotherapy alone.

² Brown, GS (Jeb), Dreis, S & Nace, D. What really makes a difference in psychotherapy outcomes? And why does managed care want to know? In Miller, S. & Hubble, M. (Eds.): *Heart and Soul of Change* pp 389-406. American Psychological Association Press: Washington, 1999.

Outcome of Care

Other indicators are relevant because they are believed to have an impact on outcomes. It is not surprising that there is almost universal agreement on the need to measure outcomes, yet implementation of outcomes as an indicator has lagged behind that of indicators from other domains. The problem is that actual measures of patient outcomes are the most difficult to implement and present the greatest challenge, from choice of measures to data collection to statistical analysis and reporting.

Clinical Outcomes

For purposes of this discussion, we have classified outcome indicators as measures of either clinical outcome or collateral outcomes. By clinical outcome, we mean the change in the symptom severity and functioning on the individual patient during the course of treatment as measured by clinician or consumer completed scales.

By collateral outcomes we mean changes in the consumers' engagement with society in general that are deemed beneficial for both the consumer and society, but have not traditionally been the focus of treatment. The most common indicators for ancillary outcomes improvement in living situation, employment status, school attendance/functioning, and decreased involvement with the criminal justice system.

Since clinical outcomes are by definition a measure of the patient's decrease in clinical symptoms and improved functionality, the first decision to be made is the choice of instruments to measure change. The following section will review the range of instruments being employed at the sites, describe how they are being employed and discuss the relative merits and disadvantages of the various strategies.

Choice of Instruments

Main Points

- **A consumer survey is necessary for assessing perception of benefit and satisfaction with services, but may be an unreliable measure of actual improvement.**
- **MHSIP consumer survey has become the standard measure of consumer perception of benefit and satisfaction with treatment.**
- **Repeated measures of outcomes instrument necessary to measure change.**
- **Patient (or parent/guardian) completed measures are economical, reliable and valid measure of improvement.**
- **Clinician completed measures necessary in certain cases, but can be burdensome if used for all cases. Cost of clinician time must be factored into the cost of using these instruments.**
- **Clinician completed measures may be susceptible to bias if used to evaluate performance.**
- **Self-report measures may not be valid for consumers in involuntary treatment or severely functionally impaired.**

Key measures

	Consumer perception satisfaction with outcome	Clinician rated measure of change		Consumer (or parent) self-report measure of change	
		<i>adults</i>	<i>children</i>	<i>Adults</i>	<i>children</i>
Ariz.	*	*		*	*
Calif.	*	*	*	*	*
Fla.	*	*	*		
Ga.	*	*	*	*	
MHS	*				
Mass.	*	*	*	!	*
Minn.	*				
N.Y.	*	*	*	*	
Ohio	*		*	*	!
PBH/ABHA	*	*		!	!
Texas	!	*	*		*
Utah	*	*		!	*
UBH	!	*	*		
VO	!		*	*	
Va.	*	*	*		

States and managed care companies are approaching the problem of outcome assessment from multiple perspectives. These efforts can best be described as in their earliest stages of implementation, and it is premature to expect the information to have any widespread impact on administrative and clinical practices. However, there are some promising beginnings. This section will attempt to summarize the main issues while comparing and contrasting the various approaches currently employed.

This summary must necessarily involve discussion of methodological and statistical concepts and issues. The authors intend that this discussion become no more technical than is necessary to address the primary issues, and refer the reader to the cited references for more in depth information.

Assessment of outcomes implies measurement of change in the patient's status over time. Some patients are expected to improve faster/more than others, and for some of the severely and persistently mentally ill maintaining some modicum of stability and functionality may even constitute a superior outcome. There are also collateral outcomes from a societal perspective, such as a reduction in homelessness and the number of mentally ill jail inmates, or increasing school attendance and employment. But for the vast majority of consumers, the goal of treatment is the reduction of psychological distress and improvement in quality of life, and therefore the measurement of outcomes requires the assessment of change over time.

The NASMHPD indicators include three broad measures of change:

1. Consumer perception of positive change as a result of services.
2. Percentage of consumers with maintained or improved levels of functioning.
3. Percentage of consumers experiencing symptom relief.

Consumer perception of change is assessed via a retrospective survey, such as the MHSIP survey, with items such as:

- I deal more effectively with daily problems.
- I am getting along better with my family.
- My symptoms are not bothering me as much.

A retrospective consumer survey is the easiest outcomes assessment methodology to implement because it only requires data collection at a single point in time and can be accomplished via mailings or telephone interviews. A variation on the retrospective survey is a “slice in time” survey, assessing all consumers at a certain point in time, such as one month out of the year. This has the advantage of looking at a sample of consumers at many stages in the treatment process, permitting inference of change by looking at difference in consumers early in treatment versus some later point.

The ValueOptions implementation of their Member Survey 2000 is an outstanding example of the slice in time survey method (see attachments in appendix). The survey samples a cross section of present and past consumers. An independent research firm conducts the survey, which is mailed out to a random sample of plan members. A smaller sample is also interviewed by phone. Items include:

- Since your treatment began, in general are you feeling...better, about the same, or worse?
- Since the beginning of treatment, are you better, about the same, or worse in getting along with family, friends and co-workers?

While the survey method is ideal for assessing satisfaction and has the advantage of ease of implementation, the value of the outcomes information obtained in this manner is limited. It is difficult to draw any conclusions about the actual magnitude of change, or how an outcome for a particular consumer compared to that of similar consumers. The authors own analyses of satisfaction and outcome data suggest that the consumer’s retrospective assessment of improvement is more a function of the consumer’s current level of psychic distress rather than actual improvement measured during treatment.

Measuring consumer improvement through assessment at two or more points in time introduces a whole new set of problems. From whose perspective should we assess change: the treating professional’s, the consumer’s, or from a parent or other collateral source? What instruments should be used? How often are they to be administered?

The greatest challenge is the actual collection of the data. Since data needs to be collected at the start of treatment, and at some later point in time, this requires changes in work flow at the provider level to administer the measures at specified points in time and changes at the administrative level to assure compliance with the data collection protocol. The additional work is often seen as needless, burdensome, and irrelevant to the job of a busy clinician “in the trenches.” Inevitably much of the data collected is unusable either because the intake or some follow-up assessment is unavailable. One without the other is not of much use.

The problems with attempting systematic collection of outcome data in the field are sufficiently daunting to lead at least one large managed company, Magellan Health Services (MHS), to conclude that it is not feasible to systematically measure change at the individual patient level across their system of care. Instead they have chosen to utilize smaller focus studies tailored to the needs of the regional operations centers.

The MHS experience should strike a cautionary note. When MHS acquired Human Affairs International (HAI) in 1998, it acquired a working clinical information system that was designed to track and report outcomes for all outpatient cases. The HAI project had found that results appeared to be moving in a strong upward trend over time, particularly at sites making full use of the decision support capabilities of the system. These sites were averaging 25% more improvement per case while utilizing fewer visits than for similar patients treated by clinicians without access to the performance feedback and decision support capabilities the system provided. After conducting a review of the system, MHS deemed that the data was unreliable due to a small percentage of cases with multiple data points. Thus, they decided that the effort to insure compliance with protocols would be too great and would encounter too much resistance to justify continued use of the HAI system.

Other states and managed care organizations such as Ohio, New York and PacifiCare Behavioral Health are pushing ahead with efforts to systematically measure change at the individual patient level. The ALERT system utilized by PBH/ABHA is similar the HAI system. The remainder of this section will examine the challenges these sites face and solutions proposed.

On the question of whose perspective to measure, virtually all of the sites measuring outcomes have chosen to include a consumer self-report measure with adult populations. This has the advantage of making intuitive sense while avoiding the need for clinicians to receive training in order to assure inter-rater reliability, not to mention the ongoing time requirement to complete forms.

There are circumstances in which self-report measures are problematic due to the response of the consumer to deny either or exaggerate symptoms. In the case of involuntary treatment, there may be significant incentives to deny problems and thus render the measure useless. The reverse can also be true if there is an incentive to malingering.

However, the same can be said of clinician completed measures. While training of raters can improve the reliability of clinician completed measures, the clinician is not disinterested in the outcome. If the measures are used to evaluate clinical performance, this inevitably creates an incentive to bias the assessment. Such systematic bias is very difficult to adjust for statistically.

While patients have their own idiosyncratic biases, this source of error is likely to be randomly distributed across providers. Further, it would be difficult for a provider to induce systematic bias on the self-report measure. In short, self-report measures of outcomes are more “cheat proof”.

Adult measures

There are a number of good self-report measures. For adults, probably the most widely used instruments are the BASIS-32, OQ-45 (including investigational versions), and the Treatment Outcome Package (TOP).

The BASIS-32 was developed by Susan Eisen, Ph.D. at McClean Hospital in Massachusetts. Two sites surveyed utilized it across their entire system (New York and Georgia). It has also been utilized by ValueOptions and MHS in some of their studies.

The OQ-45 was developed by authors Lambert and Burlingame. The instrument has been widely used for research purposes and has been employed by a number of managed care companies including HAI. It is currently utilized at a number of sites around the country, including Intermountain HealthCare and Valley Mental Health in Utah and by PBH at selected sites in California. PBH also sponsored development of a shortened version of OQ-45, the 30 item Life Status Questionnaire (LSQ), that is now being utilized across its entire system of care for both commercial and public sector business.

The Treatment Outcome Package (TOP) is a comprehensive clinical assessment and outcomes tool that has been adopted by a number of organizations during the past few years. This package of instruments is available through Access Measurement Systems, Inc. (AMS)

The TOP differs the BASIS-32 or OQ-45 in that it seeks to be both an assessment tool and outcomes measure. As such, the questionnaires are much longer than other instruments cited here. For example, the adult intake assessment is 140 items. The TOP is currently being utilized at sites in over 30 states (see Massachusetts and Utah). The company has established an impressive list of collaborating organizations including the American Psychological Association and Price Waterhouse Coopers.

One of the persistent questions regarding self-report measures is their appropriateness for the SPMI population, which due to the level of impairment may not be able to complete the questionnaires in a reliable and valid manner. For at least some of these consumers it clear that a clinician completed measure is preferable.

The Multnomah Community Ability Scale (MCAS) was developed for this purpose. Its use is mandated in Oregon, and it is also being employed by several other states including Virginia and Texas. The instrument is designed to measure functionality in a number of areas critical to SPMI populations.

Another measure of functional impairment is the Global Assessment of Functioning, or GAF. It is a part of the five Axis model of diagnosis used by the industry standard DSM-IV. However, the GAF is really a single item rating patients from 1-100 on global functionality. Because it is so global in nature, it is rarely used for research purposes, and none of the sites surveyed are using it as a primary measure of outcome.

Child/adolescent Measures

The choice of measures for children is complicated by the fact that with younger children it is obvious that a self-report measure cannot be utilized because of age-based problems such as reading comprehension. For this reason one of the most widely used measures, the Child and Adolescent Functional Assessment Scale, (CAFAS) is clinician completed (Virginia, Georgia, Ohio, California and parts of Oregon).

An alternative to a clinician completed measure is to ask the parent or guardian to complete a questionnaire about their child. Several instruments identified in this survey use this approach:

- Child Behavior Check List (CBCL), developed Thomas Achenbach, Ph.D.
- 64 item Youth Outcome Questionnaire (YOQ) and 30 item Youth Life Status Questionnaire (YLSQ), developed by Lambert and Burlingame.
- Ohio Youth Scales, developed by Ben Ogles and the Southern Consortium for Children
- TOP child measures, developed by AMS, Inc.

The CBCL is currently utilized in Texas, while the YOQ is employed in Utah and in the Texas Algorithm Project. The investigational 30-item version of this instrument, the YLSQ, is currently employed by PBH for its entire book of business. The AMS TOP child measures are in use at a growing number of sites around the country.

The Ohio experience may be instructive in this regard to choosing a clinician rated measure versus child/parent report measure. Ohio utilized the clinician completed CAFAS and Behavioral and Emotional Rating Scale (BERS) for over a year in their Outcomes Implementation Pilot. After experience with the instruments, the state opted instead to use the Ohio Youth Scales. In addition to easing the administrative burden on clinicians, the state cited the advantages that the instruments were free, presented an integrated and unified format, were shorter and less complicated than the CAFAS or BERS, and permitted the comparison of identical items across youth, parent/adult and

clinician completed versions. PBH cited similar reasons for the use of the YLSQ over the CAFAS in Oregon.

Frequency of Repeated Measures

The timing and frequency of when consumers are assessed can have significant impact of the findings. The frequency of repeated assessment in this survey ranged from every other session to once a year, with the majority of sites specifying assessments at three or six month intervals.

California elected to measure outcomes on only on SPMI patients. However, the clinician has up to six weeks to determine if the patient meets the criteria for SPMI before administrating the measure. Subsequent administrations are annual. Not surprisingly, this approach has resulted in a relatively small sample of cases with two data points.

More surprisingly perhaps, these cases have tended to show little improvement. In fact, this finding may be misleading. Since the most rapid improvement tends to occur at the start of treatment, it may be that the first administration of the instrument was occurring after the patient had also showed a positive response to treatment.

At the other end of the continuum, the ABHA/PBH site collects data at the first, third and fifth outpatient sessions, and quarterly thereafter. While this may appear burdensome, it does have several benefits. More frequent measurement may tend to result in a larger estimate of benefit from treatment.

The authors' analyses of the data repository indicates that patients with relatively short lengths of treatment (ten sessions and less) tend to have experienced much more rapid recovery and better over all outcomes. Longer lengths of treatment are associated with more difficult cases rather than superior outcomes.

Frequent measurement at the start of treatment also reveals that a significant number of cases (10-15%) show significant deterioration during the first few weeks in treatment and terminate prematurely significantly more symptomatic than when they began. Identifying these at risk patients prior to the fifth session provides the clinician with the opportunity to review the treatment plan and take steps to insure that the patient stays engaged in treatment. This information is lost with less frequent measurement.

A final argument in favor of frequent early assessments is that it results in a much larger and more representative sample of cases upon which to evaluate performance. While this might be less of an issue when aggregating results at a state level, small sample sizes seriously limit the usefulness of the data when comparing performance across programs or clinics. In light of these arguments, the ABHA decision to capture three data points within the first five outpatient encounters before dropping back to quarterly assessment is defensible, and results in greater utility for outcomes management program.

Collateral Outcomes

Main Points

- **Change in the consumer’s engagement with society in general that are deemed beneficial for both the consumer and society.**
- **May be of more interest to state legislators than clinical outcomes.**
- **Difficult to benchmark as these outcomes may be influenced by other conditions in the community beyond the control of the mental health system.**
- **May be more “cheat proof” than clinical measures.**

Key indicators

	Living situation	Employment status	Criminal justice involvement	School attendance
Ariz.	*		*	
Calif.	*	*	*	
Fla.	!	!	!	!
Ga.	*	*		
MHS				
Mass.		*		
Minn.	*	*		*
N.Y.	*	*	*	*
Ohio	*	*	*	*
PBH/ABHA				
Texas	*	*		*
Utah	*	*		*
UBH	*	*	*	
VO				
Va.	*	*	*	

Mental health professionals are understandably interested in clinical outcomes, but state legislators are likely to be interested in the broader impact of treatment of social problems such as homelessness, illegal activity and involvement with criminal justice, unemployment and school involvement. All of the sites reviewed included one or more of these indicators.

In Florida, the Office of Policy Analysis and Government Accountability (OPPAGA), an office of the Florida State Legislature, is responsible for establishing performance indicators and providing regular reports to the legislature. Because the legislature is the primary “customer” for the performance indicators, the OPPAGA system is weighted heavily for collateral outcomes, with relatively little emphasis on clinical outcomes as measured by clinical or consumer report.

Ted Harrell, Senior Management Analyst; Division of Mental Health, explained this focus very simply. The legislature was not interested in test results, and did not necessarily trust clinician reported outcomes to be accurate, particularly if funding decisions are contingent upon performance. These collateral outcomes are seen as both more cheat proof and relevant to the primary interests of the legislature.

Not surprising, many clinicians in Florida have complained that these are not true outcomes of treatment, and may be influenced by other factors beyond the control of the mental health system. Despite these objections, OPPAGA has implemented these measures and continues to refine them, while the legislature looks seriously at the performance indicators and is prepared to make funding decisions based on performance. For this reason, Florida serves as an example of best practices in measuring collateral outcomes. The reader is referred to the Florida appendix for examples of OPPAGA performance reports.

One of the primary indicators for the SPMI population in Florida is “Days in the Community” which is simply the number of days that the program clients spend in community settings and not in jail, detention facilities, crisis stabilization units, residential treatment facilities, inpatient psychiatric units, or homeless. The beauty of this indicator is that it combines elements of criminal justice involvement, housing status, and clinical stability into a single easily understood measure expressed as days per month. The Legislature adopted a standard of 300 days per year, or 25 days per month in the community.

The Community Days’ indicator is used for children services also. However, beginning in 1999-2000 clients in juvenile justice facilities were excluded from this indicator because determination of placement out of the community was based on their legal, rather than mental health status. School attendance is another of the OPPAGA indicators, reported as percent of available school days attended in the last 30 days.

Another type of collateral indicator utilized by Florida is satisfaction of “community partners,” i.e., law enforcement, judiciary, etc. The satisfaction surveyed specific concepts such as timeliness and cooperation, as well as over all satisfaction.

Reporting Clinical Outcomes

Main Points

- **Clinical outcomes are measures of change from one point in time to another**
- **Simple reporting of percentage of consumers improved is too crude a measure for effective comparisons across sites.**
- **Case mix adjustment is critical to making valid comparisons of outcomes**
- **Most powerful method for case mix adjustment is to use an outcomes prediction method, comparing actual change to predicted change using case mix adjustment formulas.**
- **Existing instruments with a large, broadly distributed normative sample of patient outcomes can provide external benchmarks for outcomes.**
- **Measures of change are most useful for measuring effectiveness of relatively discrete episodes of care such as routine outpatient treatment, crisis stabilization, or inpatient care.**
- **Patients in long term care may exhibit little improvement over time. For these cases, stability over time is a more appropriate indicator.**
- **Analyses of outpatient samples consistently shows little differentiation in treatment duration between the least symptomatic and most symptomatic consumers despite the fact that the least symptomatic show little or no improvement in treatment.**

Reporting of Outcomes

* currently reports indicator

! best practice

	Consumer perception and satisfaction with outcome	Clinician reported change based on repeat administrations of standardized measure		Consumer (patient/parent) reported change based on repeat administrations standardized measure	
		<i>adults</i>	<i>children</i>	<i>Adults</i>	<i>children</i>
Ariz.	*	*	*		
Calif.	*				
Fla.	*	*	*		
Ga.	*	*	*	*	*
MHS	*				
Mass.	*			!	!
Minn.	*				
N.Y.	*	*		*	
Ohio	*	*	*	*	*
PBH/ABHA	*			!	!
Texas	*	*	*	*	*
Utah	*			*	*
UBH	!	*	*		
VO	!		*	*	
Va.	*	*	*		

While it is a simple matter to record change scores on the test (i.e. the difference in scores from the beginning until a later follow-up point), this information alone is of little use. Many questions need to be answered: What constitutes a “good outcome”? How much change is necessary to say a patient is improved or recovered? How do you compare results for different providers with a different mix of cases?

The most widely used method to report outcomes, is to give the percentage of patients improved, stable, or deteriorated. These categories are determined by whether the points change on the test exceed the Reliable Change Index (RCI) for that test. In fact, the NASMHPD Performance Measures specify that the RCI be used to calculate the percentage of consumers with maintained or improved levels of functioning.

Following is the documentation from the NASMHPD³ indicators on the calculation and use of the RCI:

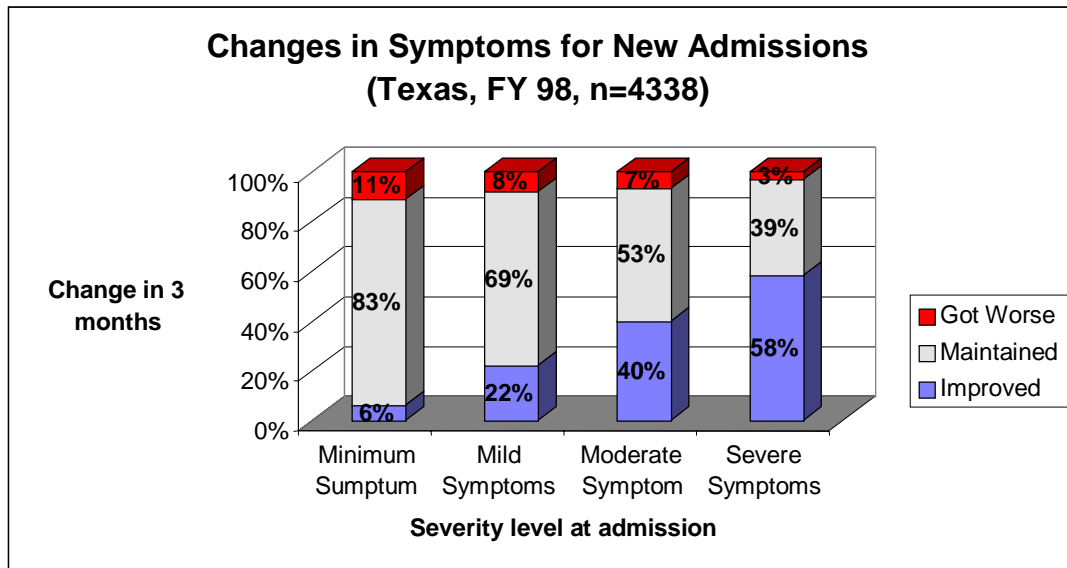
“The Reliable Change Index, (RCI) described in the literature by Jacobson and Traux, is a statistical method to calculate how much change has occurred at post-test in cases where using a cutoff point is not statistically reliable. The formula for the RCI is:

$$RCI = \frac{X - X}{Sdiff}$$

Where X represents the pretest score, X represents the same consumer’s post-test score, and Sdiff is the standard error of the difference between the two test scores. Sdiff can be computed directly from the standard error of measurement (SE); Sdiff = the square root of 2(SE). An RCI of 1.96 or greater is significant ($p < .05$) and would be unlikely to occur without actual change. In other words, using the RCI adjusts for the S.D. in the pretreatment group and the test-retest reliability of the measure. But, using the RCI alone does not allow one to differentiate change based upon the value of the pre-test score. That is, it does not differentiate the significance of a similar point change in persons starting out at different levels on a scale (a one point change from 5 to 6 versus 34 to 35). Using the log of the pre- and post-test scores and the log of Sdiff allows for this differentiation. This is useful in determining the amount of change that is needed to reach significance for people starting out at high as opposed to low scores and for ascertaining cutoffs where change is not possible, based upon the raw score change needed to achieve a significant RCI (ceiling effect).”

The NASMHPD document recognizes one of the difficulties with the RCI is that it does not differentiate the significance of change in individuals with very high levels of distress versus those with minimal symptoms. Though the NASMHPD document does provide for a statistical solution, this is often disregarded in reporting, so the percentage that improved or deteriorated is based on the number of cases that change, more than a fixed number of points as determined by the RCI, (Utah, 1999 Annual Statistical Report; King County Washington 1998 Prepaid Health Plan report Card). Texas addresses the problem by breaking out results by quartile based on level of severity at intake, as the following graph illustrates.

³ <http://www.nasmhpd.org/nri/document.htm>, page 46.



As we see here, of the patients with the fewest symptoms, 6% improved and 11% got worse while those with the most severe symptoms had 58% improved on only 3% deteriorated.

Another way of evaluating outcomes is to look at the percentage of patients that improve into a “normal range”. In this case, the normal range is determined by a representative community sample of individuals not seeking mental health treatment. By comparing the community sample with a clinical sample, a cutoff score can be estimated that constitutes a boundary between normal and clinically significant levels of life distress⁴. The formula is:

$$\text{Cut-off} = (\text{SD}_1 * \text{M}_2 + \text{SD}_2 * \text{M}_1) / (\text{SD}_1 + \text{SD}_2)$$

The use of a clinical cutoff score helps define the boundary of “normal”. While individuals below the cutoff score may be experiencing psychological distress, their “unhappiness” is within the normal range of human experience when confronted with life’s vicissitudes. The validity of this type of cutoff as a definition of normality is reinforced by the fact that consumers entering treatment in this range tend to show minimum benefit, as the Texas example of cases with minimal symptoms illustrates (6% improved, 83% unchanged and 11% worse).

The greatest problem with the use of RCI as an indicator is the case mix adjustment problem. The percentage of cases likely to show improvement based on the RCI is dependent upon the severity of the sample. Breaking the sample into severity ranges helps, but it is still difficult to compare results across multiple sites. Is a 40%

⁴ Jacobson NS, Follette WC, Revenstorf, D: Toward a standard definition of clinically significant change. Behavior Therapy 17:308-311, 1986.

improvement rate statistically different from a 35% rate? Of course it depends on the size of the two samples, but this information is difficult to convey in a simple graphic manner to a non-statistically sophisticated audience (which includes most clinicians).

There is an alternative method analyzing and reporting outcomes that gets around this problem. The approach is based on methods pioneered by Kenneth Howard, John Lyons and colleagues. The following is an excerpt from their seminal book: *The Theory and Management of Clinical Outcomes in Mental Health*.⁵

As alluded to in the section on change analysis, it is often possible to use consumer and service characteristics to predict levels of clinical status variables over time. For example, it might be possible to predict symptom level based on the initial level of symptoms and the duration of treatment. If outcomes can be predicted, they can be compared to actual outcomes, and the deviation between actual and predicted outcomes can be understood within the context of quality of service. In other words, cases that result in outcomes better than anticipated are more highly rated. However, cases with outcomes below predicted levels point to areas in need of quality improvement. (page 85)

They go on to distinguish two basic approaches to outcomes prediction:

1. Static outcomes prediction is more easily achieved and involves the prediction of outcomes at a single point in the future. End of treatment outcomes are the mostly likely target prediction.
2. Outcomes trajectory prediction, by far the more powerful and interesting approach, involves the prediction of the course of recovery. In this approach, changes in outcome are modeled over time. Thus, the resulting prediction is not a final outcome but rather a pattern of changing level of outcome over the course of a treatment episode. (page 86)

The difficulty with this second approach is that it requires a very large normative sample to calculate the various patterns of change for different diagnostic groups and treatment populations. Such a sample is difficult to obtain because it requires collecting data at multiple points in treatment from a very large population of providers and consumers. In order to benefit from this approach, it is necessary to continue to measure all cases at multiple points in treatment. Given the previously mentioned problems with achieving compliance with data collection protocols, it is not surprising that this method is not more widely employed.

Only two systems reviewed are able to produce true case mix adjusted outcomes reports: the AMS system (used by community mental health providers in over 30 states including Massachusetts, Utah, and Washington); and, the PBH ALERT system (used in all of their commercially insured population and with the Accountable Behavioral Healthcare

⁵ Lyons LS, Howard KH, O'Mahoney MT, et al: *The Measurement and Management of Clinical Outcomes in Mental Health*, New York, John Wiley & Sons, 1997, p 85 and 86.

Alliance in Oregon). In both instances case mix adjustment is achieved by benchmarking actual results against a statistical prediction of change for each individual patient. The formulas to compute these predictions are derived from analyses of large normative samples of patients with repeated measures in treatment.

The authors' data repository includes over 20,000 adult cases with repeat measures and approximately 10,000 child/adolescent cases using the OQ-45, YOQ and related investigated versions. AMS has accumulated a substantial database on the instruments (100,000+ cases) but the number of cases with repeated measures is relatively small, numbering a few thousand at most⁶.

While AMS presently has smaller repository of cases with multiple measures, the company's work on developing case mix adjustment models appears to be very sophisticated. Given the richness of their data set and the rate at which they are collecting new data, there seems to be no doubt that as their sample increases they will be able to refine the case mix model.

In principle, there exists several large samples using the BASIS-32. Unfortunately, steps have not been taken to aggregate this data at a single location for purposes of analysis and development of case mix models.⁷

Both the AMS system and the ALERT system add value to the instruments by providing a computerized system for reporting case mix adjusted outcomes and providing decision support to the clinicians.

AMS has the capability to receive a faxed protocol and respond with a faxed back clinical report within 15 minutes. In this way the clinician receives immediate decision support with regard to risk indicators, level of symptoms, and diagnostic possibilities. The system also generates a variety of reports for quality assurance and management purposes. The change for use of the instruments themselves is relatively modest, and the charge for the package of reports is assessed separately, depending on the frequency and complexity of the reports.

The PBH ALERT system likewise captured data via faxed forms. A complex set of decision support algorithms review all cases on a daily basis and generate a High Risk Case Report for all cases at risk for poor outcomes. These reports are forwarded to care managers and treating clinicians within 24 hours of receipt of the data.

Both the AMS and ALERT systems use the static outcome prediction method in reporting outcomes for closed cases. That is, the final outcome for case is compared to outcomes for similar cases, regardless of the length of treatment. The ALERT system goes one step further by tracking the improvement of active cases compared to an expected trajectory of improvement. In this manner, those cases at highest risk for a poor

⁶ Personal communication with David Kraus, Ph.D.; President, AMS Inc.

⁷ Personal communication with Susan Eisen, Ph.D.

outcome can be identified as early as the 3rd session, permitting adjustment in the treatment plan as needed.

Because the ALERT system incorporates static and trajectory of change predictions, the Accountable Behavioral Healthcare Alliance (ABHA) in Oregon serves as a case study for the implications of pursuing the outcomes prediction method to its logical consequences. The remainder of this section will utilize examples from the ALERT system and the authors' own research to illustrate the potential for applying these methods to improving clinical outcomes.

The authors wish to note that we make no proprietary claim to these methods per se and have published our results in an effort to further the science of outcomes management. . A detailed report of the statistical methods utilized in development of the ALERT system has been submitted for publication in a peer-reviewed journal.⁸ While the other systems reviewed have not yet incorporated outcomes trajectory prediction methods, we assume that they will do so in the near future if the method appears to live up to its promise.

The ALERT system provides care managers and providers a target outcome for each case. The goal of clinical management is to keep patient engaged in treatment until this target is reached or exceeded. Concern about the length of treatment is secondary to achieving the desired outcome. The patient's improvement is monitored compared to the expected trajectory of change, drawing attention to those cases that appear to be at risk for poor outcomes. This permits the treating professional to make appropriate changes in the treatment plan before the patient becomes demoralized and terminates treatment prematurely. The ALERT system is designed to permit this outcomes focused approach without increasing cost the cost of care by shaping providers towards a more effective allocation of clinical resources, as the following discussion will illustrate.

First, it is instructive to examine actual results using the Aggregate Outcomes Report from the ALERT system. This report provides case mix adjusted results using the static outcomes prediction approach for closed cases. For every individual patient, the system makes a statistical prediction of the amount of improvement expected based on data collected at intake. This target outcome stays with the case throughout the episode of care and the actual outcome for each case is compared to this predicted outcome.

The Aggregate Outcomes Report displays results by age (adults and children/adolescents) and by four severity levels based on test scores at intake. The lowest severity range is labeled "normal." This is due to the fact that scores in this range are more characteristics of a non-treatment sample than a sample of patients in treatment. For each subgroup, the report displays the average predicted improvement, the actual improvement, and the difference between the two, which is referred to as the Change Index. Change Index values greater than zero represent above average results.

The report is able to aggregate results from separate outcomes measures for different age groups by converting all raw change scores to a standardized change score, or effect size,

⁸ Brown, G.S. (Jeb), Burlingame, G.M., Lambert, M.J., Jones, E., Vacaro, J. Pushing the Quality Envelope - A New Outcomes Management System. Submitted to the Journal of Psychiatric Practice.

which is calculated by dividing the raw score change by the standard deviation of the measure. The result is a single measure of outcomes performance across the entire system of care. In the case of PBH, one of the performance targets for senior management is to achieve an overall Outcome Index score of greater than zero.

The following example of the Aggregate Outcomes Report for a large sample of PBH's commercially insured consumers. Because the measured effect size and the Change Index are continuous variables with known psychometric properties, it is relatively straightforward to calculate a confidence interval around the Index. In this case, despite the fact that the Outcome Index is positive, the difference from the expected change is not statistically significant ($p > .05$), and therefore the report indicates that the results are average, or in line with expectations.

Aggregate Outcomes Report

Date of report: 4/18/2000

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	
Adults	Normal range	885	276	6.33	-0.03	-0.15	0.12
	Mildly distressed	891	326	6.32	0.23	0.19	0.04
	Moderately distressed	902	403	6.84	0.41	0.44	-0.03
	Severely distressed	887	392	6.56	0.78	0.80	-0.01
	Combined Adult	3565	1397	6.54	0.38	0.36	0.02

Children & Adolescents							
	Normal range	195	68	6.34	-0.06	-0.09	0.03
	Mildly distressed	170	66	6.95	0.24	0.16	0.08
	Moderately distressed	129	47	6.55	0.50	0.37	0.13
	Severely distressed	334	56	5.96	0.62	0.43	0.19
	Combined Child/Adolescent	828	237	6.46	0.30	0.19	0.10

Aggregate Results for All Age Groups

Total number of cases:	4393			Change Index
Number of cases with > one data point:	1634			
% of cases with > one data point:	37%			
Sessions Per Case:	6.53			

Change		Change Index (actual-expected)
actual	expected	
0.37	0.34	0.03

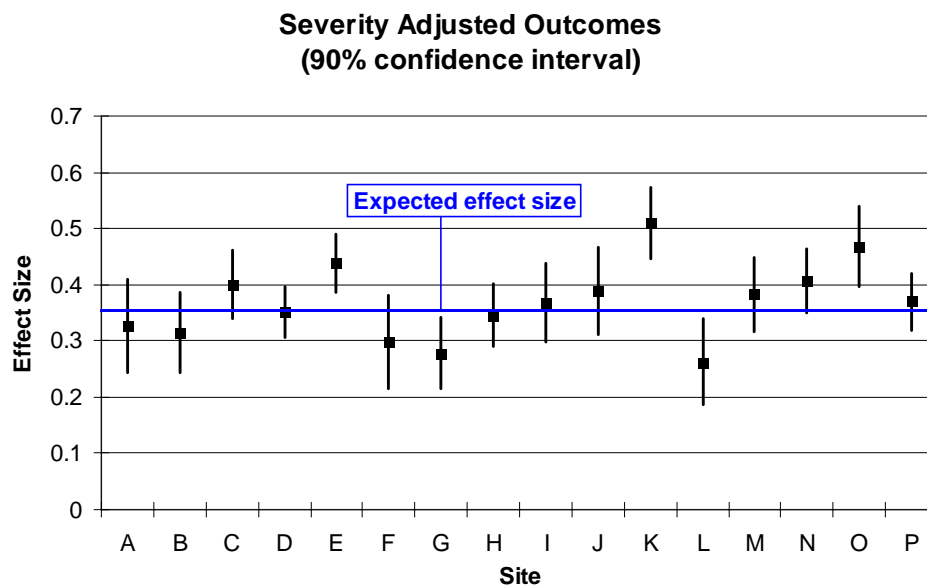
Average

This report provides evidence of how resources can be reallocated to achieve superior results without increasing cost. There is very little difference between the average length of treatment for cases in the normal range versus severe range of distress, despite the fact that cases in the severe range are realizing much greater benefit per treatment session than those in the normal range. In fact, cases in the normal range actually show a slight deterioration during treatment, though this is of no clinical significance due to the low level of distress to begin with. An increase in service cost for those cases most at risk (and most likely to benefit) can be offset by a modest shift in resources away from the least distressed consumers.

Simply reporting outcomes is not enough to drive change within the system. Annual reports are of interest to system administrators and state legislators perhaps, but at the clinic or facility level, frequent feedback combined with attention to the results is most likely to drive improvement. The Clark County Report Card provides an example of this principle using the Consumer Satisfaction Survey – 8. When measurement was first

implemented, satisfaction ratings on the various items ranged from a low of 63% to a high of 80%. However, the survey was administered on a quarterly basis and results were fed back to the facilities. The result was a clear upward trend in satisfaction over the next four quarters and satisfactions ratings ranging from 81% to 92%.

ABHA takes a similar strategy of frequent performance feedback. Each clinic site utilizing the ALERT program receives the Aggregate Outcomes Report on a monthly basis. They also receive a graph plotting their results compared to that of the other group practices. By adding each site's Change Index score to the average expected change for the entire population, it is possible to portray each site's overall case mix adjusted effect size so that the relatively magnitude of differences across sites becomes apparent. The following is an example of the effect size for different sites graphed with confidence intervals. In this case, confidence intervals are set at the 90% level. This graph conveys at a glance the relative effectiveness of the various sites. By feeding this performance information back to the providers and providing ongoing decision support tools to aid in clinician decision making, PBH and ABHA seek to foster a system of care in which outcomes will tend to improve over time.



Benchmarking outcomes

The ALERT system demonstrates the feasibility of benchmarking outcomes against a national norm. The formulas used to predict outcomes were derived from an entirely separate sample than the one being measured. In this case the normative sample consisted of commercially insured patients treated through a variety of managed care plans *other than PBH*.

By using these norms, PBH management effectively benchmarks their outcomes against a national sample of other managed care companies. Providers using the AMS system can likely benchmark their results against those of similar providers across the country.

Use of decision support to improve outcomes

Main Points

- **Monitoring trajectory of improvement on every individual case permits clinicians to make adjustment in treatment plan if patient fails to respond as expected**
- **Empirical evidence from outpatient settings supports feasibility of effectively directing resources to those most in need.**
- **Reallocation of resources based on decision rules can result in improved outcomes for highest risk cases without compromising outcomes for less distresses cases.**

There are two broad strategies for creating decision support tools to improve outcomes. One is to employ treatment algorithms designed to increase the likelihood of state of the art treatment and therefore superior outcomes. The Texas Algorithm Project is the best example of this strategy (see Texas appendix).

In a similar vein, Magellan Health Services utilizes a decision support tool with their Treatment Request Form (TRF). The TRF system employs field to field reliability algorithms that are designed to identify inconsistencies within the process of determining a diagnosis and developing a treatment plan. Results of the TRF and its field-to-field algorithms are used to support and guide treatment planning and monitoring of quality of care.

While these methods have the obvious appeal of encouraging best treatment practices, it remains to be seen if use of treatment guidelines and algorithms will result in a measurable improvement in outcomes across a system of care.

The alternative strategy is to monitor patient response to treatment on a session by session basis and use this information to adjust treatment plans as needed. One advantage of this method is that clinicians can see the immediate benefit of using the measures and monitor the response to treatment as the care is provided.

This approach of course requires frequent administration of the outcome measure, particularly in the early phase of treatment when consumers are likely to experience the most benefit but are also a greatest risk for premature termination if they fail to respond quickly. ABHA, using the ALERT system, is the only site in this survey to collect outcome data at such frequent intervals.

The increased burden of frequent data collection cannot be justified unless there is clear evidence the frequent measurement combined with timely decision support tools and

performance feedback can result in measurably improved outcomes and more efficient use of limited resources.

The ALERT decision tools for outpatient settings are intentionally designed to encourage shifting of resources from those least in need (and least likely to benefit) to the consumers with the highest levels of symptoms. It encourages clinicians to retain in treatment those cases at highest risk for a poor outcome and premature termination, while providing fewer services to patients who have already exhibited substantial benefit and are unlikely to improve further. This system-induced shift of resources has the potential to result in improved outcomes while holding steady or decreasing the cost of care. In order to achieve this, the system takes advantage of two findings from the authors' analyses of the data repository.

First, patients entering treatment already within the normal range tend to show no improvement as a group. Furthermore, patients who start treatment and then improve to be within the normal range also tend to show no further improvement once they've crossed the threshold into the normal range. Perhaps this is an "as good as it gets" phenomenon. In the example of the Aggregate Outcomes Report presented earlier, both the "Normals" and the "Severely Distressed" cases averaged between 6 and 5.5 visits per case despite the fact that those in the normal range averaged no improvement, while those in the Severe range experienced substantial benefit. From this data, it is apparent that there may be room to shift resources towards patients with greater levels of distress and increased probability of benefit, without sacrificing outcomes for less disturbed consumers.

The second finding is that patients who experienced a bad outcome (deterioration greater than the RCI on the instrument) tended to show a pattern of early deterioration and premature termination. In fact, if these cases continued in treatment beyond the fifth session, they had a good probability of improvement. However, almost half terminated prior to this point. By tracking the individual trajectory of improvement for each case, and comparing this to the normative expectation for change, the system is able to identify as early as the third session the 10-15% of cases most at risk for a poor outcomes. At the same time it provides information on cases that have done well and are not likely to benefit from continued treatment. In this way, the system seeks to divert the clinicians' time away from the least needy patients and focus instead on preventing premature termination for the at risk cases.

Keep in mind though that this capability comes at a price. It requires fairly frequent assessment at the individual consumer level. The five county Accountable Behavioral Health Alliance (ABHA) that utilizes the ALERT system with its publicly funded consumers in Oregon administers the questionnaire at the first, third, fifth and every subsequent fifth session after that.

The following graphic displays one of the ALERT decision support tools, the Change Index Report. This report provides the clinician with several critical pieces of information: the expected change for each case; the current change and change index for

each case; and, the Predicted Change Index, which is a statistical projection of the most likely outcome if the patient stays in treatment until at least the next data collection point.

Change Index Report

Provider ID: Example

Date of Report: 09/21/99

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

Average results

Bold italics 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 <small>(assumes continued treatment)</small>
<i>Patient A</i>	<i>4/23/99</i>	<i>60</i>	<i>8.12</i>	<i>7/16/99</i>	<i>4</i>	<i>99</i>	<i>-39</i>	<i>-47.1</i>	<i>-20.7</i>
	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 at risk for 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 are both safely within the normal range (using the method Jacobson method formula above). In fact, Patient C appears to be doing so well that the probability is that at the next assessment the patient will have given up some of their improvement points. It is easy to see how this tool can aid a clinician or clinical manager in better allocation of clinical resources.

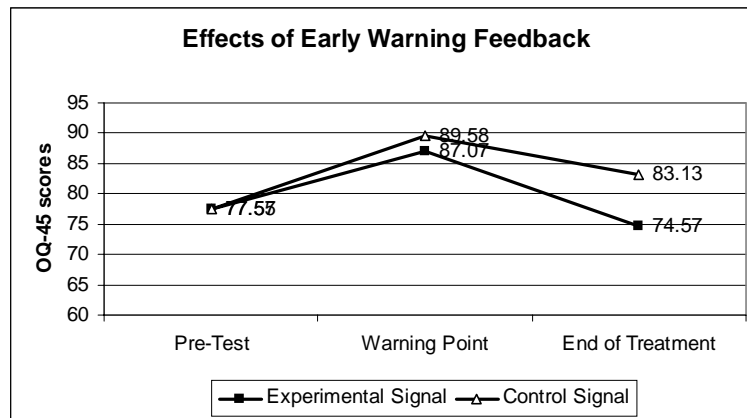
The data clearly indicates the potential to improve results if clinicians are able to act on the information. In order to test this hypothesis, Dr. Lambert and his colleagues at BYU designed a study to look at the effects of providing feedback on the patients' trajectory of

recovery.⁹ The study was conducted at the Comprehensive Counseling Center at Brigham Young University and involved 609 patients and 31 clinicians. The OQ-45 was administered at every session. Treatment continued until termination was deemed appropriate by the clinician and/or patient.

Half of the cases were randomly assigned to the experimental condition in which the clinician had the benefit of feedback on trajectory of change and severity range. The feedback was determined by a set of algorithms that were a function of the number of sessions completed, current level of distress, and assumed likelihood that the patient would fail to recover. In the control condition, the patient completed the questionnaire but the results were withheld from the clinicians.

Space does not permit a complete summary of this study, but for purposes of the present discussion, it is instructive to look at the effects of feedback for the most at-risk patients in the study. Using the algorithm logic, 35 of cases in the experimental condition (11%) and 31 (10%) from the control condition were identified as “signal cases,” those most at risk for premature termination and poor outcome. The feedback message included a warning that the patient was improving less than expected, along with suggestions to review the treatment plan and to guard against premature termination.

In the experimental condition, the feedback resulted in the signal patients receiving almost twice as many sessions of treatment than the control signal cases ($p < .001$). More importantly, the signal cases in the feedback condition showed significantly more improvement post warning than the control group ($p < .05$). The following displays the trajectory of change and outcomes for the two signal groups (higher scores mean greater levels of symptoms).



At the same time the signal cases in the experimental group were receiving additional services, those non-signal cases that were proceeding well in treatment ended up averaging fewer sessions in the experimental group than in the control group ($p < .05$). As our previous discussion would lead us to expect, this occurred without any degradation in

⁹ Lambert MJ, Whipple JL, Smart DW, et al: The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced? To appear in *Psychotherapy Research*, in press

outcomes for these cases. In fact, the non-signal cases in the experimental group averaged more change than the control group, but this difference was non-significant. The additional sessions that are provided to the signal cases was more than offset by the modest reduction in sessions to the non-signal cases, so that the experimental group utilizes overall 4% fewer sessions than the control group.

This study provides support to the premise that it is possible to focus resources and improve results for the most at risk cases without increasing the overall cost of care. Interestingly, one of the questions asked in the post research interviews with the clinicians was “How did you choose to use the feedback you received?” Of the 29 clinicians interviewed, none reported that they increased or decreased the number of sessions given to a client as a result of the feedback. When the results of the study were presented to the staff, there was considerable surprise expressed over the attendance data.

As the Lambert study suggests, this approach may in fact result in a more rational and ethical allocation of outpatient care while having the potential to actually reduce the cost of the care. The approach is rational and ethical in that it utilizes empirical data to shift the focus of care to those most in need and most likely to benefit, while offsetting this cost by encouraging judicious reductions in care for those consumers experiencing normal levels of life distress.

The ALERT system was developed based on data collected largely from commercially insured managed care populations. The case mix model is relatively simple, and geared towards tracking the episodes of care common in this kind of outpatient population. The long-term treatment of SPMI populations, where success might be measured more by stability rather than improvement over a defined period of time, will require other models. Likewise, specialized programs to assist supported living and increase job skills do not fit easily into this type of model.

The use of this system in Oregon provides an opportunity to validate the methodology with a public sector population. Prior analysis of public sector data suggests that the model is appropriate for a substantial percentage of an outpatient treatment population. In fact, the initial results from the ABHA sites support this supposition (see Aggregate Outcome Report in the Oregon/ABHA appendix). The first 96 cases monitored by the system show results as good as the commercially insured population.

While these management methods have been applied to outpatient populations, the same principles could be used in other settings characterized by discrete episode of care in response to crisis or destabilization.

Negative outcomes

Some outcomes indicators are concerned with the frequency of unwanted outcomes, such as adverse reactions to medication, or elopement from an inpatient facility. For children, out of home placement may be an appropriate indicator of an adverse outcome.

Certainly mortality related to mental illness is an important concern. This includes suicide or other consequences of dangerous behavior. One way to approach this is to require reporting all deaths of active patients. Categories for causes of death can be used to distinguish purely medical or accidental causes of death from those in which mental illness or substance abuse is a contributing or sole factor.

Most commercial managed care companies keep track of known suicides and conduct some form of quality assurance review of the care. For obvious reason, managed care companies may not be in a hurry to report suicide rates as one of their critical indicators, but many companies are able to calculate this if they so wish.

Implementation issues

While outcomes management methods may hold promise for improving the outcomes and cost effectiveness of care, collection of data requires cooperation of front line staff already feeling overburdened with paper work. The use of outcomes as a performance indicator may create anxiety among clinicians and administrators. The need for a case mix adjustment model means that the consumer of outcomes reports must have some understanding of the statistical methods utilized. Outcomes management means more work for the clinicians and administrators who must collect the data and interpret the reports. All of these factors translate into financial and psychological cost to the organization.

It is not surprising that a managed care company such as Magellan Health Services would conclude that system wide measurement of outcomes at the individual case level is not feasible across a large system of care. The state systems that have committed to outcomes assessment have not implemented the kind of frequently repeated instrument administration that a trajectory prediction approach to outcomes management requires.

PBH has been able to implement the ALERT system is due to the fact that there was a clear mandate from key members of senior management to make outcomes management a reality. The unified vision of senior decision-makers was necessary to overcome the natural resistance to such an intensive program. This sort of top down management mandate may work in the context of a for-profit managed care entity and its contracted providers, but prove problematic if attempted in other systems of care.

The experience of ABHA in Oregon may prove instructive. Data collection began in the last quarter of 1999. The initial rate of submissions of the questionnaire forms was much lower than expected. To counter this, reports of the number of submissions by county were generated weekly. The submission rate is now increasing, but the percentage of cases with two usable data points is still far below that of PBH's commercial providers (35% versus 16%). These results are preliminary, and its too early to judge the success of the implementation. Success here, as elsewhere, will depend upon the perseverance and patience of key decision-makers over a sufficient period of time to increase compliance with the assessment protocols and demonstrate the usefulness of the data to front line clinicians.

Structure/Plan Management

Main Points

- **Domain includes indicators of responsible financial management and consumer involvement in policy development and quality assurance.**
- **Primary financial indicator is medical loss ratio, i.e. percentage of revenues going directly to patient care.**
- **Cost benefit ratio, i.e. magnitude of improvement/cost of care.**

This domain as envisioned by the NASMHPD indicators has two basic components: 1) consumer involvement in policy development, quality assurance and planning and 2) indicators of responsible financial management.

Consumer involvement in policy, quality assurance and planning is reported as the percentage of policy making boards on planning committees made up by self-identified consumers or families of consumers (number of consumers/total number on board or committee).

The NASMHPD framework describes two financial indicators: medical loss ratio and average resources expended on mental health services. The medical loss ratio is the percentage of revenues going directly to patient care as opposed to administrative costs. Since this is a basic accounting function, most systems are able to report this from the administrative systems. Variations in accounting procedures may make comparisons across states questionable. This review found no external benchmarks for state systems. In commercial health plans a medical loss ratio of approximately 85% is common.

Average resources expended on mental health services can be reported both as cost per enrollee (annual direct services expenditures/total population of covered lives) and as cost per service recipient (total annual cost/number of persons receiving at least one mental health service). This should be broken down further into the following groups:

- Children and adolescents with serious emotional disturbances.
- Children and adolescents with other emotional disturbance.
- Adults with serious mental illnesses.
- Adults with other mental illnesses.
- Adults with dual diagnoses of a mental illness and substance abuse

Commercially managed care plans commonly use the average time from request to face-to-face visit and the proportion of persons that were denied services as indicators of overall plan management. This reflects the employers' wish to provide ready access to services while minimizing the need to utilize the ultimate cost control weapon, denial of requested service (usually inpatient care). This is consistent with sound clinical practice of effective intake and triage, while engaging patients in the recommended course of treatment. Effective clinical management greatly minimizes the need for denial of requested services. At present, a few states report this. New York, for example, reports numbers of denial by services category, as well as percentage of successful appeals of denials. Still, most do not have the capability to track and report these statistics. As managed care models for the public sector become more sophisticated and widespread this is likely to change.

Another approach to evaluation planning is to examine the efficiency or "value" of the services. This is done by calculating a cost/benefit ratio. In its simplest form, this can be expressed as:

$$\text{Value} = \text{Magnitude of improvement} \div \text{cost of service.}$$

This is most easily measured for outpatient services with relatively discrete episodes of care, so the statistics such as "change per session" can be estimated. As the previous discussion regarding patterns of improvement and severity of distress showed, a reallocation of services away from the healthiest consumers towards those showing the greatest level of distress and impairment can result in significant improvement per session or unit of cost. The fact that allocation of resources factors into the equation means that plan management has more direct opportunity to increase efficiency across a large system of care than to improve outcomes per se.

Despite the intuitive appeal of a "value indicator," none of the sites reviewed currently report this statistic. There are enough complexities already in reporting case mix adjusted outcomes without trying to tie in cost of service to the model. However, as outcomes measurement becomes more routine value indicators may yet prove to be useful in evaluating overall plan management.

Prevention/Early Detection

Main Points

- **Domain not yet well defined, indicators are under development**
- **3 levels of prevention: primary, secondary and tertiary**
- **Consumer surveys used to assess consumers awareness of and participation in psycho-educational or self-help program (secondary or tertiary prevention)**

The NASMHPD performance measures left this domain empty in their first set of indicators, though they have several indicators under development. Prevention efforts can be conceptualized as falling into three levels of intensity and specificity.

1. **Primary prevention:** These are broad, population based efforts, usually educational in nature. An example might be an insert on recognizing the signs of depression mailed out to health plan members.
2. **Secondary prevention:** These are efforts to target and intervene early with at risk individuals through proactive psycho-educational programs, early treatment referral, and other program interventions.
3. **Tertiary prevention:** These are services provided to individuals with an already existing disorder designed to prevent further deterioration and enhance patient stability and treatment compliance.

Some sites have conducted various prevention and psycho-educational programs and have used survey methods to assess consumer perception of the programs. Use of self-help programs is also encouraged as a form of tertiary prevention in that they increase the consumer's ability to self-manage symptoms (i.e. stay "in recovery"). New York, for example, reports the percentage of consumers reporting participation in some form of self-help group.

Other possible indicators involve use of screening tools or other methods designed to identify high-risk populations as part of a secondary prevention strategy. However, these indicators are still under development.

Data Capture and Analysis

Systematic assessment of outcomes can generate large quantities of data (or at least paper forms). This section discusses briefly the successful strategies for data capture and analysis.

The most common mechanism for data capture is a scannable form. Teleform¹⁰ was clearly the first choice as the software application for this purpose. All of the sites in this survey employing scannable forms technology used Teleform for form creation and optical mark/character recognition tasks.

Some sites have employed data entry templates for a PC, and even handheld devices (Ohio). However, it is easy to understand the attraction of scannable forms. The paper/pencil user interface is intuitively obvious and requires no special training. The software is relatively inexpensive and the process lends itself to centralization and high volume. A single worker at a PC can process hundreds of forms a day. The centralization of data capture also facilitates processes to ensure data accuracy and integrity.

Some systems have developed relatively sophisticated databases for warehousing and reporting outcomes data (Access Measurement Systems, the PBH ALERT system), while others are accumulated data files for analysis on an ad hoc basis. The two most common applications for analysis and reporting are SAS¹¹ and SPSS¹². SAS has also been used as the “backbone” for complex clinical information systems (e.g. PBH, Illinois) while interfacing its output to Windows based applications for viewing data or reports.

AMS has done an impressive job of capturing data and generating a variety of reports to its customers. AMS has utilized Teleform for data capture from faxed forms, but is developing their own application for this purpose. Clinical and outcome reports are programmed using Visual Basic. The company boasts that its system generates over 10,000 reports monthly.

¹⁰ Teleform™ distributed by Cardiff Software, facilitates the design and printing of forms for optical character and mark recognition. Teleform can recognize it's own forms from either faxed or scanned images, and performs the data capture in an accurate and efficient manner.

¹¹ SAS™ is a versatile and comprehensive program for performing data analysis, managing data, creating data warehouses, and other data processing operations. It is the product of the SAS Institute in Cary, N.C. and is widely used throughout government and industry when an “industrial strength” tool is needed.

¹² *Need info on SPSS.*

Making Use of the Information

All the effort put into collecting and analyzing performance data is of little use if it does not result in measurable improvements in the quality and outcome of care. In order to achieve this, the data must be transformed to information and delivered to the right people at the right time to affect change.

The recipient needs to perceive that the information presented is easily understood and relevant to performing his or her job well. . Tables of numbers do not achieve this. A well designed graph or brief report with critical information highlighted does.

It is necessary to remember that different stakeholders have different reasons for wanting the information. The information that a state legislature needs to perform its oversight function is clearly different from what a hospital administrator needs to improve care at that facility. With this in mind, this section will suggest a framework for organizing what and how often information is provided to various stakeholders.

In order to make a difference care delivery, the information must be actionable. That is to say, the person receiving the information needs to be in a position to act upon it. Providing more information than can be acted upon can tend to obscure in the key information that is useful. Therefore, a performance management system needs to focus the information, provided the right amount at the right time to the right people. The following suggestions have been gleaned from the interviews with informants at the various sites studied.

Legislators and funding sources

The primary interest of the legislature and other funding sources is have assurance that money allocated is well spent and achieving the stated objectives of the mental health system. The funding sources also have a legitimate interest in seeing evidence that the mental health system is taking steps to systematically improve quality and outcomes, achieving great value for the money allocated.

An annual cycle of reporting is adequate for the funding source. The legislature does not need to information more often because the primary action that needs to result from the information is the decision to renew funding. In fact, an annual report may be too frequent. Arizona for examples only conducts a consumer satisfaction survey every other year.

The indicators at this level are likely to be more global in nature, with more of a focus on financial performance and impact of the services on broader societal concerns. Most legislators probably to not have the time or interest to delve into the finer points of specific indicators for certain sub-populations.

The legislature does have an interest in knowing the system is well managed. Therefore, the legislature should expect to see evidence of improvement over time on key indicators, but it does not need all of the information necessary to drive that improvement.

System Administrators

System administrators (state level or other large geographic area) are responsible to the funding sources to assure that money is well spent in the interest of the consumers served. To accomplish this, they need information more frequently, and they need to be broken out by operational units. It isn't sufficient to know the performance on an indicator across the entire system. The administrator needs to know which facilities or clinics are producing superior results and which have room for improvement.

The reporting cycle needs to occur more frequently than for the legislature so that there are multiple opportunities to make any needed changes over the course of a year. For this reason, system administrators need information on system performance on at least a quarterly basis.

While regular reports may present relatively global measures, the administrator needs to be able to drill down into the data as needed to understand how to best act on the information. The administrator needs the ability to ask for more detailed analysis and to get the result in a timely manner.

Clinic or facility administrator

The clinic or facility administrator has responsibility for all aspects of the care provided at that site. At this level of decision making, the information needs to be timely and detailed. The administrator needs to know both how the facility is functioning at any point in time; where the problems are and what is working well. For the administrator then the information may need to be broken out by sub-populations, and by individual staff members.

The information needs to be frequent and timely. An unacceptably long wait to first appointment needs to be addressed that month. If the administrator is responsible for a quarterly report on consumer satisfaction, then it would be wise to sample at least monthly if the site has a history of low satisfaction. An annual survey is not likely to improve satisfaction, but a monthly survey with immediate feedback to the clinic staff may well do the trick.

Clinical Staff

Clinical staff is focused on treatment patients. Issues of performance and outcomes at a macro level are of little interest to them except as it impacts job stability perhaps. They are the ones most likely to bear the burden of collecting the data while perceiving the least benefit from the activity.

Clinicians do care about patients however and will work to achieve the best results they can for any particular case. They need then to perceive that information is directly relevant to their job as a clinician.

Knowing the overall outcomes for their clinic or even for their own caseload may be of great interest but it does not necessarily help them do their job better. Knowing which of their cases is getting worse and reporting a high frequency of suicidal ideation does. For the clinician then, information of this type needs to come as quickly as possible and often as needed. And at the end of the day, if the individual clinician has done a better job for a specific patient, the entire system of care has performed just a little bit better.

Summary and Conclusions

Main Points

- **Domains of Access, Appropriateness and Outcomes are relatively well defined and form the basis of a performance measurement system.**
- **Indicators for Domains of Structure/Plan Management and Prevention/Early are under development and not yet widely implemented.**
- **MHSIP Consumer Survey has become an “industry standard” and should be utilized. Most sites have chosen to use shorter versions in the 20-25 item range.**
- **Consumer satisfaction with outcome is a poor indicator of actual change.**
- **To assess change, repeated measures are necessary.**
- **Outcomes monitoring using several repeated measures in the first few weeks of outpatient care has the potential to significantly improve the allocation of care and result in improved overall outcomes.**
- **Case mix adjustment models permit use of external benchmarks for performance.**
- **Performance feedback loops that give the right information to the right person at the right time hold the greatest potential to improve care.**

This report seeks to provide a useful overview of the current state of the art in performance and outcomes measurement for mental health care. The reader is referred to documents such as the NASMHPD Performance Measures and the MHSIP Consumer Oriented Mental Health Report Card for more detail on specific indicators.

The reader is also referred to the recent report of the workgroup for Decision Support 2000+. This group has been tasked by the Center for Mental Health Services to design an information system for behavioral health care. The full copy of this report can be found at <http://mhsip.org>. The introduction and section summaries of the report are included in this binder.

The NASMHPD indicators and the associated ORYX approved NRI system for reporting to JCAHO have made strong inroads among states, with a majority of states using the system at this time. For this reason the NASMHPD indicators are the logical place to begin tailoring a performance measurement system to meet the state’s particular needs.

Indicators of Access, Appropriateness and Outcomes that can be measured through a survey method have been widely implemented, as evidenced by the MHSIP Consumer Survey. In fact, the MHSIP survey can be regarded as the industry standard. There is no compelling reason to use alternative survey, though most states seem to prefer one of the shorter versions rather than the full 40 item survey.

Instruments like the MHSIP survey provide measure of consumer satisfaction with the services and system of care. The instrument also provides a sense of the consumer's retrospective perception of benefit from the services. However, a retrospective survey is an inexact measure of outcome at best.

Real measurement of clinical outcomes requires repeated administrations of an outcome instrument. This requires a substantial commitment of energy and resources by the organization. While the actual cost of capturing the data may not be that great, the organizational work necessary to secure the cooperation of providers and consumers in the use of outcomes measures may require significant effort, focus and leadership. The organizational culture may not immediately embrace the change in business as usual that this will require. Issues of compliance with data collection protocols are significant, and can easily threaten to undermine the validity and utility of the data collected.

An outcome assessment method that makes use of repeated measures at relatively frequent intervals offers the greatest potential for managing outcomes, particularly for outpatient services. However, this is the most difficult to implement systematically and is likely to be met with clinician resistance. Nevertheless, the potential for more effective allocation of clinical resources and improved care for at risk cases may justify the added burden.

Case mix adjustment models are still relatively new, and in most cases sites are just beginning to develop models. The limiting factor is the size of the database. Any choice of instruments should give consideration to the availability of a large enough normative sample to develop case mix adjustment models. Case mix adjustment formulas predicting expected outcome can serve as external benchmarks for outcomes performance.

The domain of Structure/Plan Management is evolving. States have long experience involving consumers in policy and planning activities. Managed care influence is increasing the use of indicators like wait time till first appointment and tracking of denial of services and associated appeals.

Primary Prevention/Early Identification indicators are the least developed and utilized. Survey questions regarding use of self-help programs and awareness of prevention activities are relatively easy to implement. Screening tools and other methods of early identification and not widely employed and reported at this time.

Overall, the importance of measuring and reporting performance and outcomes indicators has expanded greatly over a relatively short period of time. It is too early to evaluate fully

the impact of this movement, there is little doubt that a new era of accountability for outcomes and quality has arrived.

The question remains will this information be used to simply give us a sense of how the system is performing, or will the performance information be used to drive real change that results in improved care for the consumer? This will be achieved by taking the data seriously and providing information in the right quantity to the right people at the right time to enable clinicians and administrators to take concrete steps to improve outcomes.