

# How Can You Improve Quality Without Measuring Outcome? *Getting From Here to There*

Mark Zimmerman, MD, Diane Young, PhD, Iwona Chelminski, PhD, and Kristy Dalrymple, PhD

## ABSTRACT

*To improve quality it is necessary to measure outcome. Psychiatry is the only medical discipline in which quantified measurements of outcome are not the standard of care. There are several potential obstacles to measuring outcome in routine clinical practice, but assessment methods have recently been developed to overcome these obstacles. One such obstacle is the determination of outcome across diagnostically heterogeneous groups of patients. The concept of psychiatric vital signs is described. The vast majority of a large, diagnostically heterogeneous group of psychiatric outpatients had clinically significant anxiety or depression upon presentation for treatment. This supports the routine assessment of anxiety and depression in clinical practice because almost all patients will have these problems as part of their initial presentation. Even for those patients without depression or anxiety, the case could be made that the measurement of depression and anxiety is relevant and is analogous to measuring certain physiological statistics in medical practice such as blood pressure and body temperature regardless of the reason for the visit. Another obstacle towards utilizing measurement to evaluate quality is the difficulty in aggregating data across large numbers of patients. A Web-based system of outcome assessment is described that allows response and remission rates to be calculated in a group of patients. The tools and technology now exist to overcome the challenges posed by a measurement-based care approach towards care, and this will hopefully accelerate the incorporation of measurement into routine clinical practice and enable quality improvement efforts to be tested in a cost-effective manner.*

## FOCUS POINTS

- Quality improvement efforts in psychiatry will require a paradigm shift toward the incorporation of quantifiable outcome assessments in clinical practice.
- Six potential obstacles toward measurement-based care are: patient acceptability, clinician acceptability, clinical utility of the measure, cost, diagnostic heterogeneity, and data aggregation.
- To address the heterogeneity problem, the concept of psychiatric vital signs is introduced. It is recommended that depression and anxiety be routinely measured, analogous to the routine assessment of body temperature and blood pressure in medical settings.
- A Web-based system for outcome assessment administration is described which allows for the aggregation of outcome information across a group of patients. The reliability and validity of Web-based administration is described.
- Several clinical situations are described illustrating how measurement-based care could improve outcome.

## INTRODUCTION

The *U.S. News and World Report*<sup>1</sup> annually publishes a list of the top programs in psychiatry (and other medical specialties). Noteworthy in the accompanying article is the absence of a discussion of data describing treatment response rates. Are psychiatric patients treated in programs at the top of this list more likely to achieve better outcomes than patients treated in programs lower on the list, and do patients who are treated in programs that are not on the list have even poorer outcome? Perhaps in some medical specialties data exist demonstrating that a positive outcome is more likely in some

Dr. Zimmerman is associate professor of Psychiatry and Human Behavior at Brown University Medical School in Providence, Rhode Island. Dr. Young is staff psychologist in the Department of Psychiatry at Rhode Island Hospital and assistant professor of Research in the Department of Psychiatry at Alpert Medical School at Brown University. Dr. Chelminski is clinical assistant professor in the Department of Psychiatry at Rhode Island Hospital. Dr. Dalrymple is staff psychologist in the Department of Psychiatry at Rhode Island Hospital and assistant professor of Research in the Department of Psychiatry at Alpert Medical School.

Disclosures: Dr. Zimmerman is officer/director of Psych Products Press. Drs. Young, Chelminski, and Dalrymple report no affiliation with or financial interest in any organization that may pose a conflict of interest.

Please direct all correspondence to: Mark Zimmerman, MD, Bayside Medical Center, 235 Plain St, Providence, RI 02905; Tel: 401-444-7098; Fax: 401-444-7107; E-mail: mzimmerman@lifespan.org.

medical centers than others, but the authors of this article are not aware of such data in psychiatry.

Considering the issue of outcome more broadly, the field of psychiatry is only beginning to ask fundamental questions regarding the effectiveness of currently available treatments in real-world clinical practice. How well do they work? For whom do they work best, and for whom are they ineffective? How many patients are and are not receiving evidence-based care, and is the provision of evidence-based care associated with better outcome? Are some clinicians more effective than others, and, assuming there are differences in patient outcomes between clinicians, is it possible to improve outcomes in patients treated by those clinicians who perform below average? Each of these and other questions can only be addressed if outcome is measured, and measurement is incorporated into routine clinical practice.

During the past 15 years the authors' clinical research group has been conducting the Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS) project.<sup>2,3</sup> As suggested by its name, the overarching goal of the MIDAS project has been to improve the quality of care. To achieve this goal the core assumption was that it was not possible to improve quality unless outcome was measured. Therefore, scales were developed that were feasible to use in clinical practice.<sup>4,6</sup> The first part of our mission was to develop methods of measuring outcome that could be readily adopted by clinicians working outside of this program. This article describes the practical issues confronted and addressed in establishing a behavioral health outcomes program.

## POTENTIAL OBSTACLES IN MEASURING OUTCOME

Table 1 lists six obstacles to be overcome in the implementation of outcomes assessment in clinical practice:

**Patient acceptability.** If measurement is overly burdensome to patients, then they may be dissatisfied with their care and either drop out from treatment or seek care elsewhere.

**Clinician acceptability.** If measurement interferes with a clinician's usual work flow, then it is less likely to be adopted.

**Clinical utility.** Measurement that is clinically meaningful and improves the efficiency of conducting clinical evaluations will be more likely adopted than tools that do not inform clinical decision making.

**Cost.** Instruments that have higher acquisition costs, or support staff costs, are less likely to be utilized.

**Diagnostic heterogeneity.** Most clinicians treat patients with a variety of diagnoses. Measurement that is diagnostic specific will only be appropriate for a limited number of patients; therefore, scales must be used that apply to patients with a variety of diagnoses.

**Data aggregation.** If the goal of measuring outcome is to improve quality, then it will be necessary to aggregate information on outcome across patients in order to determine the impact of a change in service delivery.

Each of these obstacles is discussed below.

## THE HETEROGENEITY PROBLEM AND THE CONCEPT OF PSYCHIATRIC VITAL SIGNS

In describing the measurement-based care approach towards treatment used in the Sequenced Treatment Alternatives to Relieve Depression (STAR\*D) study, Trivedi and colleagues<sup>7</sup> limited their focus to depressed patients. However, most psychiatrists treat heterogeneous groups of patients with a variety of diagnoses. The *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition,<sup>8</sup> lists >400 diagnoses in 17 diagnostic categories. Diagnostic heterogeneity poses a challenge towards the adoption of measurement-based care in clinical practice.

There are three possible approaches towards addressing the heterogeneity issue in measuring outcome in clinical practice. One approach is to use global ratings that would be appropriate for all patients regardless of diagnosis. The measure can reflect changes in clinical status such as the Clinical Global Index (CGI) of Improvement,<sup>9</sup> which is an overall assessment of how much a patient has improved or deteriorated during treatment. Alternatively, the measure can be a diagnostically nonspecific rating of symptoms or functioning such as the Global Assessment of Functioning (GAF).<sup>8</sup> A limitation with this type of approach, though, is that these types of scales do not provide information about specific symptoms, upon which treatment decisions are often based. In contrast to the use of global measures, a second approach would be to use diagnosis-specific symptom rating scales. Scales have been developed and validated for most of the common major psychiatric disorders, and the recently updated *Handbook of Psychiatric Measures*<sup>10</sup> is a compendium of many of these instruments. While the ideal approach towards monitoring outcome would include disorder-specific measurement, many

TABLE 1

### POTENTIAL OBSTACLES IN IMPLEMENTING AN OUTCOMES ASSESSMENT PROGRAM IN CLINICAL PRACTICE

Patient acceptability
Clinician acceptability
Clinical utility
Cost
Diagnostic heterogeneity
Data aggregation

Zimmerman M, Young D, Chelminski I, Dalrymple K. *Primary Psychiatry*. Vol 17, No 7. 2010.

scales are time consuming to administer, and the authors of this article are skeptical that they will be used by many clinicians in routine clinical practice. A third strategy, which the authors are proposing herein, is the routine measurement of a couple of central and frequently occurring psychiatric constructs, analogous to the routine measurement of medical vital signs. These are therefore referred to as psychiatric vital signs. Specifically, the authors suggest that the measurement of depression and anxiety be considered psychiatric vital signs.

Medical vital signs are measures of basic physiologic functions that are routinely determined in medical settings. Vital signs are so named because they reflect essential indicators of life. It is not being suggested that the measurement of depression and anxiety is analogous in this sense. Rather, it is the routine assessment of body temperature, blood pressure, heart rate, and respiration rate at the beginning of the office visit, regardless of the patient's chief complaint, that is the basis of using the term "vital signs."

Medical vital signs are, at times, a primary outcome measure (eg, blood pressure in patients treated for hypertension), and, at other times, an adjunctive measurement. Likewise, the measurement of depression and anxiety might be either a primary outcome or an adjunctive measurement. However, because mood and anxiety disorders are the most common psychiatric disorders, the routine assessment of depression and anxiety could be considered core elements of the clinical encounter that provide clinically meaningful and quantifiable information on the effectiveness of treating most patients with psychiatric disorders. Of course, the measurement of anxiety and depression would not preclude the use of additional disorder-specific assessments. However, considering the low frequency with which standardized scales are currently used by

psychiatrists to measure outcome, the authors believe that the adoption of a couple of outcome indicators would move forward efforts towards measurement-based mental health care.

As part of the MIDAS project, the authors examined the frequency of depression and anxiety in 3,000 psychiatric outpatients interviewed with the Structured Clinical Interview for *DSM-IV* (SCID).<sup>11</sup> The SCID was supplemented with symptom items from the Schedule for Affective Disorders and Schizophrenia (SADS),<sup>12</sup> two of which were depressed mood and psychic anxiety. Based on the SADS ratings, 79.3% (n=2,378) of the patients reported clinically significant depression of at least mild severity, 64.4% (n=1,932) reported anxiety of at least mild severity, and 87.4% (n=2,621) reported either anxiety or depression. Data in Table 2 show that the high frequency of anxiety and depression cut across all diagnostic classes. Thus, these findings support the routine assessment of anxiety and depression in outpatient clinical practice because almost all patients will have these problems as part of their presenting complaints. Even for those patients without depression or anxiety, the case could be made that the measurement of depression and anxiety is relevant and is analogous to measuring certain physiologic parameters in medical practice such as blood pressure and body temperature regardless of the reason for the visit.

A vital signs approach towards measurement not only refers to what to measure, but also when to measure. In outpatient settings, two approaches can be considered regarding the frequency and timing of outcome assessments—fixed time points versus every visit. In the fixed time points approach, assessment occurs at predefined time points such as baseline and every 3 months, or baseline and the end of treatment (ie, pre-treatment and post-treatment). A practical problem with this approach is that

TABLE 2

### FREQUENCY OF CLINICALLY SIGNIFICANT DEPRESSION AND ANXIETY IN 3,000 PSYCHIATRIC OUTPATIENTS WITH PRINCIPAL DIAGNOSES IN DIFFERENT DIAGNOSTIC CATEGORIES

<i>Diagnostic Category</i>	<i>n</i>	<i>Depression % (n)</i>	<i>Anxiety % (n)</i>	<i>Depression or Anxiety % (n)</i>
Mood disorder	1,702	90.9 (1,547)	66.4 (1,130)	93.9 (1,598)
Anxiety disorder	546	71.2 (389)	77.5 (423)	89.2 (487)
Substance use disorder	56	53.6 (30)	46.4 (26)	71.4 (40)
Psychotic disorder	51	54.9 (28)	62.7 (32)	74.5 (38)
Somatoform disorder	65	70.8 (46)	66.2 (43)	81.5 (53)
Impulse control disorder	53	64.2 (34)	50.9 (27)	73.6 (39)
Eating disorder	20	75.0 (15)	55.0 (11)	85.0 (17)
Adjustment disorder	180	70.0 (126)	53.9 (97)	83.3 (150)
Attention-deficit disorder	77	41.6 (32)	40.3 (31)	54.5 (42)
Personality disorder	92	72.8 (67)	53.3 (49)	79.3 (73)

Zimmerman M, Young D, Chelminski I, Dalrymple K. *Primary Psychiatry*. Vol 17, No 7. 2010.

patients often leave treatment without advanced warning, thereby resulting in the loss of important information about clinical status at treatment termination. Another practical problem is the need to keep track of when the follow-up assessments are to be conducted. Moreover, this approach implicitly suggests that the determination of the patient's status is not an essential component of all treatment visits. A recommendation that the measurement of outcome be evaluated only at certain time points implies that such measurement is burdensome, and to minimize this burden it need not be evaluated routinely. The perceived burdensomeness of outcome measurement will impede its widespread adoption. On the other hand, according to the "every visit" or "vital signs approach," assessment is conducted at all follow-up visits and is considered integral to treatment. To be sure, clinicians already routinely evaluate outcome at follow-up visits, albeit in an unstandardized, nonquantitative, manner. Because of the availability of brief, albeit reliable and valid, scales to measure depression and anxiety, it is as feasible to incorporate assessments of these constructs into a busy clinician's work flow as it is to incorporate the measurement of medical vital signs.

## ACCEPTABILITY OF OUTCOME MEASUREMENT

There is no shortage of measures to be used to monitor outcome. Two perspectives are of primary importance in deciding which measure to choose—that of the patient and clinician. Patients should find the measures user-friendly and the directions easy to follow. The questions should be understandable and relevant to the patient's problem. The scales should be brief, taking no more than 2–3 minutes to complete, so that upon routine administration at follow-up visits patients are not inconvenienced by the need to come for their appointment 10–15 minutes early in order to complete the measure.

The instrument should provide clinicians with clinically useful information, and improve the efficiency of conducting their clinical evaluation; thus, the measure should have practical value to the practicing clinician. Of course, clinicians need to be able to trust the information provided by any instrument they use. Consequently, outcome measures should have a sound basis in psychometrics, demonstrating good reliability, validity, and sensitivity to change. Clinicians and clinics should also find the instrument user-friendly; it should be easy to administer and score with minimal training.

The authors' group<sup>13</sup> examined the feasibility and acceptability of using a self-administered depression questionnaire to measure outcome in routine clinical practice in two studies of depressed psychiatric outpatients who were in ongoing treatment. The patients completed a questionnaire assessing how burdensome it was to complete the scale during the visit (0=very little burden; 3=a large burden), and their willingness to complete the scale at

every visit to help monitor the progress of their treatment (0=not at all willing; 3=very willing to fill it out at every visit). Almost all patients considered questionnaire completion very little or only a little burdensome (98%, n=49), and no patient perceived it as very burdensome. More than 90% of patients indicated a willingness to complete the scale at every visit in the future if their clinician believed that it was helpful (94%, n=47).

Of course, acceptability and feasibility may vary by scale. Some measures consist of  $\geq 100$  statements, whereas others contain  $< 10$  items. The study summarized above was of a scale the authors developed in the MIDAS project—the Clinically Useful Depression Outcome Scale (CUDOS),<sup>5,14</sup> a brief measure that asks respondents to rate 16 symptom items on a 5-point Likert scale. In the second study of feasibility,<sup>15</sup> a separate sample of 50 depressed outpatients completed both the CUDOS and the Beck Depression Inventory (BDI) during a follow-up visit. In contrast to the CUDOS, the BDI has respondents read groups of four statements and select the item that best describes how they had been feeling during the preceding week. Thus, the BDI includes more information to read than the CUDOS. After completing the two questionnaires the patients completed a questionnaire asking which of the two measures took less time to complete, was easier to understand, was less burdensome to complete, and was more acceptable to complete at every follow-up appointment. Significantly more patients indicated that the CUDOS took less time to complete than the BDI (64% vs. 12%,  $P < .05$ ), and was less of a burden to complete (50% vs. 10%,  $P < .05$ ). Nearly three times as many patients indicated that they would prefer to complete the CUDOS than the BDI to monitor the outcome of treatment (40% vs. 14%,  $P < .05$ ). These studies suggest that patients did not find scale completion burdensome, especially when the scale is brief. In addition to the CUDOS, another brief measure of depression that is feasible to incorporate into routine clinical practice is the 9-item Patient Health Questionnaire.<sup>16</sup>

## THE DATA AGGREGATION PROBLEM

This article began with the question of whether patients are more likely to achieve a positive outcome in medical centers ranked highly in the *U.S. News and World Report* annual survey. To answer this question it is necessary to not only measure outcome, but also to collect the data across patients and across clinicians. Above, the authors referred to this as the data aggregation problem. While data aggregation is not necessary to realize the clinical benefit of measurement-based care in the treatment of individual patients, data aggregation is necessary for quality improvement efforts in which outcome across a case load is compared before and after a change in service delivery. Data aggregation could be labor intensive, and thus costly. To address the potential obstacle posed by the cost of data entry, the authors developed a Web-based system to administer and score

the administration of the CUDOS and aggregate data across a clinician's caseload to determine response and remission rates at the end of the acute phase of treatment.

## WEB-BASED ASSESSMENT

A Web-based platform for the administration of outcome assessments offers several advantages over paper-and-pencil assessments (Table 3). Web-based scales can be completed by patients at their convenience in their home rather than arriving early or staying after their clinical appointment to complete the measure. A computer-administered survey can prompt respondents to ensure all questions are answered, thereby reducing missing data. The administrative costs associated with the copying, handing out, and scoring of paper questionnaires are reduced with a Web-based system. Similarly, the high costs of establishing and maintaining a data base to evaluate treatment outcome for a large sample of patients based on administration, scoring, and data entry of paper questionnaires could be markedly reduced with a Web-based system. Moreover, because the data collected via the Internet is automatically entered into a database, data entry errors are reduced.<sup>17</sup>

It cannot be assumed that paper- and computer-based administrations of the same test produce equivalent results.<sup>18-20</sup> Despite equivalent content, differences in format can influence results.<sup>21,22</sup> For example, some computer-administered tests present items one at a time on a screen whereas paper tests present multiple items on a page. Presentation of items singly versus as a group can result in differential attention to the items and thus result in different item-scale correlations. Some computer-administered tests do not allow a return to previous items, whereas paper tests allow prior responses to be changed. This could produce respondent frustration with computer administration, thereby influencing responses to the remaining items on the scale. Also, respondents might reconsider prior answers after answering subsequent questions. Some computer-administered scales do not allow missing answers. The "candor hypothesis" suggests that

respondents are more truthful when responding to computer-administered tests. Each of these factors could influence the psychometric properties of a test, and the cutoff values used to determine caseness (or remission status), thereby warranting the demonstration of equivalence.<sup>18,19</sup>

Internet administration of scales adds another potential source of error variance because patients might complete questionnaires while watching television, talking on the telephone, eating, or participating in another distracting activity.<sup>23</sup> The nonstandardized conditions under which scales might be completed on the Internet can alter responses, thereby further requiring demonstration of equivalence to in-office administration.<sup>17</sup>

Despite the potential sources of error, studies comparing paper and computer or Internet scale administration have found high correspondence between the two assessment methods.<sup>24-27</sup> However, few studies examined Web-based scale administration, and the authors are unaware of a single study of psychiatric patients in real world ongoing treatment. Thus, caution should remain before extrapolating the results of paper-and-pencil scale administration to Web-based administration, and the prevailing recommendation of demonstrating equivalence should continue. As part of the MIDAS project, the authors have conducted the first study comparing paper and Internet administration of a depression scale to depressed psychiatric outpatients in ongoing treatment.

Preliminary findings of an ongoing study of the comparability of paper and Web versions of the CUDOS are presented here. Forty-eight hours before their appointment, 30 depressed patients in ongoing outpatient treatment received an automatically generated e-mail reminding them of their forthcoming appointment as per the system's appointment reminder function, and directed them to complete the Web-administered version of the CUDOS (CUDOS-W). At the end of their appointment, the patients were asked if they would complete the paper version of the CUDOS. It was explained to them that the circumstances and setting of scale completion sometimes influences responses to a scale, therefore it was important to examine the comparability of computer and paper administrations of a scale. At the visit the clinician completed the Montgomery-Asberg Depression Rating Scale (MADRS)<sup>28</sup> and rated the GAF and CGI Severity (CGI-S).<sup>29</sup>

The formats of the paper and Internet versions of the CUDOS were identical. The Web administration presents the entire scale at once, rather than one item at a time. Patients are able to change their responses after answering a question. The questionnaire cannot be submitted unless all items are completed.

The sample included six (20%) men and 24 (80%) women who ranged in age from 22–85 years ( $M=47.0$ ,  $SD=12.2$ ). The mean score on the MADRS (11.9,  $SD=12.3$ ) and CGI-S (1.2,  $SD=1.3$ ) indicated a mild level of depression severity. The mean score on the GAF was 66.1 ( $SD=10.2$ ).

TABLE 3

### POTENTIAL ADVANTAGES OF WEB-BASED ASSESSMENT

Patient convenience
Reduced missing data
Reduced costs
Automatic scoring
Automatic database generation
Aggregation of data across patients

Zimmerman M, Young D, Chelminski I, Dalrymple K. *Primary Psychiatry*. Vol 17, No 7. 2010.

The average interval between the completion of the paper and Internet versions of the scale was 1.1 days (SD=0.8). The correlation between the CUDOS and CUDOS-W was high (ICC=.95,  $P<.001$ ). The mean scores were similar on the paper and Internet administrations (18.9±14.5 vs. 18.5±13.8, paired  $t=0.4$ , n.s.). In previous validation studies of the paper version of the CUDOS, the authors found a cutoff of 20 identified patients who were in remission. Based on this cutoff score there was 94.7% agreement between the paper and Internet administrations in determining patients' remission status ( $k=.90$ ).

The internal consistency of the paper and Internet administrations of the CUDOS was high (Cronbach's  $\alpha=.94$  and .93, respectively). For each item the correlation between the CUDOS and CUDOS-W was significant (median=.86).

Both the paper and Internet versions of the CUDOS were significantly correlated with the MADRS ( $r=.92$  and .89, respectively), CGI-S ( $r=.94$  and .93, respectively), and GAF ( $r=.94$  and .90, respectively). None of the differences in the correlations between the paper and Internet administrations and the validity scales were significant.

The preliminary results of this ongoing study support the reliability and validity of Internet administration of the CUDOS. Internal consistency, item-scale correlations, and correlations with external validators were as high with Internet administration as with paper administration of the scale. The Web site, which includes both the CUDOS, and the authors' Clinically Useful Anxiety Outcome Scale,<sup>4-6</sup> can be found at [www.outcometracker.org](http://www.outcometracker.org)<sup>30</sup> and is currently available for clinicians to use at no cost.

## HOW MIGHT OUTCOME ASSESSMENT IMPROVE THE CARE OF DEPRESSED PATIENTS?

The authors believe that tools now exist to enable clinicians working in outpatient practices to measure outcome in a manner that does not interfere with their work flow. Why should clinicians change their behavior and adopt a measurement-based care approach towards treatment? The authors wish they could point to empirical research demonstrating the therapeutic benefit of measurement based care. In fact, there is little data demonstrating that measurement-based care improves outcome. The STAR\*D researchers asserted that measurement-based care was beneficial because treatment response during Phase I of the STAR\*D study was similar to the response rates typically found in industry-sponsored efficacy trials despite having much broader inclusion criteria in STAR\*D that resulted in the inclusion of more difficult-to-treat patients who had comorbid conditions.<sup>7</sup>

There are several reasons why measurement-based care is likely to improve outcome. One reason is the improved detection of residual symptoms in patients who have improved in treatment,

but have not remitted. A number of studies have demonstrated that the presence of residual symptoms in depressed patients who have improved with treatment predicts poorer long-term outcome.<sup>31-34</sup> How well do clinicians detect such residual symptoms? The authors are not aware of studies that have addressed this question. However, as demonstrated in the STAR\*D, residual symptoms are common with 33% of the patients experiencing mild-moderate levels of symptoms at the end of the acute phase of treatment with citalopram.<sup>7</sup> While it is true that the remission rate during the acute phase in the STAR\*D was modest despite the use of measurement to guide treatment decision making, cumulative remission rates after multiple levels of treatment are >60%.<sup>35,36</sup> We agree with the STAR\*D researchers' speculation that quantified measurement enhanced outcome because incomplete response could not be ignored. The use of a measurement tool should make it less likely that residual symptoms which leave patients at greater risk for relapse are unrecognized, as well as reduce the likelihood of incorrectly concluding that symptomatic patients have positively responded to treatment based on global attestations that they are doing "fine" or "okay."

It is also important to measure outcome at the beginning of treatment. Several studies have shown that patients who have not improved by at least 20% after 2 weeks of treatment are unlikely to improve after 6–10 weeks.<sup>37,38</sup> Early nonresponse to treatment is the best predictor of failure to respond or remit with treatment and, therefore, could suggest to clinicians that treatment should be altered. Caution, though, is warranted because no studies have established that changing treatment in nonresponders after 2 weeks results in improved outcome.

Many depressed patients drop out of treatment within the first 6 months of initiating care,<sup>39,40</sup> and many stay in treatment for years because of the chronic nature of the illness. Routine outcome assessment with self-report scales can enhance therapeutic effectiveness for different reasons depending on the stage of treatment. The completion of self-administered scales increases patients' active participation in their care, and this might facilitate participation in other therapeutic activities such as exercise or pleasant activities. Patients who are more active in their treatment, and who believe that their clinicians better understand their clinical status, may be more likely to continue with treatment. Valid symptom assessment may help clinicians identify for patients areas of improvement that had not been recognized. For example, consider a patient who is still depressed, pessimistic, unmotivated, and self-deprecatory who, at the beginning of the follow-up visit, states that he or she is no better, but who in fact is sleeping better, feeling somewhat more energetic, eating better, and concentrating better. Identification of some areas of improvement could reduce patients' therapeutic nihilism, thereby increasing treatment retention. Thus, more accurate symptom assessment might not only improve the detection of residual symptoms in patients who report that

they are feeling better, but it can also improve the detection of mild levels of improvement (eg, 20% to 30% improvement over baseline), which might be a harbinger of future improvement in patients who are not yet doing well.<sup>37,38</sup>

Patients followed longitudinally, over the course of years, may uniquely benefit from the routine use of scales. For example, it may be easier to detect seasonal patterns of symptom fluctuation when looking at graphs of symptom scores. Patients who relapse, and distort the effectiveness of treatment by minimizing or overlooking periods of sustained remission because of state-dependent cognitive biases, may be more open to veridical reports of their longitudinal course when shown the forms they had completed which indicated minimal levels of symptoms.

There are thus multiple theoretical reasons as to why measurement-based care might improve treatment outcome. While the fact remains that there is no research demonstrating the benefits of measurement-based care, the cost in adopting a measurement-based approach towards treatment is so small, and the potential benefits significant, that it is difficult to justify not trying it out for a period of time.

## HOW TO GET CLINICIANS TO ADOPT MEASUREMENT-BASED CARE

This article has been addressed to the practicing clinician and has attempted to make a rational case for modifying clinical care to include quantifiable assessments. Some clinicians will conduct experiments to determine whether they can appreciate an improvement in efficiency or outcome, and at what cost. Will their patients be put off by completing scales? Will patients object, express dissatisfaction, or embrace it? Will work flow be impeded?

However, there are other stakeholders to consider in facilitating a paradigm shift towards measurement-based care, and stakeholders such as health insurance companies or employers may be in positions to encourage clinicians to change. How might an insurance company get clinicians to administer a scale of their choosing to all patients in order to measure outcome? They could demand or require it, though this approach is likely to engender resistance.

Instead of behavioral change by edict, insurance companies could provide incentives to encourage outcome measurement. Possible incentives include directing referrals, reducing levels of micromanagement and oversight, increased reimbursement rates, or underwriting costs of electronic medical record systems. The authors would predict that the collaborative, incentive-based approach is more likely to achieve success.

The purchasers of health insurance, employers, might likewise incentivize the health insurance industry to determine how well their healthcare dollars are being spent, and this can only be done if outcome is being measured and aggregated over large groups.

Another stake holder is the healthcare organization. The first organization in a geographic area to adopt measurement-based care approaches may be better able to compete for contracts. Such organizations could advertise their outcome data and quality of service to both the public as well as health insurers. Thus, large practices might get their clinicians to measure outcomes if it is perceived that future financial viability is dependent on the establishment of an outcome measurement program.

## CONCLUSION

The authors predict that during the next few years increasing numbers of individual clinicians and institutional programs will consider whether to incorporate standardized assessment tools into their practice. The ultimate goal of an outcomes assessment program is to improve the quality of care. For an outcomes assessment program to be successful, the information collected should be clinically meaningful and quantifiable, the methods of data collection should not significantly interfere with clinical work flow, and the data collected should be used to evaluate the impact of various parameters on outcome and whether changes in service delivery improve outcome. The tools and technology now exist to overcome the challenges posed by a measurement-based approach towards care, and this will hopefully accelerate the incorporation of measurement into routine clinical practice. **PP**

## REFERENCES

1. America's Best Hospitals. *U.S. News and World Report*. 2009;140(7):89-112.
2. Zimmerman M. Integrating the assessment methods of researchers in routine clinical practice: The Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS) project. In: First MB, ed. *Annual Review of Psychiatry*. Washington, DC: American Psychiatric Press, Inc.; 2003:29-34.
3. Zimmerman M, Mattia JI. Psychiatric diagnosis in clinical practice: is comorbidity being missed? *Compr Psychiatry*. 1999;40(3):182-191.
4. Zimmerman M, Mattia JI. A self-report scale to help make psychiatric diagnoses: The Psychiatric Diagnostic Screening Questionnaire (PDSQ). *Arch Gen Psychiatry*. 2001;58(8):787-794.
5. Zimmerman M, Chelminski I, McGlinchey JB, Posternak MA. A clinically useful depression outcome scale. *Compr Psychiatry*. 2008;49(2):131-140.
6. Zimmerman M, Chelminski I, Young D, Dalrymple KL. A clinically useful anxiety outcome scale. *J Clin Psychiatry*. 2010;71(5):534-542.
7. Trivedi M, Rush A, Wisniewski S, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR\*D implications for clinical practice. *Am J Psychiatry*. 2006;163(1):28-40.
8. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: American Psychiatric Association; 1994.
9. Guy W. *ECDEU Assessment Manual for Psychopharmacology*. Rockville, MD: National Institute of Mental Health; 1976.
10. Rush AJ, First MB, Blacker D. *Handbook of Psychiatric Measures*. 2nd ed. Washington, DC: American Psychiatric Publishing; 2008.
11. First MB, Spitzer RL, Gibbon M, Williams JB. *Structured Clinical Interview for DSM-IV Axis I Disorders - Patient edition (SCID-IP, version 2.0)*. New York, NY: Biometrics Research Department, New York State Psychiatric Institute; 1995.
12. Endicott J, Spitzer RL. A diagnostic interview: the schedule for affective disorders and schizophrenia. *Arch Gen Psychiatry*. 1978;35(7):837-844.
13. Zimmerman M, McGlinchey J. Depressed patients' acceptability of the use of self-administered scales to measure outcome in clinical practice. *Ann Clin Psychiatry*. 2008;20(3):125-129.
14. Zimmerman M, Posternak M, Chelminski I. Using a self-report depression scale to identify remission in depressed outpatients. *Am J Psychiatry*. 2004;161(10):1911-1913.
15. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry*. 1961;4:561-571.
16. Kroenke K, Spitzer R, Williams J. The PHQ-9. Validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606-613.
17. Coles ME, Cook LM, Blake TR. Assessing obsessive compulsive symptoms and cognitions on the internet: evidence for the comparability of paper and internet administration. *Behav Res Ther*. 2007;45(9):2232-2240.
18. Buchanan T. Online assessment: desirable or dangerous? *Prof Psychol Res Pr*. 2002;33:148-154.

19. Buchanan T. Internet-based questionnaire assessment: appropriate use in clinical contexts. *Cogn Behav Ther*. 2003;32:100-109.
20. McDonald A. The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments. *Comput Educ*. 2002;39:299-312.
21. Webster J, Compeau D. Computer-assisted versus paper-and-pencil administration of questionnaires. *Behav Res Methods Instrum Comput*. 1996;28:567-576.
22. *Guidelines for Computer-based Tests and Interpretations*. Washington, DC: American Psychological Association; 1986.
23. Barbeite F, Weiss E. Computer self-efficacy and anxiety scales for an internet sample: Testing measurement equivalence of existing measures and development of new scales. *Comput Human Behav*. 2004;20:1-15.
24. Ogles BM, France CR, Lunnen KM, Bell MT, Goldfarb M. Computerized depression screening and awareness. *Community Ment Health J*. 1998;34(1):27-38.
25. Merten T, Ruch W. A comparison of computerized and conventional administration of the German versions of the Eysenck Personality Questionnaire and the Carroll Rating Scale for Depression. *Pers Individ Dif*. 1996;20:281-291.
26. Peterson L, Johannsson V, Carlsson S. Computerized testing in a hospital setting: Psychometric and psychological effects. *Comput Human Behav*. 1996;12:339-350.
27. Cook IA, Balasubramani GK, Eng H, et al. Electronic source materials in clinical research: acceptability and validity of symptom self-rating in major depressive disorder. *J Psychiatr Res*. 2007;41(9):737-743.
28. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 1979;134:382-389.
29. Guy W. *ECDEU Assessment Manual for Psychopharmacology*. Revised. Rockville, MD: National Institute of Mental Health; 1976.
30. Outcome Tracker. Available at: [www.outcometracker.org](http://www.outcometracker.org). Accessed May 28, 2010.
31. Paykel ES, Ramana R, Cooper Z, Hayhurst H, Kerr J, Barocka A. Residual symptoms after partial remission: An important outcome in depression. *Psychol Med*. 1995;25(6):1171-1180.
32. Thase ME, Simons AD, McGeary J, et al. Relapse after cognitive behavior therapy of depression: Potential implications for longer courses of treatment. *Am J Psychiatry*. 1992;149(8):1046-1052.
33. Simons A, Murphy G, Levine J, Wetzel R. Cognitive therapy and pharmacotherapy for depression. Sustained improvement over one year. *Arch Gen Psychiatry*. 1986;43(1):43-48.
34. Faravelli C, Ambonetti A, Pallanti S, Pazzagli A. Depressive relapses and incomplete recovery from index episode. *Am J Psychiatry*. 1986;143(7):888-891.
35. Quitkin F, McGrath P, Stewart J, et al. Remission rates with 3 consecutive antidepressant trials: Effectiveness for depressed outpatients. *J Clin Psychiatry*. 2005;66(6):670-676.
36. Rush AJ, Trivedi MH, Wisniewski SR, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR\*D report. *Am J Psychiatry*. 2006;163(11):1905-1917.
37. Szegedi A, Jansen WT, van Willigenburg AP, van der Meulen E, Stassen HH, Thase ME. Early improvement in the first 2 weeks as a predictor of treatment outcome in patients with major depressive disorder: a meta-analysis including 6562 patients. *J Clin Psychiatry*. 2009;70(3):344-353.
38. Tadic A, Helmreich I, Mergl R, et al. Early improvement is a predictor of treatment outcome in patients with mild major, minor or subsyndromal depression. *J Affect Disord*. 2010;120(1-3):86-93.
39. Lin EH, Von Korff M, Katon W, et al. The role of the primary care physician in patients' adherence to antidepressant therapy. *Med Care*. 1995;33(1):67-74.
40. Demyttenaere K, Enzlin P, Dewe W, et al. Compliance with antidepressants in a primary care setting, 1: beyond lack of efficacy and adverse events. *J Clin Psychiatry*. 2001;62 suppl 22:30-33.