Why Don't Psychiatrists Use Scales to Measure Outcome When Treating Depressed Patients?

Mark Zimmerman, M.D., and Joseph B. McGlinchey, Ph.D.

Objective: A survey of psychiatrists in the United Kingdom found that only a minority routinely used standardized measures to assess outcome when treating depression and anxiety disorders. The goals of the present study were to determine how frequently psychiatrists in the United States use scales to measure outcome when treating depressed patients and, for those clinicians who do not regularly use such scales, to ascertain the reasons for the lack of use.

Method: The subjects were 314 psychiatrists who attended a continuing medical education conference in California, Massachusetts, New York, or Wisconsin in 2006 or 2007. Prior to a lecture, the subjects completed a questionnaire that included 2 questions regarding the use of rating scales to monitor outcome when treating depression.

Results: More than 80% of the psychiatrists indicated that they did not routinely use scales to monitor outcome when treating depression. The most frequent reasons psychiatrists gave for not using scales were that they did not believe scales would be clinically helpful, that scales take too much time to use, and that they were not trained in the use of such measures.

Conclusions: The majority of psychiatrists indicated that they do not routinely use standardized measures to evaluate outcome when treating depressed patients. The Centers for Medicare and Medicaid Services' Physician Quality Reporting Initiative is intended to improve quality of care by providing physicians financial incentives to document outcomes reflecting best practices. If standardized outcome assessment is to assume increasing importance in this country, either educational efforts or payor mandates, or both, will be necessary to change clinicians' behavior.

(J Clin Psychiatry 2008;69:1916–1919)

Imagine going to your primary care doctor with a fever and symptoms of an upper respiratory tract infection. Your primary care provider puts his or her palm to your forehead and agrees that you feel warm. A course of treatment is recommended, you return in a couple of days, and he or she again feels your forehead and notes that you are cooler. Would you continue to see a doctor who evaluated your body temperature in this way?

To determine the impact of treatment, it is necessary to evaluate outcome. In mental health clinical settings, evaluation typically is based on unstructured interactions that yield *unquantified* judgments of progress. This practice is at variance with other areas of medical care in which outcome is determined, in part, on the change of a numerical value. Body temperature, blood pressure, cholesterol values, blood sugar levels, cardiac ejection fraction, and white blood cell counts are examples of quantifiable variables that are used to evaluate treatment progress. In the mental health field, standardized, quantifiable outcome measures exist for most major psychiatric disorders, yet they are rarely used in routine clinical practice.

Trivedi and colleagues[1] suggested that one of the primary lessons learned from the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) project on the effectiveness of treating depression was the importance of using quantitative, reliable, and valid scales to measure outcome. They referred to this strategy as "measurement-based care" and recommended that it become the standard of care.

Gilbody et al.[2] surveyed 340 psychiatrists in the United Kingdom regarding their use of outcome measures. Only 11.2% of the psychiatrists routinely used standardized measures to assess outcome when treating depression and anxiety disorders. More than half of the clinicians indicated that they never used standardized measures to evaluate outcome. The authors did not ask the respondents why they were disinclined to use scales to measure outcome; however, the authors noted that several respondents included comments on the questionnaires indicating that they thought such scales were simplistic, not useful in clinical practice, of questionable reliability and validity, or overly burdensome and costly to implement routinely.

We are not aware of comparable surveys of psychiatrists in the United States regarding their use of standardized scales to measure outcome. The goals of the present

Received March 28, 2008; accepted May 13, 2008. From the Department of Psychiatry and Human Behavior, Brown University School of Medicine, Providence, R.I.

Drs. Zimmerman and McGlinchey have no personal affiliations or financial relationships with any commercial interest to disclose relative to the article.

Corresponding author and reprints: Mark Zimmerman, M.D., Bayside Medical Center, 235 Plain St., Providence, RI 02905 (e-mail: mzimmerman@lifespan.org).

For Clinical Use

- ◆ In mental health clinical settings, assessments of outcome are typically based on unstructured interactions that yield *unquantified* judgments of progress.
- ◆ The results of 2 surveys found that psychiatrists typically do not use standardized scales with established reliability and validity to monitor outcome when treating patients with depression.
- ◆ Self-report questionnaires are a cost-effective option to monitor patient outcome because, although they are inexpensive in terms of professional time needed for administration, they correlate highly with clinician ratings.
- ◆ Brief self-report scales have been developed that are reliable and valid measures of depression and are feasible to incorporate into routine clinical practice.

study were to determine how frequently psychiatrists use scales to measure outcome when treating depressed patients and, for those clinicians who do not regularly use such scales, to ascertain the reasons for the lack of use.

METHOD

The subjects were 314 psychiatrists who attended a continuing medical education conference in California, Massachusetts, New York, or Wisconsin in 2006 or 2007. The subjects completed a questionnaire before the first author (M.Z.) delivered a lecture on the treatment of depression. The conferences were half-day or full-day events with multiple speakers. The title of the lecture did not suggest that it would address the topic of measuring outcome when treating depression.

The first part of the questionnaire elicited subjects' demographic characteristics (age, sex) and professional background (medical specialty, profession, practice setting, and years in practice). The second part of the questionnaire included 6 questions, 2 of which addressed the use of rating scales. The first question was "How often do you use a rating scale to monitor the course of treatment for depression? (a) never, (b) rarely, (c) sometimes, (d) frequently, (e) almost all the time." The second question was intended to determine the reasons clinicians do not regularly use scales to monitor outcome. Accordingly, only subjects who responded "a," "b," or "c" to the first question were asked to respond to the second question. The second item was "Please indicate why you do not routinely use rating scales to monitor the course of depression (circle all that apply): (a) do not believe it would be clinically helpful, (b) do not know what measure to use, (c) takes too much time, (d) too disruptive of clinical practice, (e) wasn't trained to use them, (f) other." The response alternatives were based on conversations with psychiatrists about why they do not use scales in clinical practice. The protocol was reviewed by the Rhode Island Hospital Institutional Review Board, and, because the study did not involve more than minimal risk or disclosure of protected health information, written informed consent was not necessary. This was announced before the respondents were asked to complete the questionnaire.

RESULTS

The majority of the 314 subjects were male (64.5%) and worked in an outpatient setting (71.7%). The mean age of the subjects was 55.8 years (SD = 11.3 years), and they had been in practice a mean of 22.8 years (SD = 11.9 years).

The vast majority of psychiatrists indicated that they did not routinely use scales to monitor outcome of treating depression (Table 1). More than half of the psychiatrists indicated that they never or rarely used scales to monitor outcome, and less than 10% almost always used scales to monitor outcome of depression treatment.

Subjects who reported never, rarely, or sometimes using scales to monitor outcome were asked the reasons for not routinely using scales in their clinical practice. More than one quarter of the subjects indicated that they did not believe using scales would be clinically helpful, that the scales take too much time to use, and that they were not trained in their use (Table 2). More than one quarter of the subjects noted reasons other than the ones listed and wrote in responses to the open-ended question. The most commonly noted free-form responses were that the psychiatrists preferred to assess outcome clinically and that they were not in the habit of using scales.

We compared the characteristics of psychiatrists who reported using scales frequently or almost always to the rest of the group. There were no differences between the 2 groups in gender, age, years of practice, or practice setting.

DISCUSSION

Similar to the results of the survey of British psychiatrists conducted by Gilbody and colleagues,[2] our results indicated that most psychiatrists in the United States are not routinely using standardized measures to evaluate outcome when treating depressed patients. Psychiatrists did

Table 1. Reported Frequency of Use of Standardized Scales by a Sample of U.S. Psychiatrists to Measure Outcome in the Treatment of Depression (N = 306)[a]

Frequency	%	N
Never	28.8	88
Rarely	32.0	98
Sometimes	21.2	65
Frequently	11.4	35
Almost all the time	6.5	20

[a] Data were not included for 8 subjects because of missing data (N = 7) or because more than 1 response alternative was checked (N = 1).

not believe that the regular use of scales would be clinically helpful, and they thought that scales would take too much time to administer. In addition, the psychiatrists indicated that a lack of previous training was a reason for not using scales. We suspect that the psychiatrists were considering clinician rating scales such as the Hamilton Rating Scale for Depression[3] and the Montgomery-Asberg Depression Rating Scale[4] when answering why they did not use scales, rather than thinking of self-report scales, which require little training and do not take much clinician time to administer and score.

Outcome assessment is assuming increasing importance in this country, and while psychiatrists have not yet embraced the use of standardized scales in clinical practice, payor mandates may accelerate a change in clinicians' behavior. The Centers for Medicare and Medicaid Services' Physician Quality Reporting Initiative (PQRI),[5] signed into law in 2006, is intended to improve quality of care by providing physicians financial incentives to document outcomes reflecting best practices. In 2007, the first year of the PQRI, 74 indicators were listed, 1 of which was related to the treatment of depression: percentage of patients aged 18 years and older diagnosed with a new episode of major depressive disorder and documented as treated with antidepressant medication during the entire 84-day (12-week) acute treatment phase. For 2008, the PQRI list of indicators was expanded to 134 items, with 2 additional indicators related to the treatment of depression: (1) percentage of patients aged 18 years and older with a new diagnosis or recurrent episode of major depressive disorder who met the DSM-IV criteria during the visit in which the new diagnosis or recurrent episode was identified and (2) percentage of patients aged 18 years and older with a new diagnosis or recurrent episode of major depressive disorder who had a suicide risk assessment completed at each visit. Thus far, the PQRI indicators related to depression have referred to the adequacy of assessment and treatment duration. Perhaps a future version of the PQRI list of indicators will include a determination of the effectiveness of treatment. This will require psychiatrists to measure outcome in a systematic manner. The results of the present study suggest that educational efforts will probably be required to acquaint psychiatrists with reliable, valid measures that are feasible to incorporate into clinical practice.

Table 2. Reasons Selected by a Sample of U.S. Psychiatrists for Not Using Standardized Scales to Measure Outcome in the Treatment of Depression (N = 248)[a]

Reason	%	N
Do not believe it would be clinically helpful	27.8	69
Do not know what measure to use	20.6	51
Takes too much time	33.9	84
Too disruptive of clinical practice	19.0	47
Wasn't trained to use them	34.3	85
Other	28.6	71

[a] Only subjects who indicated that they never, rarely, or sometimes used scales were asked to respond to this question. Subjects could check off more than 1 response. Three subjects who indicated that they never, rarely, or sometimes used scales did not respond to this question.

The results of the present survey should not be interpreted as a criticism of clinicians' behavior. The current community standard of care does not include quantified measurement of outcome. It is our hope, however, that the standard of care will change over the next decade.

In the future, if the delivery of mental health treatment increasingly requires the measurement of outcome, then the user friendliness of measurement tools as well as their reliability and validity will be critical to their widespread adoption. Clinicians are already overburdened with paperwork, and adding to this load by requiring repeated detailed evaluations with such instruments as the Hamilton Rating Scale for Depression is unlikely to meet with success. Self-report questionnaires are a cost-effective option because they are inexpensive in terms of professional time needed for administration and they correlate highly with clinician ratings. To be sure, there are also limitations with self-report questionnaires, such as response set biases, and their use may be limited by the readability of the scale and literacy of the respondent. However, self-report scales are free of clinician bias and are therefore free from clinician overestimation of patient improvement (which might occur when there is incentive to document treatment success).

There is no shortage of self-report questionnaires that assess the severity of depression. In fact, enough depression scales have been developed to warrant the publication of a compendium of these measures.[6] While the reliability and validity of many scales have been demonstrated, there are important differences between scales that might impact the feasibility of their use in routine clinical practice. Some measures lack adequate coverage of the DSM-IV diagnostic criteria,[7,8] some are expensive to purchase,[9] some are somewhat complicated to score,[7] and some may take too long to complete and are therefore less convenient for use in clinical practice. Scales such as the Beck Depression Inventory,[9] the Diagnostic Inventory for Depression,[10] and the Quick Inventory of Depressive Symptomatology[11] assess symptoms with groups of 4 or 5 statements and are thus composed of 60 or more statements. These scales take respondents 5 to 15 minutes to complete, and this may be too long for regular use in clinical practice in which the scale would be routinely administered at follow-up appointments. Because of this problem, our research group developed the Clinically Useful Depression Outcome Scale (CUDOS), a self-administered questionnaire that was designed to be brief (on average, completed in less than 2 minutes), quickly scored (in less than 15 seconds), clinically useful (fully covering the DSM-IV symptoms of major depressive disorder and dysthymic disorder), reliable, valid, and sensitive to change.[12] In order to keep the CUDOS brief, a Likert scale was used to rate the 16 symptom statements. In a study[13] comparing the feasibility and acceptability of the CUDOS and the Beck Depression Inventory to assess outcome in clinical practice, significantly more patients indicated that the CUDOS took less time to complete, was less of a burden to complete, and was preferred as a measure to complete during ongoing treatment.[13]

Another brief instrument in which items are rated on a Likert scale is the 9-item Patient Health Questionnaire (PHQ-9).[14] Because of ease-of-use considerations, we would recommend that either the CUDOS or the PHQ-9 be used by clinicians at every visit to monitor the course of depression. In fact, because it contains fewer items than the CUDOS, the PHQ-9 probably takes a little less time to complete. However, the advantage offered by being somewhat briefer is offset by some loss of information. The PHQ-9 adheres to the construction of the DSM-IV criteria; thus, compound DSM-IV criteria that refer to more than 1 symptom (e.g., insomnia or hypersomnia, increased or decreased appetite) are represented by a single item on the PHQ-9. Since treatment decision-making might be influenced by whether a patient has insomnia or is sleeping too much, or has a reduced appetite or is eating too much, the PHQ-9 does not capture potentially clinically significant information. The CUDOS is also unique in that it is the only self-administered depression scale that not only evaluates the symptoms of depression but also assesses both psychosocial impairment due to depression and quality of life with reliable and valid single-item questions for these domains.[15] The importance of these constructs has been increasingly recognized during the past decade.[16] However, more important than which scale clinicians use to measure outcome is that some quantifiable index is used to track the progress of treatment.

The results of the present study should be interpreted with caution because we did not conduct a random survey of psychiatrists practicing in the United States. Rather, we surveyed psychiatrists attending medical education conferences at geographically diverse regions of the country. Psychiatrists attending these programs may not be representative of all psychiatrists, although we are unsure in which direction the findings might be biased. Are such attendees better informed and more likely to have incorporated measurement-based care into their practices? Or are they less likely? Considering that the results of the present study were so clear-cut—there is a widespread lack of use of scales in clinical practice—we would expect that the results of a study of a nationally representative sample would produce very similar results. The similarity of the results of the present study to those of Gilbody and colleagues,[2] based on a nationally representative sample of psychiatrists in the United Kingdom, further supports the validity of the present findings.

Disclosure of off-label usage: The authors have determined that, to the best of their knowledge, no investigational information about pharmaceutical agents that is outside U.S. Food and Drug Administration–approved labeling has been presented in this article.

REFERENCES

1. 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;163:28–40
2. Gilbody S, House A, Sheldon T. Psychiatrists in the UK do not use outcomes measures. Br J Psychiatry 2002;180:101–103
3. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;23:56–62
4. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979;134:382–389
5. US Department of Health and Human Services, Centers for Medicare and Medicaid Services. Physician Quality Reporting Initiative. Available at: http://www.cms.hhs.gov/PQRI/30_EducationalResources.asp. Accessed Mar 3, 2008
6. Nezu A, Ronan G, Meadows E, et al. Practitioner's Guide to Empirically Based Measures of Depression. New York, NY: Kluwer Academic/Plenum Publishers; 2000
7. Zung WWK. A self-rating depression scale. Arch Gen Psychiatry 1965;12:63–70
8. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Applied Psychological Measurement 1977;1:385–401
9. Beck AT, Ward CH, Mendelson M, et al. An inventory for measuring depression. Arch Gen Psychiatry 1961;4:561–71
10. Zimmerman M, Sheeran T, Young D. The Diagnostic Inventory for Depression: a self-report scale to diagnose DSM-IV major depressive disorder. J Clin Psychol 2004;60:87–110
11. Rush AJ, Trivedi MH, Ibrahim HM, et al. The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biol Psychiatry 2003;54:573–583
12. Zimmerman M, Chelminski I, McGlinchey JB, et al. A Clinically Useful Depression Outcome Scale. Compr Psychiatry 2008;49:131–140
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;20:125–129
14. Kroenke K, Spitzer R, Williams J. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16:606–613
15. Zimmerman M, Ruggero C, Chelminski I, et al. Developing brief measures for use in clinical practice: the reliability and validity of single-item self-report measures of depression symptom severity, psychosocial impairment due to depression, and quality of life. J Clin Psychiatry 2006;67:1536–1541
16. Zimmerman M, McGlinchey JB, Posternak M, et al. Remission in depressed patients: more than just symptom resolution. J Psychiatr Res 2008;42:797–801

For the CME Posttest for this article, see pages 2004–2006.