Developing Brief Scales 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

Mark Zimmerman, M.D.; Camilo J. Ruggero, Ph.D.; Iwona Chelminski, Ph.D.; Diane Young, Ph.D.; Michael A. Posternak, M.D.; Michael Friedman, M.D.; Daniela Boerescu, M.D.; and Naureen Attiullah, M.D.

Objective: Reliable, valid, user-friendly measurement is necessary to successfully implement an outcomes evaluation program in clinical practice. Self-report questionnaires, which generally correlate highly with clinician ratings, are a cost-effective assessment option. However, even self-administered questionnaires can be burdensome to patients because many are lengthy. Consequently, we developed and determined the reliability and validity of ultra-brief, single-item assessments of 3 domains important to consider when treating depressed patients: symptom severity, psychosocial functioning, and quality of life.

Method: In the first study (conducted June 1997 to March 2002), 1278 psychiatric outpatients with various DSM-IV diagnoses completed single-item assessments of psychosocial functioning and quality of life as well as more detailed measures of these constructs. In the second study (conducted August 2003 to July 2004), 562 psychiatric outpatients who were in ongoing treatment for a DSM-IV major depressive episode completed a depression symptom scale and a measure of global severity of depression.

Results: The test-retest reliability of the psychosocial functioning and quality-of-life items was high. The single-item measures of symptom severity, psychosocial functioning, and quality of life were significantly correlated with the total scores and individual item scores of longer measures of the same constructs ($p < .001$). The single-item measures significantly discriminated between depressed patients in full remission, in partial remission, and in a current depressive episode ($p < .001$).

Conclusion: These studies provide evidence of the reliability and validity of single-item measures of symptom severity, psychosocial functioning, and quality of life. Very brief measures, such as the ones described in the present report, are not burdensome for patients to complete and can be easily incorporated into a busy clinical practice in order to collect data on treatment effectiveness.

(J Clin Psychiatry 2006;67:1536–1541)

Received Jan. 16, 2006; accepted April 24, 2006. From the Department of Psychiatry and Human Behavior, Brown University School of Medicine, Rhode Island Hospital, Providence, R.I.

The authors report no financial or other affiliations relevant to the subject of this article.

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

During the past decade, the treatment literature increasingly has emphasized the importance of going beyond symptom assessments to include evaluations of quality of life and psychosocial functioning. Some scales assessing symptoms, quality of life, and psychosocial functioning are quite lengthy, and repeated completion of all of these scales can be overly burdensome to patients. For example, scales such as the Beck Depression Inventory,[1] Diagnostic Inventory for Depression[2] (DID), and Inventory of Depressive Symptomatology[3] assess symptoms with groups of 4 or 5 statements and are thus composed of 80 or more statements. These scales take respondents 10 to 20 minutes to complete. The Social Adjustment Scale[4] is composed of 48 items, and the Quality of Life Enjoyment and Satisfaction Scale[5] is composed of 60 items. Briefer measures of symptoms, functioning, and quality of life also exist, though these, too, often consist of 15 or more statements or items.[6–7]

There are several reasons why it would be desirable to develop very brief, valid measures of these concepts, ideally as short as single items. Both clinical and community-based studies that are not primarily focused on depression nonetheless often incorporate an index of depression, though, because this is only a peripheral component of the main protocol, the measure is brief. Even when the primary focus is on depression, it might be beneficial, or even necessary, to use brief measures. Researchers of the effectiveness of treatment in clinical practice desire brief measures of outcome to avoid disrupting the flow of work in routine clinical care. For

example, in our clinical research laboratory, we recently conducted a study of remission from depression in depressed outpatients receiving ongoing treatment.[8,9] Because we were interested in multiple constructs that might be related to remission, we needed to use brief measures so as to not overly burden the patients.

As part of the Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS) project, we recently completed a survey of the factors that depressed patients judged to be important in defining remission from depression.[9] We found that patients considered symptom resolution, functional improvement, and life satisfaction as all being important in determining remission. These findings confirmed our hypothesis that the evaluation of treatment outcome should not be limited to assessments of symptoms. We were further interested in examining the independent and additive association between level of severity of depressive symptoms, functional impairment, and quality of life and depressed patients' subjective evaluation of their remission status. In order to design a study that was acceptable to both patients and clinicians, we needed to use very brief assessments. Prior research from our laboratory[2,8] had incorporated information on global, single-item assessments of depression symptom severity, psychosocial impairment due to depression, and quality of life. However, we had never demonstrated the reliability and validity of these assessments. In the present report from the MIDAS project, we examined the test-retest reliability and validity of single-item global measures of the severity of depression (GSEVDEP), psychosocial functioning (GPF), and quality of life (GQOL) in 2 studies. In the first study, we examined the test-retest reliability and concurrent validity of the GPF and GQOL scales, and in the second study, we examined the validity of the GSEVDEP measure.

METHOD

Study 1

In the first study, 1278 psychiatric outpatients with various DSM-IV Axis I diagnoses completed the DID before their intake appointment at the Rhode Island Hospital (RIH) Department of Psychiatry outpatient practice between June 1997 and March 2002. This private practice group predominantly treats individuals with medical insurance on a fee-for-service basis, and it is distinct from the hospital's outpatient residency training clinic that predominantly serves lower-income, uninsured, and medical-assistance patients. Patients are referred from a variety of sources, the most common being primary care physicians and therapists in the community, though data on referral source was not systematically recorded. Not all patients who presented for treatment participated in the study. Because one of the goals of the MIDAS project is to develop and study the reliability and validity of self-administered questionnaires, patients with significant cognitive limitations were not included; thus, we disproportionately excluded elderly patients. Patients who did and did not participate in the study were similar in scores on self-administered symptom questionnaires.

Table 1. Current DSM-IV Axis I Diagnoses of 1278 Psychiatric Outpatients[a]

DSM-IV Diagnosis	N	%
Major depressive disorder	582	45.5
Bipolar I depression	13	1.0
Bipolar II depression	32	2.5
Dysthymic disorder	90	7.0
Generalized anxiety disorder	266	20.8
Panic disorder without agoraphobia	48	3.8
Panic disorder with agoraphobia	184	14.4
Social phobia	365	28.6
Specific phobia	129	10.1
Obsessive-compulsive disorder	94	7.4
Posttraumatic stress disorder	147	11.5
Adjustment disorder	75	5.9
Schizophrenia	8	0.6
Schizoaffective disorder	6	0.5
Bulimia nervosa	14	1.1
Binge-eating disorder	31	2.4
Alcohol abuse/dependence	127	9.9
Drug abuse/dependence	66	5.2
Somatization disorder	8	0.6
Undifferentiated somatoform disorder	31	2.4
Hypochondriasis	17	1.3

[a]Individuals could be given more than 1 diagnosis.

The sample included 488 men (38.2%) and 790 women (61.8%) who ranged in age from 18 to 79 years (mean = 37.3, SD = 12.28). Approximately two fifths of the subjects were married (39.2%, N = 501); the remainder were single (31.2%, N = 399), divorced (14.0%, N = 179), separated (6.5%, N = 83), widowed (1.7%, N = 22), or living with someone as if in a marital relationship (7.4%, N = 94). About 10 percent (N = 133) of the subjects did not graduate from high school, 62.4% (N = 797) graduated from high school or achieved equivalency, and 27.2% (N = 348) graduated from college. The data in Table 1 show the diagnostic composition of the sample.

All patients were interviewed by a trained diagnostic rater who administered the Structured Clinical Interview for DSM-IV[10] (SCID), supplemented with questions from the Schedule for Affective Disorders and Schizophrenia[11] assessing the severity of symptoms during the week prior to the evaluation. Patients were also rated on the Global Assessment of Functioning (GAF) scale. The RIH institutional review committee approved the research protocol, and all patients provided informed, written consent.

Subjects completed the DID as part of their initial paperwork. When scheduling their appointments, the subjects were told to arrive early to complete some standard forms. The DID takes approximately 15 to 20 minutes to complete. Test-retest reliability was examined in a consecutive series of 101 of the 1278 participants. These sub-

  © Copyright 2006 Physicians Postgraduate Press, Inc.

Table 2. Single-Item Global Measures of Severity of Depression (GSEVDEP), Psychosocial Functioning (GPF), and Quality of Life (GQOL)

GSEVDEP
Rate the current level of severity of your symptoms of depression during the past week.
0 None
1 Minimal
2 Mild
3 Moderate
4 Severe
GPF
Overall, how much have symptoms of depression interfered with or caused difficulties in your life during the past week?
0 Not at all
1 A little bit
2 A moderate amount
3 Quite a bit
4 Extremely
GQOL
In general, how would you rate your overall quality of life during the past week?
0 Very good, my life could hardly be better
1 Pretty good, most things are going well
2 The good and bad parts are about equal
3 Pretty bad, most things are going poorly
4 Very bad, my life could hardly be worse

jects were given the scale at the conclusion of the intake evaluation and asked to mail it back in a preaddressed postage-paid envelope. They were told that the purpose of the second administration was to test the performance of the scale, not to question the truthfulness or accuracy of their responses. All patients completed the second administration within 1 week of the first evaluation. A problem with examining the test-retest reliability of a state measure in psychiatric patients who present for treatment is that the patient's state often changes quickly. A review of studies of treatment response in psychopharmacology has shown that response is often early.[12] Clinical experience indicates that patients often feel much less distressed (and depressed) after the initial evaluation. Consequently, to study test-retest reliability over a longer interval in patients presenting for treatment would be inappropriate, because improvement is to be expected in many patients over a short interval.

The DID includes both a psychosocial functioning and quality-of-life subscale. The 6-item psychosocial functioning subscale assesses the amount of difficulty that symptoms of depression have caused in usual daily responsibilities, relationships with significant others such as a spouse, relationships with close family members, relationships with friends, and participation in leisure activities. There is also a global item of overall level of functional impairment due to depression, which we will refer to as the GPF scale (Table 2). All items are rated on a 5-point Likert scale (0 = no difficulty, 4 = extreme difficulty). The quality-of-life subscale assesses satisfaction with the same areas covered by the psychosocial functioning subscale, as well as general satisfaction with mental health and physical health. Items are rated on a 5-point Likert scale (0 = very satisfied, 4 = very dissatisfied). In addition, the DID includes a global quality-of-life question (Table 2). We will refer to this as the GQOL scale. For all psychosocial functioning and quality-of-life items, the instructions indicate that the respondent is to pick the item that best describes how he or she has been feeling during the past week.

Study 2

In the second study, participants were 562 psychiatric outpatients who were in ongoing treatment for a DSM-IV major depressive episode in the RIH Department of Psychiatry outpatient practice between August 2003 and July 2004. The sample included 191 men (34.0%) and 371 women (66.0%) who ranged in age from 18 to 80 years (mean = 44.0, SD = 11.7). The RIH institutional review committee approved the research protocol, and all patients provided informed, written consent.

Patients completed 2 questionnaires. One of the questionnaires was the Clinically Useful Depression Outcome Scale (CUDOS), a self-administered scale that has been found to be a valid indicator of remission status.[13,14] The CUDOS contains 16 items assessing all of the DSM-IV inclusion criteria for major depressive disorder. Compound DSM-IV symptom criteria referring to more than 1 construct (e.g., problems concentrating or making decisions, insomnia or hypersomnia) are subdivided into their respective components thus requiring 16 items to cover the 9 DSM-IV symptom criteria. Items are rated on a 5-point Likert scale indicating "how well the item describes you during the past week, including today" (0 = not at all true/0 days, 1 = rarely true/1–2 days, 2 = sometimes true/3–4 days, 3 = usually true/5–6 days, 4 = almost always true/every day).

The second questionnaire assessed patients' opinions regarding the importance of different factors in determining remission from depression. As part of this questionnaire, background demographic and clinical information was collected. One of the items was a single-item question regarding current level of severity of depression (GSEVDEP) as rated on a 5-point rating scale (Table 2). A subset of patients (N = 146) was also rated by the treating clinician on the Clinical Global Impressions-Severity of Illness (CGI-S) scale.[15]

Data Analysis

Study 1. We computed intraclass correlation coefficients to determine the test-retest reliability of the GPF and GQOL scales. We computed Spearman correlation coefficients between the GPF and GQOL with the ratings of items assessing individual components of these domains as well as total scores from these scales. When computing the total scores on the functional impairment

and quality-of-life subscales, the scores from the global items were not included. We also computed the correlation between the 1-item scales and the clinician-rated GAF. These analyses were conducted for the entire sample of 1278 patients as well as for the 582 patients who met DSM-IV criteria for a major depressive episode at the time of their evaluation.

At the time of presentation, 128 patients had major depressive disorder that was in partial remission, and 127 patients had prior episodes of depression that had resolved. We used an analysis of variance (ANOVA) to compare the scores on the GPF and GQOL scales in the 3 groups (currently depressed, partial remission, full remission). If the ANOVA was significant, Tukey follow-up tests were used for 2-group comparisons.

Study 2. We computed the Spearman correlation coefficients between the GSEVDEP rating and the CUDOS total score, the individual CUDOS item scores, and the clinician-rated CGI-S. At the time of the evaluation, 77 patients still were in their depressive episode, 146 were in partial remission, and 330 were in full remission (data on remission status were missing for 9 patients). We used an ANOVA to compare the scores on the GSEVDEP in the 3 groups and Tukey follow-up tests for 2-group comparisons.

RESULTS

Study 1

Test-retest reliability and validity of single-item global measurements of psychosocial functioning (GPF) and quality of life (GQOL). Test-retest reliability was determined in 101 patients. The intraclass correlation coefficients for the 1-item GPF (.76, $p < .001$) and GQOL (.81, $p < .001$) were high.

Twenty-eight (2.2%) of the 1278 patients did not answer the GPF, and 22 (1.7%) omitted the GQOL. There were no demographic differences between the patients who did and did not answer these questions. For the entire sample of patients, the mean (SD) score on the GPF was 2.43 (1.19), and the mean (SD) score on the GQOL was 2.33 (.94). Compared with nondepressed patients, depressed patients scored significantly higher on both the GPF (mean ± SD score = 3.02 ± .88 vs. 1.93 ± 1.19, $t = 18.14$, $p < .001$) and GQOL (mean ± SD score = 2.82 ± .75 vs. 1.93 ± .89, $t = 19.01$, $p < .001$).

The data in Table 3 show that the GPF was significantly correlated with each of the specific areas of functioning as well as the total impairment score. The GPF was significantly correlated with the clinician-rated GAF in the entire sample ($r = –.41$, $p < .001$) as well as in the subsample of depressed patients ($r = –.30$, $p < .001$). In fact, the correlation between the GPF and the GAF was as high as the correlation between the GAF and the total score on the psychosocial functioning scale (total sample: $r = –.42$, $p < .001$; depressed patients: $r = –.29$, $p < .001$).

Table 3. Correlation Between the Single-Item Global Measure of Psychosocial Functioning (GPF) and Specific Areas of Functioning in the Total Sample and Subsample of Depressed Patients[a,b]

Domain	Total Sample (N = 1250)	Depressed Patients (N = 570)
Work performance	.65	.50
Marital relationship	.50	.33
Family relationships	.50	.33
Friendships	.51	.35
Leisure activities	.66	.41
Total score[c]	.70	.51

[a]28 patients, 12 of whom had current major depressive disorder, did not complete the GPF, thereby reducing the total sample size to 1250 and the depressed patient sample size to 570.
[b]All correlations are significant at $p < .001$.
[c]Total score was calculated without the GPF included.

Table 4. Correlation Between the Single-Item Global Measure of Quality of Life (GQOL) and Specific Areas of Quality of Life in the Total Sample and Subsample of Depressed Patients[a,b]

Domain	Total Sample (N = 1256)	Depressed Patients (N = 573)
Work performance	.54	.29
Marital relationship	.46	.27
Family relationships	.42	.24
Friendships	.43	.27
Leisure activities	.55	.33
Mental health	.60	.37
Physical health	.41	.18
Total score[c]	.70	.46

[a]22 patients, 9 of whom had current major depressive disorder, did not complete the GQOL, thereby reducing the total sample size to 1256 and the depressed patient sample size to 573.
[b]All correlations are significant at $p < .001$.
[c]Total score was calculated without the GQOL included.

The data in Table 4 show that the GQOL was significantly correlated with each of the specific domains as well as the total quality-of-life score. Similar to the strength of the correlations between the GPF and the specific areas of function, the correlations between the GQOL and the specific domains were moderate for the total sample and lower for the depressed patients. The lower correlations in the depressed subsample were due to the more restricted range of scores.

Scatter plots of the single-item measures against their total subscale measures of the DID and the GAF did not reveal any clustering of scores around a floor or ceiling. Rather, scores were evenly distributed in a linear fashion across the range of the single-item measure. This linear pattern would suggest that both single-item measures have not sacrificed the original measure's ability to discriminate responses across the range of psychosocial functioning and quality of life, respectively.

We compared scores on the GPF and GQOL in depressed patients who were currently in a major depressive

Table 5. Correlation Between the Single-Item Global Measure of Severity of Depression and Individual Symptoms of Depression in 551 Depressed Outpatients[a]

Depression Symptom	Correlation Coefficient[b]
Depressed mood	.76
Decreased interest in usual activities	.71
Decreased appetite	.48
Increased appetite	.28
Insomnia	.50
Hypersomnia	.30
Psychomotor agitation	.46
Psychomotor retardation	.58
Decreased energy	.63
Guilt	.52
Worthlessness	.64
Decreased concentration	.60
Indecisiveness	.62
Thoughts about death	.44
Suicidal ideation	.38
Hopelessness	.62
Total score	.78

[a]11 of the 562 outpatients did not answer the single-item global measure of severity of depression.
[b]All correlations are significant at $p < .001$.

episode (N = 582), in partial remission (N = 128), and in full remission (N = 127). Both ANOVAs were significant (GPF, 3.02 ± .88 vs. 2.16 ± 1.03 vs. 1.86 ± 1.06, F = 107.27, $p < .001$; GQOL, 2.82 ± .74 vs. 1.81 ± .72 vs. 1.88 ± .80, F = 146.32, $p < .001$ for patients currently in episode, in partial remission, or in full remission, respectively). Tukey follow-up tests confirmed that the currently depressed patients reported significantly poorer psychosocial functioning than the patients in partial remission, and the patients in partial remission were significantly different than the patients in full remission. For the quality-of-life item, currently depressed patients scored significantly higher, indicating lower perceived quality of life, than the patients in partial or full remission; there was no significant difference between patients in partial versus full remission.

Study 2

Validity of a single-item global measure of severity of depression (GSEVDEP). Eleven (2.0%) of the 562 outpatients in the second study did not answer the GSEVDEP. Patients who did not answer the question were older on average than those who did answer the question (mean ± SD age = 51.54 ± 9.00 vs. 43.86 ± 11.69 years, t = 2.17, $p < .05$), but did not differ by race, sex, or education.

The mean (SD) score on the GSEVDEP was 1.90 (1.17). The data in Table 5 show that the GSEVDEP was significantly correlated with each of the individual symptom items on the CUDOS as well as the total scale score. The GSEVDEP was significantly correlated with the clinician-rated CGI-S ($r = .64$, $p < .001$).

In addition to the 11 patients missing data on the GSEVDEP item, data on remission status were missing for 9 other patients. We compared scores on the GSEVDEP in depressed patients who were currently in episode (N = 76), in partial remission (N = 139), and in full remission (N = 327); the ANOVA was significant (3.34 ± .62 vs. 2.57 ± .69 vs. 1.29 ± .95, F = 234.39, $p < .001$, respectively). Tukey follow-up tests confirmed that the currently depressed patients reported significantly greater symptom severity than the patients in partial remission, and the patients in partial remission scored significantly higher than the patients in full remission.

DISCUSSION

We believe that the optimal delivery of mental health treatment depends on measuring outcome. If outcome assessment is essential to determining treatment effectiveness, then quantitative, reliable, valid, user-friendly measurement will be necessary to successfully implement an outcomes evaluation program in clinical practice. 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[16] has not met with success.[17] 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. However, even self-administered questionnaires could be burdensome, particularly to the patients asked to complete them. Measures such as the Beck Depression Inventory or Social Adjustment Scale can take upwards of 20 minutes each to complete and routinely take 5 to 10 minutes to fill out. Expecting patients to routinely complete multiple scales at their visits is unlikely to meet with success. Consequently, we sought to develop reliable and valid ultra-brief, single-item, global assessments of 3 domains important to consider when treating depressed patients—symptom severity, psychosocial functioning, and quality of life. Such measures should be easier to incorporate into routine clinical practice because they would not be burdensome to patients.

The 2 studies presented evidence of the reliability and convergent validity of single-item measures of symptom severity, psychosocial functioning, and quality of life. Presumably, clinicians treating depressed patients already routinely assess the presence of depressive symptoms, as well as level of functioning and patients' satisfaction with their progress and life situation. However, such unstructured assessments do not lend themselves to quantitative evaluations of outcome. There are many valid, detailed measures of each of these constructs,[18] but they are rarely incorporated into clinical practice. The question is why.

We have speculated that part of the reason for the lack of widespread utilization of existing measures is the time burden on patients. It is also important to acknowledge the operational burden to administer and score such scales

  © Copyright 2006 Physicians Postgraduate Press, Inc.

that could interfere with the functioning of a busy clinical practice. Another obstacle in the use of existing measures is that some come with a financial cost, and this might dissuade clinicians from using them. However, alternatives are available that are free-of-charge. Finally, clinicians might not believe that the use of any type of standardized assessment adds value beyond their clinical evaluation. Although we are not aware of any research demonstrating that the use of standardized assessment tools to monitor the course of treatment improves the outcome of care, we hypothesize that the use of instruments to quantify a patient's status in clinical practice would heighten clinicians' sensitivity to patients' progress and reduce the likelihood that less than optimal outcomes are overlooked. It remains an empirical question whether this prediction is correct.

Some limitations of the present study should be noted. The study was conducted in a single outpatient practice in which the majority of the patients were white, female, and had health insurance. The generalizability to samples with different demographic characteristics needs to be demonstrated. Single-item global assessments of symptom severity, functioning, and quality of life provide clinicians with limited information regarding patient status. While more detailed assessments may be desirable, this must be balanced against the practicality of ascertaining such information. Brevity may come at a cost of detail; nonetheless, the results of the present studies suggest that brief measures are also reliable and valid.

REFERENCES

1. Beck AT, Ward CH, Mendelson M, et al. An inventory for measuring depression. Arch Gen Psychiatry 1961;4:561–571
2. Zimmerman M, Sheeran T, Young D. The Diagnostic Inventory for Depression: a self-report scale to diagnose DSM-IV for major depressive disorder. J Clin Psychol 2004;60:87–110
3. Rush AJ, Gullion CM, Basco MR, et al. The Inventory of Depressive Symptomatology (IDS). Psychol Med 1996;26:477–486
4. Weissman M, Bothwell S. Assessment of social adjustment by patient self-report. Arch Gen Psychiatry 1976;33:1111–1115
5. Endicott J, Nee J, Harrison W, et al. Quality of Life Enjoyment and Satisfaction Questionnaire: a new measure. Psychopharmacol Bull 1993;29:321–326
6. Rush A, Trivedi M, Ibrahim H, 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
7. Ware J, Sherbourne C. The MOS 36-item Short-Form Health Survey (SF-36)1 : conceptual framework and item selection. Med Care 1992;30: 473–483
8. Zimmerman M, McGlinchey J, Posternak M, et al. Discordance between self-reported symptom severity and psychosocial functioning ratings in depressed outpatients: implications for how remission from depression should be defined. Psychiatry Res 2006;141:185–191
9. Zimmerman M, McGlinchey JB, Posternak MA, et al. How should remission from depression be defined? the depressed patient's perspective. Am J Psychiatry 2006;163:148–150
10. First MB, Spitzer RL, Gibbon M, et al. Structured Clinical Interview for DSM-IV Axis I Disorders, Patient Edition (SCID-I/P, version 2.0). New York, NY: Biometrics Research Department, New York State Psychiatric Institute; 1995
11. Endicott J, Spitzer RL. A diagnostic interview: the Schedule for Affective Disorders and Schizophrenia. Arch Gen Psychiatry 1978;35:837–844
12. Posternak MA, Zimmerman M. Is there a delay in the antidepressant effect? a meta-analysis. J Clin Psychiatry 2005;66:148–158
13. Zimmerman M, Posternak M, Chelminski I. Using a self-report depression scale to identify remission in depressed outpatients. Am J Psychiatry 2004;161:1911–1913
14. Zimmerman M, Posternak M, McGlinchey J, et al. Validity of a self-report depression symptom scale for identifying remission in depressed outpatients. Compr Psychiatry 2006;47:185–188
15. Guy W. ECDEU Assessment Manual for Psychopharmacology. US Dept Health, Education, and Welfare publication (ADM) 76-338. Rockville, Md: National Institute of Mental Health;1976:218–222
16. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;23:56–62
17. Gilbody S, House A, Sheldon T. Psychiatrists in the UK do not use outcomes measures. Br J Psychiatry 2002;180:101–103
18. American Psychiatric Association. Handbook of Psychiatric Measures. Washington, DC: American Psychiatric Association; 2000