How Can We Use Depression Severity to Guide Treatment Selection When Measures of Depression Categorize Patients Differently?

Mark Zimmerman, MD; Jennifer H. Martinez, BA; Michael Friedman, MD; Daniela A. Boerescu, MD; Naureen Attiullah, MD; and Cristina Toba, MD

ABSTRACT

Objective: Treatment guidelines for depression suggest that severity should be taken into account when initiating treatment. If clinicians are to consider illness severity in selecting among treatment options for depression, then it is important to have reliable, valid, and clinically useful methods of distinguishing between levels of depression severity. In the present report from the Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS) project, we compared 3 self-report scales that assess the *DSM-IV* criteria for major depressive disorder on the basis of how these scales distribute patients into severity categories.

Method: From June 2010 to November 2011, 245 depressed outpatients completed the Clinically Useful Depression Outcome Scale (CUDOS), Quick Inventory of Depressive Symptomatology (QIDS), and Patient Health Questionnaire (PHQ-9). The study was conducted at Rhode Island Hospital, Providence, Rhode Island. The patients were subdivided into severity categories according to the cutoff scores recommended by each scales' developers. The patients were also rated on the 17-item Hamilton Depression Rating Scale (HDRS-17).

Results: The correlations between the HDRS-17 and the 3 self-report scales were nearly identical. Yet the scales significantly differed in their distribution of patients into severity categories. On the CUDOS and HDRS-17, moderate depression was the most frequent severity category, whereas on the PHQ-9 and QIDS, the majority of the patients were classified as severe. Significantly fewer patients were classified as severely depressed on the CUDOS compared to the PHQ-9 (McNemar = 153.8; $P < .001$) and QIDS (McNemar = 114.0; $P < .001$).

Conclusions: If clinicians are to follow treatment guidelines' recommendations to base initial treatment selection on the severity of depression, then it is important to have a consistent method of determining depression severity. The marked disparity between standardized scales in the classification of depressed outpatients into severity groups indicates that there is a problem with the use of such instruments to classify depression severity. Caution is warranted in the use of these scales to guide treatment selection until the thresholds to define severity ranges have been empirically established.

J Clin Psychiatry 2012;73(10):1287–1291

© Copyright 2012 Physicians Postgraduate Press, Inc.

Submitted: *March 12, 2012; accepted May 8, 2012.*
Online ahead of print: *September 4, 2012 (doi:10.4088/JCP.12m07775).*
Corresponding author: *Mark Zimmerman, MD, Bayside Medical Center, 235 Plain St, Providence, RI 02905 (mzimmerman@lifespan.org).*

Treatment guidelines for depression suggest that severity should be taken into account when initiating treatment. The recently revised American Psychiatric Association guidelines for the treatment of major depressive disorder (MDD) recommend both psychotherapy and pharmacotherapy as monotherapies for mild and moderate depression and pharmacotherapy (with or without psychotherapy) for severely depressed patients.[1] The National Institute for Health and Clinical Excellence updated guidelines for the treatment and management of depression discourage the use of antidepressant medication as the initial treatment option for mild depression and recommend medication together with empirically supported psychotherapy for moderate and severe depression.[2] As reported by van der Lem and colleagues,[3] the Netherlands treatment guidelines also recommend pharmacotherapy as the first treatment option for severely depressed patients and either pharmacotherapy or psychotherapy for mildly and moderately depressed patients. If clinicians are to consider illness severity in selecting among treatment options for depression, then it is important to have available reliable, valid, and clinically useful methods of distinguishing between levels of depression severity.

Many scales have been developed to measure the severity of depression.[4] In clinical practice, self-report questionnaires may be preferable to clinician-rated scales such as the Hamilton Depression Rating Scale (HDRS)[5] or the Montgomery-Asberg Depression Rating Scale[6] because they are less expensive in terms of professional time needed for administration. Zimmerman et al[7] discussed the use of self-report scales in routine clinical practice and recommended measures that assess the *Diagnostic and Statistical Manual of Mental Disorders,* Fourth Edition (*DSM-IV*) criteria for MDD and that are available for clinical use at no cost. Three such measures, the Clinically Useful Depression Outcome Scale (CUDOS),[8] Quick Inventory of Depressive Symptomatology (QIDS),[9] and Patient Health Questionnaire (PHQ-9)[10] each recommend cutoff scores to distinguish patients with mild, moderate, and severe depression.

Because of the significance accorded severity by treatment guidelines, it is important to compare different scales on the basis of their allocation of patients into severity groups. If scales markedly differ in the distribution of patients into severity categories, then this would pose a problem to clinicians who wish to use such scales to inform treatment selection. Accordingly, in the present report from the Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS) project, we compared 3 self-report scales that assess the *DSM-IV* symptom criteria for MDD on the basis of how these scales distribute patients into severity categories.

© 2012 COPYRIGHT PHYSICIANS POSTGRADUATE PRESS, INC. NOT FOR DISTRIBUTION, DISPLAY OR COMMERCIAL PURPOSES.

METHOD

As part of an ongoing study of the validity of a new measure to assess remission from depression conducted at Rhode Island Hospital, Providence, Rhode Island, from June 2010 to November 2011, 245 outpatients with *DSM-IV* MDD who presented for treatment or who were in ongoing treatment and had their medication changed due to lack of efficacy completed the CUDOS, PHQ-9, and QIDS at the initiation of treatment and were evaluated with the 17-item HDRS (HDRS-17) blind to the completion of the self-report scales. The sample included 74 men (30.2%) and 171 women (69.8%) who ranged in age from 18 to 79 years (mean = 41.9, standard deviation [SD] = 13.1). Approximately two-fifths of the subjects were married (41.2%, n = 101); the remainder were single (25.3%, n = 62), divorced (11.8%, n = 29), separated (6.9%, n = 17), widowed (2.4%, n = 6), or living with someone as if in a marital relationship (12.2%, n = 30). More than half of the patients attended school beyond high school (57.6%, n = 141), although only one-third graduated from a 4-year college (34.3%, n = 84). The racial composition of the sample was 77.1% white (n = 189), 9.4% black (n = 23), 9.4% Hispanic (n = 23), 1.6% Asian (n = 4), and 2.4% other (n = 6).

The CUDOS contains items assessing all of the *DSM-IV* inclusion criteria for MDD. The respondent is instructed to rate the symptom items 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). Compound *DSM-IV* symptom criteria referring to more than 1 construct (eg, problems concentrating or making decisions, insomnia, or hypersomnia) were subdivided into their respective components, and a CUDOS item was written for each component. Total scores range from 0 to 64. In the original study[8] of the scale's validity, score ranges were empirically derived corresponding to depression severity categories: no depression, 0 to 10; minimal depression, 11 to 20; mild depression, 21 to 30; moderate depression, 31 to 45; and severe depression, 46 and above.

Similar to the CUDOS, the QIDS uses 16 items to assess the *DSM-IV* MDD symptom criteria. However, the format of the 2 questionnaires differs. On the QIDS, each symptom is assessed by a group of 4 statements, and the respondent selects the item that best describes how he or she has been feeling. Not every item contributes to the total score. In scoring the QIDS, the highest score is used of the 4 items assessing sleep disturbance (initial, middle, or terminal insomnia or hypersomnia), the 2 items assessing psychomotor disturbance (agitation, retardation), and the 4 items assessing appetite and weight disturbance. Total scores on the scale range from 0 to 27, and the recommended severity score ranges are no depression, 0–5; mild depression, 6–10; moderate depression, 11–15; severe depression, 16–20; and very severe depression, 21–27.[11]

Clinical Points

- Treatment guidelines for depression suggest that severity be taken into account when initiating treatment.
- There is a marked disparity between the way standardized scales classify depressed outpatients into severity groups. This disparity suggests that using such scales to classify severity subtypes is problematic.
- Recommending any one scale to measure depression severity is premature.

The PHQ-9 contains 9 items corresponding to the *DSM-IV* MDD criteria. Unlike the CUDOS and QIDS, the PHQ-9 assesses compound symptom criteria with a single item. For example, the PHQ-9 assesses insomnia and hypersomnia, and reduced or increased appetite, with a single item. The respondent is instructed to rate the symptom items on a 4-point Likert scale indicating how often he or she has been bothered by the symptom over the past 2 weeks (0 = not at all; 1 = several days; 2 = more than half the days; 3 = nearly every day). Total scores on the scale range from 0 to 27, and recommended severity score ranges are no depression, 0–4; mild depression, 5–9; moderate depression, 10–14; moderately severe depression, 15–19; and severe depression, 20–27.[10]

Statistical Analysis

Each of the 3 scales subdivides patients into 5 severity categories, although they do so in different ways. The CUDOS has an extra category at the lower end of severity by distinguishing between the absence of clinically significant depression and minimal depression. In contrast, the QIDS and PHQ-9 have an extra category at the severe end of the severity continuum. The PHQ-9 distinguishes between moderately severe and severe depression, whereas the QIDS distinguishes between severe depression and very severe depression. In our analyses, we collapsed the 5 groups into 4. For the CUDOS, we combined the minimally depressed group with the mild depression group, because minimal depression better reflects the lower end of the mild depression category than the absence of depression. For the QIDS, we combined the 2 highest groups (severe and very severe) into the severe group. Similarly, for the PHQ-9, we also combined the 2 highest groups (moderately severe and severe) into the severe group. We used the McNemar test to compare the percentage of patients classified as severe on the self-report measures.

The cutoff scores to identify severity groups on the 17-item HDRS have varied. Experts on the treatment of severe depression have generally been consistent in recommending a cutoff of 25.[12–14] Similarly, there is a relative consensus in the field that patients scoring 7 and below are considered to be in remission, and we therefore used this threshold to define the no depression group.[15] Because a 17-item HDRS score of 18 is the most commonly used threshold for inclusion in antidepressant treatment trials, we used this cutoff to distinguish between mild and moderate depression (ie, scores of 8–17 indicating mild depression and 18–24 indicating moderate depression).

© 2012 COPYRIGHT PHYSICIANS POSTGRADUATE PRESS, INC. NOT FOR DISTRIBUTION, DISPLAY, OR COMMERCIAL PURPOSES

RESULTS

For the HDRS-17 and CUDOS, the mean score of the 245 patients fell in the moderate range (CUDOS: mean = 34.7, SD = 11.6; HDRS-17: mean = 20.3, SD = 6.0). In contrast, the mean scores on the PHQ-9 (mean = 17.1, SD = 5.6) and QIDS (mean = 15.9, SD = 4.6) fell into the moderately severe and severe ranges, respectively. The data in Table 1 show that the correlations between the HDRS-17 and the 3 self-report scale scores were nearly identical, and the mean correlation among the 3 self-report scales was 0.73.

Table 1. Intercorrelation Matrix Among Depression Measures

Scale	HDRS-17	CUDOS	PHQ-9	QIDS
Clinically Useful Depression Outcome Scale (CUDOS)	0.61	1.00		
Patient Health Questionnaire (PHQ-9)	0.61	0.73	1.00	
Quick Inventory of Depressive Symptomatology (QIDS)	0.62	0.73	0.75	1.00

Abbreviation: HDRS-17 = 17-item Hamilton Depression Rating Scale.

Table 2 shows the distribution of patients into severity categories on the depression measures. A small number of patients fell into the nondepressed range on each of the 4 measures. Approximately one-third of the patients scored in the mild range on the HDRS-17 and CUDOS, whereas approximately 10% of the patients were mildly depressed according to the PHQ-9 and QIDS. On the CUDOS and HDRS-17, moderate depression was the most frequent severity category, whereas on the PHQ-9 and QIDS, the majority of the patients were classified as severe. Significantly fewer patients were classified as severely depressed on the CUDOS compared to the PHQ-9 (McNemar = 153.8; $P < .001$) and QIDS (McNemar = 114.0; $P < .001$). Significantly more patients were severe on the PHQ-9 than the QIDS (McNemar = 46.3; $P < .001$). It could be argued that we should have included the moderately severe group with the moderate group rather than the severe group. Had we grouped the patients in this manner, then 51.6% of the patients would have been classified as moderately depressed, and 38.9% would have been classified as severely depressed, a rate that is still higher than the rate of severe depression based on the CUDOS (McNemar = 95.9; $P < .001$), although now lower than the rate of severe depression according to the QIDS (McNemar = 30.5; $P < .001$). Had we subdivided the PHQ-9 moderately severe group into moderate depression (scores of 15, 16, and 17) and severe depression (scores of 18 and 19), then the overall rates of moderate and severe depression would have been 40.2% and 50.4%, respectively.

Table 2. Prevalence of Severity Subtypes According to Different Measures of Depression

	None		Mild		Moderate		Severe	
Scale	n	(%)	n	(%)	n	(%)	n	(%)
Clinically Useful Depression Outcome Scale (CUDOS)[a]	5	(2.1)	80	(33.1)	111	(45.9)	46	(19.0)
Patient Health Questionnaire (PHQ-9)[b]	5	(2.0)	18	(7.4)	52	(21.3)	169	(69.3)
Quick Inventory of Depressive Symptomatology (QIDS)[c]	2	(0.8)	31	(12.8)	81	(33.5)	128	(52.9)
Hamilton Depression Rating Scale, 17-item (HDRS-17)	4	(1.6)	74	(30.2)	106	(43.3)	61	(24.9)

[a]CUDOS data were missing for 3 participants, leaving a final sample of n = 242. Percentages do not sum to 100 due to rounding.
[b]PHQ-9 data were missing for 1 participant, leaving a final sample of n = 244.
[c]QIDS data were missing for 3 participants, leaving a final sample of n = 242.

Table 3. Percentage of Depressed Patients of Mild and Moderate Severity According to the 17-Item Hamilton Depression Rating Scale (HDRS-17) Who Were Classified as Severe on Different Self-Report Measures

	HDRS Mild Depression (n = 74)		HDRS Moderate Depression (n = 106)	
Scale	n	(%)	n	(%)
Clinically Useful Depression Outcome Scale (CUDOS)[a]	1	(1.4)	20	(19.0)
Patient Health Questionnaire (PHQ-9)[b]	27	(37.0)	82	(77.4)
Quick Inventory of Depressive Symptomatology (QIDS)[b]	17	(23.3)	64	(60.4)

[a]CUDOS data were missing for 2 participants with HDRS mild depression and for 1 participant with HDRS-17 moderate depression. The final sample was n = 72 and n = 105 for mild and moderate depression, respectively.
[b]PHQ and QIDS data were missing for 1 participant with HDRS-17 mild depression, leaving a final sample of n = 73.

The majority of the patients in the moderate range on the HDRS-17 were in the severe range on the PHQ-9 and QIDS, whereas less than 20% of these patients scored in the severe range on the CUDOS (Table 3). Of the 74 patients rated in the mild range on the HDRS-17, only 1 scored in the severe range on the CUDOS, whereas approximately one-quarter scored in the severe range on the QIDS, and approximately one-third scored in the severe range on the PHQ-9.

DISCUSSION

Treatment guidelines for depression suggest that it is important to consider severity when selecting a patient's initial treatment modality.[1–3] While it can be debated whether the empirical evidence is sufficient to support one treatment modality over another as a function of depression severity,[16] there should be little debate that there is a problem with recommendations to link treatment selection to severity when severity classification greatly depends on the scale used.

Standardized scales are typically not used in clinical practice.[17,18] In the past few years there have been increasing calls for the utilization of such measures,[1,19,20] and it is likely that self-report scales are more likely to be used than clinician-rated scales such as the HDRS-17. We anticipate future studies examining how well clinicians adhere to official treatment guidelines, and the impact of baseline severity on initial treatment selection is a potential topic of interest. The results of the present study suggest that the scale used to measure severity could have an impact on treatment selection. Measures such as the QIDS and PHQ-9, which broadly define the severe category, could result in fewer psychotherapy referrals and greater reliance on medication as a first-line treatment option.

© 2012 COPYRIGHT PHYSICIANS POSTGRADUATE PRESS, INC. NOT FOR DISTRIBUTION, DISPLAY OR COMMERCIAL PURPOSES.

In the present study, we found significant differences between 3 scales that presumably measure the same construct (ie, the symptom criteria of *DSM-IV* MDD) in the distribution of patients into severity categories. The scales differ somewhat in how they are scored, and the CUDOS and PHQ-9 assess severity in terms of symptom frequency, whereas the QIDS assesses severity in terms of both symptom frequency and symptom intensity. However, the item content is largely the same. What then might account for the marked differences between scales of similar content in the distribution of patients into severity groups?

The cutoff scores on the 3 scales to define the severity groups were derived in different ways. We could not find a definitive article establishing the severity cutoffs on the QIDS. Several authors refer to the 2003 article by Rush et al[9]; however, this study derived QIDS cutoffs corresponding to the definition of remission on the HDRS-17 and did not derive cutoff scores corresponding to severity ranges. In an article published in 2006, Rush et al[11] identified QIDS scores corresponding to severity ranges and noted the correspondence between these QIDS scores and 17-item HDRS scores based on data from their 2003 article. Of note, the 17-item HDRS score used by Rush et al to delineate the lower bound of the severe range was 18, a score that is lower than the usual 17-item HDRS score indicating severe depression.[12–14]

The cutoff scores on the PHQ-9 were chosen for the pragmatic reason of making them easier for clinicians to recall.[10] The authors also noted that alternative cutoffs did not increase the association between increasing PHQ-9 severity and indices of construct validity. When selecting the cutoff scores to define the severity ranges on the PHQ-9, the authors did not consider the potential impact of the broadness by which severity ranges were defined and how this might impact treatment selection based on recommendations of official treatment guidelines.

The severity ranges on the CUDOS were the only ones specifically derived from empirical study.[8] A large sample of psychiatric patients completed the scale and were rated on the Clinical Global Impressions-Severity of Illness (CGI-S) scale.[21] The means and SDs of CUDOS scores were computed for each CGI-S rating, and these values, along with "clinical experience," were used to establish the range of scores for the severity descriptors. The authors did not, however, compute diagnostic efficiency statistics such as sensitivity and specificity to determine the optimal threshold values to define the severity score ranges.

The present study focused on the distribution of patients into severity groups and de-emphasized the issue of validity. To be sure, each of the 3 self-report scales was equally highly correlated with the 17-HDRS, thereby suggesting that each measure was equally valid as a dimensional measure of depression symptoms. In deriving the scoring ranges for the PHQ-9, Kroenke et al[10] suggested that, when severity groupings based on different cutoff scores are equally associated with external variables, the cutoff values can be chosen on the basis of their ease of recall. We disagree with this logic. For all scales measuring the severity of depressive symptoms, the thresholds distinguishing patients with mild, moderate, and severe depression do not represent well-demarcated lines separating the severity subtypes. As with other areas of psychopathology, the severity of depression better corresponds to a dimensional rather than a categorical model of classification.[22] However, the choice of cutoff is important, insofar as it impacts the relative broadness of each of the severity categories. If severity distinctions are used in treatment selection, then the relative broadness of the definitions of mild, moderate, and severe depression will have significant clinical implications.

The majority of patients scoring in the moderate range on the HDRS-17 and a substantial minority of patients scoring in the mild range on the HDRS-17 fell into the severe range on the PHQ-9 and QIDS. In contrast, a small number of patients scoring in the mild and moderate ranges on the HDRS-17 fell into the severe range on the CUDOS. In light of the equally high correlations between all 3 self-report scales and the HDRS-17, we do not believe that the CUDOS is any more valid than the PHQ-9 and QIDS as a measure of severity but rather that the cutoff scores on these latter 2 scales need to be adjusted to more accurately identify severely depressed patients.

Each of the 3 questionnaires that we examined delineated 5 categories of severity; however, we collapsed these 5 categories into 4 to facilitate comparison between the scales. All 3 scales identify a mild, moderate, and severe group. The CUDOS distinguishes between no depression and minimal depression. This distinction was made because research from our clinical research group has found that the presence of mild residual symptoms in patients who are considered to be in remission on the HDRS-17 is associated with increased psychosocial impairment and reduced quality of life.[23,24] We combined the minimal depression group with the mild depression group and thus broadened the mild category. The QIDS distinguishes between severe and very severe depression, and we combined these 2 groups. The PHQ-9 defined a moderately severe group, lying between the moderate and severe groups. It is not clear why this group was identified other than to make it possible to retain 5-point scoring ranges for the qualitative descriptors (eg, 0–4 vs 5–9 vs 10–14). When we distributed the patients in the moderately severe group into moderate and severe groups rather than combining them with the severe group, the PHQ-9 still classified more patients as severe than the other measures.

Before concluding, the limitations of the study should be considered. The present study was conducted in a single clinical practice in which the majority of the patients were white and female and had health insurance. Replication in samples with different demographic characteristics is warranted. However, the generalizability of the findings is enhanced by the lack of inclusion and exclusion criteria to select patients. The focus of the study was on the distribution of patients into severity categories and not on validity. Future research should examine if the scales are equally valid in predicting treatment outcome and other clinically relevant constructs such as psychosocial morbidity. In the present

© 2012 COPYRIGHT PHYSICIANS POSTGRADUATE PRESS, INC. NOT FOR DISTRIBUTION, DISPLAY, OR COMMERCIAL PURPOSES

study, severity was defined according to scores on symptom severity measures. Other methods, such as hospitalization, presence of melancholia, suicidality, psychosis, and level of functional impairment, have also been used as indicators of severity.[15] There are advantages and disadvantages to each of these approaches in determining the severity of depression, although scores on standardized rating scales have been the most commonly used index of severity in research studies. This approach might contrast with how severity is measured in clinical practice, where the focus might be more on functional impairment or suicidality. Also, clinicians might differentially weight symptoms in determining severity, whereas most scales weight symptoms similarly. The study was limited to 4 scales—the HDRS-17 and 3 self-report scales. Future studies of the comparability of measures in classifying severity should also include the CGI-S, a simple, widely used global measure of severity that probably most closely corresponds to how clinicians classify severity, at least informally, in their practices.

In conclusion, if clinicians are to follow treatment guidelines' recommendations regarding the impact of severity on initial treatment selection, then it is important to apply a consistent method of determining depression severity. The marked disparity between standardized scales in the classification of depressed outpatients into severity groups is disconcerting and indicates that there is a problem with the use of such instruments to classify depression severity. While we agree with recommendations to use quantitative measures of depression in clinical practice, we also caution against the use of these scales to guide treatment selection until the thresholds to define severity ranges have been well established empirically. It is important for the developers of depression measures not to be cavalier in recommending thresholds corresponding to severity levels of depression because of the potential implications of symptom severity on initial treatment selection. The *DSM-5* Work Group for Mood Disorders is considering recommending the PHQ-9 to measure depression severity. We believe that it is premature to recommend any one scale to measure depression severity, especially one that lacks empirically derived thresholds to identify grades of severity.

Author affiliations: Department of Psychiatry and Human Behavior, Brown Medical School, and the Department of Psychiatry, Rhode Island Hospital, Providence.
Potential conflicts of interest: None reported.
Funding/support: The research was supported, in part, by Eli Lilly USA, LLC.

REFERENCES

1. American Psychiatric Association. *Practice Guideline for the Treatment of Patients With Major Depressive Disorder*. 3rd ed. Washington, DC: American Psychiatric Association; 2010.
2. National Center for Mental Health and Clinical Excellence. Depression: The treatment and management of depression in adults. In: *National Institute for Health and Clinical Excellence*, ed. London, England: National Institute for Health and Clinical Excellence; 2009:64.
3. van der Lem R, van der Wee NJ, van Veen T, et al. The generalizability of antidepressant efficacy trials to routine psychiatric out-patient practice. *Psychol Med*. 2011;41(7):1353–1363.
4. 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.
5. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960;23(1):56–62.
6. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 1979;134(4):382–389.
7. Zimmerman M, McGlinchey JB, Chelminski I. An inadequate community standard of care: lack of measurement of outcome when treating depression in clinical practice. *Prim Psychiatry*. 2008;15:67–75.
8. Zimmerman M, Chelminski I, McGlinchey JB, et al. A clinically useful depression outcome scale. *Compr Psychiatry*. 2008;49(2):131–140.
9. 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(5):573–583.
10. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606–613.
11. Rush AJ, Bernstein IH, Trivedi MH, et al. An evaluation of the Quick Inventory of Depressive Symptomatology and the Hamilton Rating Scale for Depression: a sequenced treatment alternatives to relieve depression trial report. *Biol Psychiatry*. 2006;59(6):493–501.
12. Hirschfeld RM. Efficacy of SSRIs and newer antidepressants in severe depression: comparison with TCAs. *J Clin Psychiatry*. 1999;60(5):326–335.
13. Montgomery SA, Lecrubier Y. Is severe depression a separate indication? ECNP Consensus Meeting; September 20, 1996; Amsterdam. European College of Neuropsychopharmacology *Eur Neuropsychopharmacol*. 1999;9(3):259–264.
14. Schatzberg AF. Antidepressant effectiveness in severe depression and melancholia. *J Clin Psychiatry*. 1999;60(suppl 4):14–21, discussion 22.
15. Frank E, Prien RF, Jarrett RB, et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder: remission, recovery, relapse, and recurrence. *Arch Gen Psychiatry*. 1991;48(9):851–855.
16. Driessen E, Cuijpers P, Hollon SD, et al. Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? a meta-analysis. *J Consult Clin Psychol*. 2010;78(5):668–680.
17. Zimmerman M, McGlinchey JB. Why don't psychiatrists use scales to measure outcome when treating depressed patients? *J Clin Psychiatry*. 2008;69(12):1916–1919.
18. Gilbody SM, House AO, Sheldon TA. Psychiatrists in the UK do not use outcomes measures: national survey. *Br J Psychiatry*. 2002;180(2):101–103.
19. Morris DW, Trivedi MH. Measurement-based care for unipolar depression. *Curr Psychiatry Rep*. 2011;13(6):446–458.
20. Harding KJ, Rush AJ, Arbuckle M, et al. Measurement-based care in psychiatric practice: a policy framework for implementation. *J Clin Psychiatry*. 2011;72(8):1136–1143.
21. Guy W. *ECDEU Assessment Manual for Psychopharmacology*. Rockville, MD: National Institute of Mental Health; 1976.
22. Ruscio J, Zimmerman M, McGlinchey JB, et al. Diagnosing major depressive disorder XI: a taxometric investigation of the structure underlying *DSM-IV* symptoms. *J Nerv Ment Dis*. 2007;195(1):10–19.
23. Zimmerman M, Posternak MA, Chelminski I. Is the cutoff to define remission on the Hamilton Rating Scale for Depression too high? *J Nerv Ment Dis*. 2005;193(3):170–175.
24. Zimmerman M, Martinez J, Attiullah N, et al. Further evidence that the cutoff to define remission on the 17-item Hamilton Depression Rating Scale should be lowered. *Depress Anxiety*. 2012;29(2):159–165.

© 2012 COPYRIGHT PHYSICIANS POSTGRADUATE PRESS, INC. NOT FOR DISTRIBUTION, DISPLAY OR COMMERCIAL PURPOSES.