Defining Remission on the Montgomery-Asberg Depression Rating Scale

Mark Zimmerman, M.D.; Michael A. Posternak, M.D.; and Iwona Chelminski, Ph.D.

Objective: In antidepressant efficacy trials, it is common to define treatment remission as a score below a cutoff on symptom severity measures. No consensus has emerged regarding an appropriate cutoff for defining remission on the Montgomery-Asberg Depression Rating Scale (MADRS). The goal of the present study was to establish an empirically based cutoff on the MADRS for defining remission.

Method: 303 psychiatric outpatients being treated for a DSM-IV major depressive episode were rated on the Standardized Clinical Outcome Rating for Depression, an index of DSM-IV remission status; the MADRS; and the Global Assessment of Functioning (GAF) scale. We examined the sensitivity, specificity, and overall classification rate of the MADRS for identifying a broad and narrow interpretation of the DSM-IV definition of remission, as well as the association between the breadth of the definition of remission and psychosocial functioning.

Results: On the basis of a narrow definition of remission, which requires a complete absence of clinically significant symptoms of depression, the optimal MADRS cutoff was ≤ 4. On the basis of a broader definition, the optimal cutoff was ≤ 9. Compared with patients scoring 5 through 9 on the MADRS, those who met the narrow definition of remission were rated higher, indicating better functioning, on the GAF and reported significantly less psychosocial impairment ($p < .05$).

Conclusion: Our results support the use of a low cutoff on the MADRS to define remission. Because the choice of cutoff will impact the percentage of patients who are considered to be in remission and thus impact conclusions about treatment effectiveness, more empirical study should be directed toward this issue.

(J Clin Psychiatry 2004;65:163–168)

Received March 15, 2003; accepted July 21, 2003. From the Department of Psychiatry and Human Behavior, Brown University School of Medicine, Providence, R.I.

The authors report no financial affiliation or other relationship relevant to the subject matter of this article.

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

In antidepressant efficacy trials, outcome is typically measured on standardized instruments, the 2 most common being the Hamilton Rating Scale for Depression (HAM-D)[1] and the Montgomery-Asberg Depression Rating Scale (MADRS).[2] For the past 30 years, the HAM-D has been the most widely used outcome measure in antidepressant efficacy trials,[3] although during the past decade, the MADRS has been increasingly used.[4] Whereas the HAM-D was intended as a measure of the severity of depressive symptoms, the MADRS was designed to be particularly sensitive to change in patients treated with antidepressant medication.

In characterizing treatment outcome in antidepressant efficacy trials, it is common to define treatment *response* as a 50% or more improvement in score on the HAM-D or MADRS and treatment *remission* as a score below a predetermined cutoff score on the scale. Through the years, many cutoff scores have been used on the HAM-D to define remission[5]; however, since the publication of the recommendations of Frank and colleagues,[6] a consensus has emerged to define remission on the HAM-D as a score of 7 or less on the 17-item version of the HAM-D.

No such consensus has yet emerged in defining remission on the MADRS. Investigators using the MADRS to define remission have used cutoff scores ranging from 6 to 12.[7–12] The lack of convention in defining remission on the MADRS is problematic because it creates difficulty in combining and contrasting findings across studies. Also, the variability of choosing among multiple cutoff scores to define remission leaves open the possibility that investigators examine different threshold scores to define remission and report only the most favorable findings.

To date, 3 groups of investigators have recommended thresholds on the MADRS to define remission on the basis of the association between the MADRS and ratings on a global severity index or the HAM-D, although each study had methodological limitations. Mittmann et al.[13] collected 262 MADRS–HAM-D pairs of ratings in 77 depressed outpatients, conducted a regression analysis to derive a formula for converting scores on the 17-item HAM-D into MADRS scores, and found that a MADRS cutoff of ≤ 8 was equivalent to the HAM-D remission definition of ≤ 7. However, as the authors noted, patients contributed multiple data points; thus, most of the rating couplets were not statistically independent.

Snaith et al.[14] rated 80 depressed inpatients and outpatients on the MADRS and an 11-point severity scale that was collapsed into 4 categories, 1 of which represented recovered/absence of symptoms. They derived MADRS score ranges corresponding to each severity range on the basis of the median scores of successive ranges and reported that a MADRS score of 6 was the upper limit of the recovered range. Perhaps because the sample included psychiatric inpatients as well as outpatients, the sample size of the recovered group was small (N = 13). Also, they defined *recovered* as no need for treatment, rather than as an absence of symptoms.

Most recently, Hawley and colleagues[15] made 1114 pairs of Clinical Global Impressions-Severity of Illness (CGI-S) and MADRS ratings in 684 outpatients. They analyzed their data in 2 ways. First, they examined patients who were rated as 1 (i.e., not ill) and as 3 or more (i.e., ill) on the CGI-S and determined the number of patients who would be misclassified as remitters according to different MADRS thresholds. On the basis of this analysis, the lowest misclassification rate was found for a cutoff score of ≤ 8. In a second analysis, the distribution of MADRS scores was examined in patients rated 2 on the CGI-S (i.e., borderline ill). The authors hypothesized that the optimal remission cutoff would bisect this distribution and found that the best cutoff was ≤ 9. They therefore concluded that remission on the MADRS should be defined as a score of ≤ 8 or ≤ 9.

Questions can be raised, however, about how the evaluators in this study[15] rated the CGI-S. It is difficult to reconcile a rating of 1 on the CGI-S, which reflects an absence of depressive symptoms, with a MADRS score of 10 or more, as was necessary for more than 10% of the patients. The authors did not describe how they interpreted the CGI-S rating points, and they did not report reliability data. The relatively high frequency of patients rated 1 on the CGI-S who scored 10 or more on the MADRS suggests that the authors rated patients as 1 on the CGI-S in the presence of low levels of depression symptoms. In our interpretation of the CGI-S, we would rate patients with mild symptoms of depression as at least 2. In fact, the authors presented some data suggesting that they might have sometimes underrated the CGI-S.

In a separate sample of 200 patients, the authors compared clinician and patient ratings on the CGI-S and found higher patient ratings in nearly 40%, whereas the clinician rating was higher only 10% of the time.[15] This is consistent with the hypothesis that the clinicians in their study tended to rate CGI-S scores on the low side.

A potential problem with using the MADRS to define remission is that it does not include some of the defining features of major depressive disorder (MDD). Specifically, the MADRS does not assess reverse vegetative symptoms such as increased appetite and hypersomnia or indecisiveness. It is therefore possible for patients to have several symptoms of depression yet achieve low scores on the MADRS. This might be more likely to occur in patients with atypical depression, which is characterized by overeating and oversleeping.[16]

The definition of remission is addressed in the introductory chapter to DSM-IV on the use of the manual. DSM-IV indicates that a disorder is in full remission when "there are no longer any symptoms or signs of the disorder."[17(p2)] The mood disorders section of DSM-IV is 1 of 2 sections (the other is substance dependence) that provide additional guidelines to determine remission status. The DSM-IV remission definition for MDD requires the absence of "*significant* [italics added] signs or symptoms of the disturbance."[17(p378)] No guidelines are provided for interpreting the meaning of "significant."

In an unpublished report from the Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS) project, we described the reliability and validity of the Standardized Clinical Outcome Rating Scale for Depression (SCOR-D, available from the authors by request), an outcome measure that was designed to reflect the DSM-IV definition of remission. In the present report, we compare MADRS scores with ratings on the SCOR-D to determine the cutoffs on the MADRS that correspond to narrow and broad interpretations of the DSM-IV definition of remission. After identifying the optimal cutoff points for broad and narrow definitions of remission on the MADRS, we compared the psychosocial functioning and quality-of-life ratings in patients who met the 2 definitions of remission according to the MADRS.

METHOD

Participants were 303 psychiatric outpatients who were being treated by the first 2 authors (M.Z., M.A.P.) for a DSM-IV major depressive episode in the Rhode Island Hospital Department of Psychiatry outpatient practice (Providence). Other than a diagnosis of a major depressive episode, there were no inclusion criteria, and no exclusion criteria were used. Thus, the patients represent a consecutive series of outpatients being treated for depression. This private practice group predominantly treats individuals with medical insurance (including Medicare but not Medicaid) 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.

The sample included 114 men (37.6%) and 189 women (62.4%) who ranged in age from 18 to 79 years (mean = 42.9, SD = 12.7). Almost half of the subjects were married (47.9%, N = 145); the remainder were single (23.4%, N = 71), divorced (19.8%, N = 60), separated (5.6%, N = 17), widowed (2.0%, N = 6), or living with someone as if in a marital relationship (1.3%, N = 4). The racial composition of the sample was 86.8% white

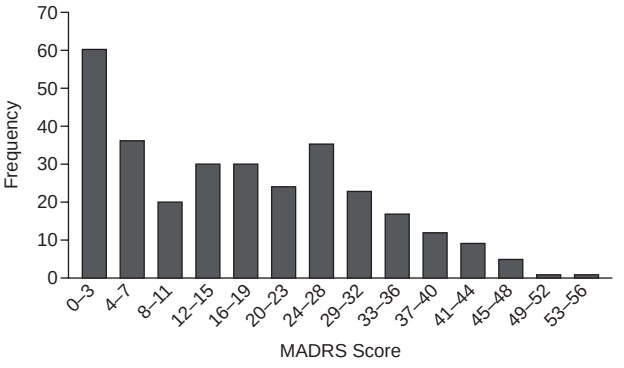
(N = 263), 2.6% black (N = 8), 4.3% Hispanic (N = 13), 0.7% Asian (N = 2), and 5.6% other (N = 17). The Rhode Island Hospital institutional review committee approved the research protocol, and all patients provided informed, written consent.

Although we did not systematically record the treatment received at the time of the evaluation, almost all patients were being treated with antidepressant medication. Patients may have also been treated with concomitant hypnotic agents, mood stabilizers, benzodiazepines, or psychotherapy.

The patients were rated by the first 2 authors (M.Z., M.A.P.) on the MADRS,[2] the SCOR-D (described below), and the DSM-IV Global Assessment of Functioning (GAF) scale.[17] Approximately one third of the patients completed a questionnaire that included a question regarding functional impairment due to depression ("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") and quality of life ("How would you rate your overall quality of life? 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."). Only a minority of the patients completed the questionnaire because it was introduced late in the study. Diagnoses were based on the Structured Clinical Interview for DSM-IV.[18] Interrater reliability on the MADRS and SCOR-D was obtained in 16 patients, with one of the authors interviewing the patient while the other observed and made independent ratings. For the MADRS, the intraclass correlation coefficient was .96.

As part of the MIDAS project, we have modified and expanded upon the Psychiatric Status Ratings used in the Longitudinal Interval Follow-Up Evaluation (LIFE).[19] SCORs have been developed for the most common DSM-IV disorders presenting for treatment in outpatient practice. The SCOR-D is a 6-point rating scale based on the number of DSM-IV criterion symptoms for a major depressive episode and level of psychosocial impairment present during the past week. The Collaborative Depression Study, for which the LIFE was developed, has alternately defined recovery broadly (a LIFE rating of 1 or 2) and narrowly (a LIFE rating of 1).[20] (Recovery and remission are not synonymous, differing in terms of the persistence of the period of improvement. From a cross-sectional perspective, however, the clinical status of patients in remission and recovery is the same.) Patients rated 2 on the SCOR-D may have 1 or 2 symptoms of depression to a mild degree, or they may not have any symptoms meeting the threshold to indicate symptom presence (i.e., the symptom is not present daily), but they do not consider themselves to be back to their normal selves. Patients rated 1 on the SCOR-D have no clinically significant symptoms of depression and they judge that they have returned to their normal selves. Thus, the broad definition of remission was defined as a SCOR-D rating of 1 or 2, and the narrow definition of remission was defined as a SCOR-D rating of 1. The intraclass reliability coefficient for the SCOR-D dimensional rating was .95. Dichotomizing the ratings as remitted or not, the kappa coefficient of agreement was 1.0 for the broad definition of remission and 0.71 for the narrow definition.

Figure 1. Distribution of Montgomery-Asberg Depression Rating Scale (MADRS) Scores in 303 Depressed Outpatients



The association between the MADRS and SCOR-D was examined in several ways. With the SCOR-D as the independent variable, an analysis of variance was conducted on the MADRS scores, with Tukey follow-up tests comparing each adjacent level of the SCOR-D. We also computed the Pearson correlation between the SCOR-D and the MADRS. We examined the performance of the MADRS as a measure of remission across the range of cutoff scores by conducting receiver operating curve (ROC) analyses.[21] An ROC is a plot of a measure's sensitivity versus 1 minus specificity at each cutoff score. The area under this curve (AUC) is the evaluative measure, which can range from 0.5 (random performance) to 1.0 (perfect performance). We plotted separate ROCs for the broad and narrow DSM-IV definitions of remission. For both the broad and narrow definitions of remission on the SCOR-D, we examined the sensitivity, specificity, and overall correct classification rate associated with different MADRS values.

RESULTS

The mean score on the MADRS for the entire sample was 17.2 (SD = 12.7). Following the approach of Hawley and colleagues,[15] we grouped the MADRS scores into 4-point intervals, and the distribution of ratings is provided in Figure 1.

Slightly more than one third (N = 114) of the sample met MDD criteria (SCOR-D rating of 5 or 6) at the time

Table 1. Montgomery-Asberg Depression Rating Scale (MADRS) Mean Values for Each of the Standardized Clinical Outcome Rating Scale for Depression (SCOR-D) Ratings[a]

SCOR-D Rating	N	MADRS Score Mean	MADRS Score SD
6	11	43.2	5.1
5	103	28.4	6.8
4	35	21.3	4.5
3	52	13.8	3.7
2	50	5.3	2.8
1	52	1.5	2.2

[a]$F = 347.6$, $df = 5,297$; $p < .001$. All follow-up Tukey tests between each adjacent SCOR-D level were significant at $p < .01$.

of the evaluation, slightly less than one third (N = 87) was in partial remission (rating of 3 or 4), and one third (N = 102) was in remission according to the broad definition (rating of 1 or 2). The Pearson correlation between the MADRS and the SCOR-D was .91. The overall ANOVA was significant ($F = 347.6$, $df = 5,297$; $p < .001$), and Tukey follow-up tests found that the difference on the MADRS between each adjacent level of severity on the SCOR-D was significant (Table 1).

The AUCs were significant for both definitions of remission (broad definition: AUC = .99, $p < .001$; narrow definition: AUC = .97, $p < .001$). The sensitivity, specificity, and overall classification rate of the MADRS for identifying remission according to the broad and narrow definitions of remission are presented in Table 2. A cutoff of ≤ 9 on the MADRS maximized the level of agreement with the broad definition of remission, although the agreement rates associated with cutoffs of 7, 8, and 10 were very similar. Based on a narrower definition of remission, which requires the absence of clinically significant symptoms of depression, the optimal MADRS cutoff was ≤ 4.

We compared the GAF scores and responses to the self-reported quality of life and psychosocial impairment questions in patients whose scores on the MADRS were 0 through 4 (i.e., remitted according to the narrow definition) and 5 through 9 (i.e., remitted according to the broad but not the narrow definition). Patients who met the narrow definition of remission on the MADRS were rated higher, indicating better functioning and fewer symptoms, on the GAF (73.5 ± 5.7 vs. 68.1 ± 5.5; $t = 4.56$, $df = 100$, $p < .005$) and reported significantly less impairment on the self-rated question (0.3 ± 0.6 vs. 0.8 ± 0.8; $t = 2.26$, $df = 30$, $p < .05$). There was no difference between the groups on the quality-of-life rating.

DISCUSSION

Variation in the cutoff used to define remission on scales such as the MADRS creates practical and theoretical problems. At a practical level, it is difficult to assimilate the treatment literature when different cutoffs are used. Cross-study comparisons of absolute rates of remission are hampered when markedly different thresholds are used to define remission.

The failure to establish an accepted cutoff to define remission also opens the possibility of data mining. Researchers can analyze their data using multiple cutoffs, report only the most favorable results, and find a prior study to cite as justification for their choice of cutoff to define remission. The implications of this possibility were illustrated several years ago by one of our group who examined the association between melancholic subtyping and response to electroconvulsive therapy and demonstrated that the choice of cutoff score to define treatment response impacted the conclusions that could be drawn[5]; that is, whether the findings were statistically significant depended on the cutoff point chosen. In that article, it was pointed out that a misleading conclusion could have been drawn had the presentation of the results been limited to the single cutoff point associated with a significant finding.

From a statistical perspective, the problem with analyzing more than one cutoff point is that multiple statistical tests are being conducted without a corresponding adjustment in the alpha level. Consequently, this approach increases the possibility of type I error.

In the present study, we examined the relationship between MADRS scores and 2 interpretations of the DSM-IV definition of remission. The cutoff score for the broader definition was more than twice as high as the cutoff for the narrow definition (9 vs. 4). The choice of cutoff will obviously have an impact on the percentage of patients who are considered to be in remission. In our sample, the respective remission rates were 33.7% and 22.4%. The question is, which definition is more valid?

We are aware of only 1 study that has compared the validity of different definitions of response and remission on symptom severity rating scales. Riso and colleagues[22] examined the prognostic validity of multiple definitions of response based on HAM-D and CGI ratings in 90 outpatients treated with 16 weeks of cognitive-behavioral therapy. In one of their analyses, they examined the association between response at 6 weeks of treatment and outcome at the end of 16 weeks of treatment. Compared with patients who scored 7 through 10 on the HAM-D at week 6, patients who scored ≤ 6 were less depressed, had fewer dysfunctional attitudes, and scored higher on the GAF at week 16.

Although little systematic study has compared the validity of different cutoff scores on the HAM-D and MADRS to define remission, there are several studies suggesting that residual symptoms in patients who have been identified as treatment responders place those patients at greater risk for relapse. For example, Paykel and colleagues[23] followed 64 treatment responders for 15 months. Treatment response was defined as failure to

Table 2. Sensitivity and Specificity of the Montgomery-Asberg Depression Rating Scale (MADRS) as a Measure of Remission

MADRS Cutoff Score	Cumulative N	Broad Definition of Remission (N = 102)[a]				Narrow Definition of Remission (N = 52)[b]			
		Sensitivity, %	Specificity, %	Overall % Correct Rate	Kappa Value	Sensitivity, %	Specificity, %	Overall % Correct Rate	Kappa Value
0	23	22.5	100.0	73.9	.28	40.4	99.2	89.1	.51
≤ 1	34	33.3	100.0	77.6	.40	57.7	98.4	91.4	.65
≤ 2	52	51.0	100.0	83.5	.58	80.8	96.0	93.4	.77
≤ 3	60	58.8	100.0	86.1	.66	88.5	94.4	93.4	.78
≤ 4	68	66.7	100.0	88.8	.73	96.2	92.8	93.4	.79
≤ 5	78	75.5	99.5	91.4	.80	98.1	89.2	90.7	.72
≤ 6	88	85.3	99.5	94.7	.88	98.1	85.3	87.4	.65
≤ 7	96	91.2	98.5	96.0	.91	98.1	82.1	84.8	.60
≤ 8	99	93.1	98.0	96.4	.92	98.1	80.9	83.8	.58
≤ 9	102	95.1	97.5	96.7	.93	98.1	79.7	82.8	.56
≤ 10	106	96.1	96.0	96.0	.91	98.1	78.1	81.5	.54
≤ 11	116	98.0	92.0	94.0	.87	98.1	74.1	78.2	.48
≤ 12	121	98.0	89.6	92.4	.84	98.1	72.1	76.6	.47
≤ 13	126	99.0	87.6	91.4	.82	98.1	70.1	74.9	.44
≤ 14	135	100.0	83.6	89.1	.77	100.0	66.9	72.6	.41
≤ 15	146	100.0	78.1	85.5	.71	100.0	62.5	69.0	.36
≤ 16	148	100.0	77.1	84.8	.69	100.0	61.8	68.3	.36
≤ 17	155	100.0	73.6	82.5	.65	100.0	59.0	66.0	.33
≤ 18	164	100.0	69.2	79.5	.60	100.0	55.4	63.0	.30
≤ 19	176	100.0	63.2	75.6	.54	100.0	50.6	59.1	.26
≤ 20[c]	178	100.0	62.2	74.9	.52	100.0	49.8	58.4	.25

[a]A Standardized Clinical Outcome Rating for Depression (SCOR-D) rating of 1 or 2.
[b]A SCOR-D rating of 1.
[c]For MADRS cutoff scores above 20, sensitivity continues to equal 100%; specificity, overall correct classification rate, and kappa values continue to decrease as the cutoff increases.

meet full major depression criteria for 2 months. Patients who scored above 8 on the HAM-D were 3 times more likely to relapse during the follow-up interval than were patients scoring 7 or below (76% vs. 25%). Thase et al.[24] followed 48 depressed patients who responded to 16 weeks of cognitive-behavioral therapy for 1 year after the completion of treatment. Responders scored 10 or less on the HAM-D and their scores improved at least 50% from baseline. The responders were subdivided into those who did and did not score 6 or less on the HAM-D for the last 2 months of treatment. Patients who scored 6 or less were 5 times less likely to relapse than were patients who scored 7 through 10 (9% vs. 52%). Van London and associates[25] conducted a 3- to 5-year follow-up study of 56 depressed patients and defined remission according to the number of DSM-III-R symptoms present and their score on the MADRS. Patients with residual symptoms were at significantly increased risk of relapse compared with patients in full remission. Other follow-up studies have similarly found that the presence of residual symptoms in patients who responded to treatment predicted poorer outcome.[19,26]

One of the goals of a definition of remission is to predict future morbidity; that is, a more valid definition of remission should be more likely to identify patients who are less likely to relapse than should a less valid definition of remission. This perspective is analogous to how treatment goals for hypertension and hypercholesterolemia are derived (i.e., the prediction of future adverse health events). If a definition of remission should have prognostic significance, then the studies demonstrating poorer outcome in treatment responders with residual symptoms suggest that remission should be defined narrowly. Exactly how low the threshold on the MADRS (or HAM-D) should be to define remission is uncertain because these studies did not examine whether patients scoring very low (e.g., 0–3) differed from patients scoring in the minimal symptom range (e.g., 4–7).

Another method of validating the threshold to define symptomatic remission is to examine its association with concurrent and future psychosocial functioning. We found that patients who met the narrow definition of remission on the MADRS reported less concurrent impairment than did patients who met the broad, but not the narrow, definition of remission. Previous authors have suggested that functional status may improve more slowly than symptoms.[27] An alternative explanation of these findings is that patients who were identified as being symptomatically recovered were in fact a heterogeneous group, and, when the recovered group is defined more narrowly, then improvement in functional status corresponds to symptom improvement. We believe that a conceptually valid definition of remission should reflect both symptomatic and functional remission. The findings of the present study therefore suggest that remission on the MADRS be defined as a score of 4 or less.

Yet another approach toward deriving a valid cutoff score for defining remission is to determine whether a patient's level of symptoms falls within the normal range of values after treatment. We recently conducted a literature review of studies of the MADRS in healthy con-

trols to determine the normal range of values.[28] We identified 10 studies of 14 samples that included data on the MADRS for 569 controls. Across all studies the weighted mean MADRS score, adjusting for sample size, was 4.0 (95% confidence interval = 3.5 to 4.5). Thus, in a sample of healthy controls, as well as within a sample of remitted depressed patients, there are some fluctuations of mild symptom levels. The presence of such mild perturbations is not inconsistent with the concept of remission; consequently, remission is not defined so restrictively that a score of 0 is required on the MADRS.

Before concluding, 5 limitations of the present study require discussion. First, the sample was drawn from a single, large, general adult outpatient private practice setting in which the majority of the patients were white, female, and in their 30s and 40s. The generalizability to samples with different demographic characteristics needs to be demonstrated.

Second, medication side effects and comorbid medical and psychiatric disorders might result in false positive symptom ratings and elevate scores on both the MADRS and SCOR-D. We followed the common practice of making ratings without regard to possible etiology. The advantage of this approach is that it improves reliability and is consistent with how these ratings are generally made. The disadvantage is that validity might be reduced and the percentage of patients who are truly in remission is underestimated. Comorbidity might delay improvement in functioning as well as falsely elevate symptom scores. Unfortunately, we did not record comorbid conditions and thus are unable to evaluate the impact of comorbidity on the ratings.

A third limitation is that only a subset of patients completed the measure of functioning, although there were no differences between patients who did and did not complete the ratings.

Fourth, we did not do a prospective follow-up study in order to examine validity from the perspective of prediction of relapse and future morbidity.

And finally, the sample size may have been too small to detect differences in validity between the individual cutoff scores on the MADRS. While the overall sample size was respectable, the sample size at most of the individual MADRS scores was only approximately 10. Future studies of the validity of different cutoff scores should focus on patients who score relatively low on the MADRS (e.g., below 15) in order to generate sufficiently sized samples at each MADRS score.

REFERENCES

1. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;23:56–62
2. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979;134:382–389
3. Prien RF, Carpenter LL, Kupfer DJ. The definition and operational criteria for treatment outcome of major depressive disorder. Arch Gen Psychiatry 1991;48:796–800
4. Khan A, Khan SR, Shankles EB, et al. Relative sensitivity of the Montgomery-Asberg Depression Rating Scale, the Hamilton Depression Rating Scale and the Clinical Global Impressions rating scale in antidepressant clinical trials. Int Clin Psychopharmacol 2002;17:281–285
5. Zimmerman M, Coryell W, Pfohl B. The treatment validity of DSM-III melancholic subtyping. Psychiatry Res 1985;16:37–43
6. Frank E, Prien RF, Jarrett RB, et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder. Arch Gen Psychiatry 1991;48:851–855
7. Guelfi JD, Ansseau M, Timmerman L, et al. Mirtazapine versus venlafaxine in hospitalized severely depressed patients with melancholic features. J Clin Psychopharmacol 2001;21:425–431
8. Kyle CJ, Petersen HEH, Overo KF. Comparison of the tolerability and efficacy of citalopram and amitriptyline in elderly depressed patients treated in general practice. Depress Anxiety 1998;8:147–153
9. Nierenberg AA, Feighner JP, Rudolph R, et al. Venlafaxine for treatment-resistant unipolar depression. J Clin Psychopharmacol 1994;14:419–423
10. Schweitzer I, Burrows G, Tuckwell V, et al. Sustained response to open-label venlafaxine in drug-resistant major depression. J Clin Psychopharmacol 2001;21:185–189
11. Forlenza OV, Almeida OP, Stoppe A Jr, et al. Antidepressant efficacy and safety of low-dose sertraline and standard-dose imipramine for the treatment of depression in older adults: results from a double-blind, randomized, controlled clinical trial. Int Psychogeriatr 2001;13:75–84
12. Levine S, Deo R, Mahadevan K. A comparative trial of a new antidepressant, fluoxetine. Int Clin Psychpharmacol 1989;4:41–45
13. Mittmann N, Mitter S, Borden EK, et al. Montgomery-Asberg severity gradations [letter]. Am J Psychiatry 1997;154:1320–1321
14. Snaith RP, Harrop FM, Newby DA, et al. Grade scores of the Montgomery-Asberg Depression and the Clinical Anxiety Scales. Br J Psychiatry 1986;148:599–601
15. Hawley CJ, Gale TM, Sivakumaran T. Defining remission by cut off score on the MADRS: selecting the optimal value. J Affect Disord 2002;72:177–184
16. Posternak M, Zimmerman M. The validity of atypical depression. Arch Gen Psychiatry 2002;59:70–76
17. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV). Washington, DC: American Psychiatric Association; 1994
18. 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: Biometric Research, New York State Psychiatric Institute; 1995
19. Keller MB, Lavori PW, Friedman B, et al. The Longitudinal Interval Follow-Up Evaluation: a comprehensive method for assessing outcome in prospective longitudinal studies. Arch Gen Psychiatry 1987;44:540–548
20. Judd LL, Akiskal HS, Maser JD, et al. Major depressive disorder: a prospective study of residual subthreshold depressive symptoms as predictor of rapid relapse. J Affect Disord 1998;50:97–108
21. Hsiao JK, Bartko JJ, Potter WZ. Diagnosing diagnoses: receiver operating characteristic methods and psychiatry. Arch Gen Psychiatry 1989;46: 664–667
22. Riso LP, Thase ME, Howland RH, et al. A prospective test of criteria for response, remission, relapse, recovery, and recurrence in depressed patients treated with cognitive behavior therapy. J Affect Disord 1997;43:131–142
23. Paykel ES, Ramana R, Cooper Z, et al. Residual symptoms after partial remission: an important outcome in depression. Psychol Med 1995;25:1171–1180
24. 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;149:1046–1052
25. Van Londen L, Molenaar RPG, Goekoop JG, et al. Three- to 5-year prospective follow-up of outcome in major depression. Psychol Med 1998;28:731–735
26. Faravelli C, Ambonetti A, Pallanti S, et al. Depressive relapses and incomplete recovery from index episode. Am J Psychiatry 1986;143:888–891
27. Mintz J, Mintz LI, Arruda MJ, et al. Treatments of depression and the functional capacity to work. Arch Gen Psychiatry 1992;49:761–768
28. Zimmerman M, Chelminski I, Posternak M. A review of studies of the Montgomery-Asberg Depression Rating Scale in controls: implications for the definition of remission in treatment studies of depression. Int Clin Psychopharmacol 2004;19:1–8

