

It is illegal to post this copyrighted PDF on any website.

### CME Background

Articles are selected for credit designation based on an assessment of the educational needs of CME participants, with the purpose of providing readers with a curriculum of CME articles on a variety of topics throughout each volume. Activities are planned using a process that links identified needs with desired results.

To obtain credit, read the article, correctly answer the questions in the Posttest, and complete the Evaluation. A \$5 processing fee will apply.

### CME Objective

After studying this article, you should be able to:

- Use quantitative measures of depression in clinical practice with an understanding of how remission score cutoffs differ among scales

### Accreditation Statement

The CME Institute of Physicians Postgraduate Press, Inc., is accredited by the Accreditation Council for Continuing Medical Education to provide continuing medical education for physicians.



### Credit Designation

The CME Institute of Physicians Postgraduate Press, Inc., designates this journal-based CME activity for a maximum of 1 *AMA PRA Category 1 Credit*<sup>™</sup>. Physicians should claim only the credit commensurate with the extent of their participation in the activity.

*Note:* The American Academy of Physician Assistants (AAPA) accepts certificates of participation for educational activities certified for *AMA PRA Category 1 Credit*<sup>™</sup> from organizations accredited by ACCME or a recognized state medical society. Physician assistants may receive a maximum of 1 hour of Category I credit for completing this program.

### Date of Original Release/Review

This educational activity is eligible for *AMA PRA Category 1 Credit*<sup>™</sup> through February 28, 2019. The latest review of this material was December 2016.

### Financial Disclosure

All individuals in a position to influence the content of this activity were asked to complete a statement regarding all relevant personal financial relationships between themselves or their spouse/partner and any commercial interest. The CME Institute has resolved any conflicts of interest that were identified. In the past year, Alan J. Gelenberg, MD, Editor in Chief, has been a consultant for Zynx Health, has been a stock shareholder of Healthcare Technology Systems, and has been owner and editor of the *Biological Therapies in Psychiatry* newsletter. No member of the CME Institute staff reported any relevant personal financial relationships. **Faculty financial disclosure appears at the end of the article.**

## Identifying Remission From Depression on 3 Self-Report Scales

Mark Zimmerman, MD<sup>a,b,\*</sup>; Emily Walsh, BA<sup>a,b</sup>;  
Michael Friedman, MD<sup>a,b</sup>; Daniela A. Boerescu, MD<sup>a,b</sup>;  
and Naureen Attiullah, MD<sup>a,b</sup>

### ABSTRACT

**Objective:** Advocates of measurement-based care approaches toward treatment recommend the use of self-report questionnaires. Many self-report scales have been developed to measure the severity of depression. Because of the significance accorded remission by experts, it is important to compare different scales in their identification of remitted patients. 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 criteria for major depression in the identification of remission in patients treated in routine practice.

**Methods:** From June 2011 to November 2012, 153 depressed outpatients with *DSM-IV* major depressive disorder completed the Clinically Useful Depression Outcome Scale (CUDOS), Quick Inventory of Depressive Symptomatology–Self-Report (QIDS-SR), and Patient Health Questionnaire (PHQ-9). The patients were considered to be in remission according to the cutoff scores recommended by each scale's developers. The patients were also rated on the 17-item Hamilton Depression Rating Scale (HDRS).

**Results:** When the HDRS was used as the "gold standard" definition of remission, the CUDOS had the highest sensitivity for detecting remission (87%) and the QIDS-SR the highest specificity (97%). Overall, though, the level of agreement between the 3 self-report scales and HDRS in determining remission was approximately the same (79%–84%). The rate of remission was significantly higher on the HDRS compared to the QIDS-SR (35% vs 23%, McNemar  $P < .001$ ), significantly lower than the rate on the CUDOS when a cutoff score of 19 was used (35% vs 47%, McNemar  $P < .001$ ), and not significantly different from the rate on the PHQ-9 (31%) or the CUDOS when a cutoff score of 10 was used (34%).

**Conclusions:** There are significant differences between standardized scales in determining remission from depression. It is important for the developers of depression measures to empirically derive cutoff scores that define important constructs such as remission.

*J Clin Psychiatry* 2017;78(2):177–183

<https://doi.org/10.4088/JCP.16m10641>

© Copyright 2017 Physicians Postgraduate Press, Inc.

<sup>a</sup>Department of Psychiatry and Human Behavior, Brown Medical School, Providence, Rhode Island

<sup>b</sup>Department of Psychiatry, Rhode Island Hospital, Providence

\*Corresponding author: Mark Zimmerman, MD, 146 West River St, 11B, Providence, RI 02904 (mzimmerman@lifefspan.org).

Determining the impact of treatment is not simply a matter of *evaluating* outcome but rather a matter of *measuring* outcome. In mental health clinical settings, the effectiveness of treatment is typically based on unstructured interactions that yield unquantified judgments of progress, as clinicians rarely use scales in their practice.<sup>1,2</sup> This approach is at variance with other areas of medical care in which outcome is determined, in part, on the change of a numerical value. Body temperature, blood pressure, cholesterol values, blood sugar levels, cardiac ejection fraction, and white blood cell counts are examples

It is illegal to post this copyrighted PDF on any website.

**It is illegal to post this copyrighted PDF on any website.**

of quantifiable variables that are used to evaluate medical treatment progress. Standardized, quantifiable outcome measures exist for most major psychiatric disorders, yet they are infrequently used in routine clinical practice.<sup>1,2</sup>

The quantitative measurement of treatment outcome has long been the standard of research investigations of the efficacy and effectiveness of care. Recently, some investigators and treatment guidelines have suggested that scales should be used to monitor the course of treatment in routine clinical practice.<sup>3-6</sup> If the optimal delivery of mental health treatment depends, in part, on systematically assessing outcome, then reliable, valid, informative, and user-friendly measurement is critical to evaluating the quality and efficiency of care 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 Depression Rating Scale (HDRS)<sup>7</sup> is unlikely to meet with success. Self-report questionnaires are a cost-effective option because they are inexpensive in terms of professional time needed for administration, and they correlate highly with clinician ratings.<sup>8</sup> It is not surprising that self-administered and clinician-rated scales are highly correlated because clinician ratings are heavily reliant on patients' reports. An advantage of self-report scales is that they are free of clinician bias and are therefore immune from clinician overestimation of patient improvement, which might occur when there is an incentive to demonstrate favorable outcomes.

Experts in the treatment of depression have emphasized the importance of striving for remission.<sup>9-13</sup> This recommendation stems from studies that have consistently demonstrated that residual symptoms in patients who have responded to treatment have negative prognostic significance. That is, the presence of residual symptoms in treatment responders is associated with a much greater likelihood of recurrence of a full depressive syndrome.<sup>14,15</sup> In antidepressant efficacy trials, remission is usually defined by a score that is below a threshold value on an interview-based measure of depression severity such as the HDRS.<sup>16</sup>

Many self-report scales have been developed to measure the severity of depression.<sup>17</sup> Zimmerman et al<sup>18</sup> discussed the use of self-report scales in routine clinical practice and recommended measures that assess the *DSM-IV* criteria for major depressive disorder (MDD) and that are available for clinical use at no cost. Three such measures, the Clinically Useful Depression Outcome Scale (CUDOS),<sup>19</sup> Quick Inventory of Depressive Symptomatology–Self-Report (QIDS-SR),<sup>20</sup> and Patient Health Questionnaire (PHQ-9),<sup>21</sup> recommend cutoff scores to identify patients who are in remission from their depression. The National Quality Foundation recommends that depression remission rates be used as a performance measure.<sup>22,23</sup> Because of the significance accorded to remission by treatment guidelines and policy agencies, it is important to determine whether scales differ in their identification of remitted patients. If scales markedly differ in their definition of remission, then it would pose a problem in comparisons based on different scales. In addition, clinicians, who already are skeptical

- Many self-report scales have been developed to measure the severity of depression. Because of the significance accorded remission by experts, it is important to compare how different scales identify remitted patients.
- The results of the study comparing 4 symptom scales of depression found that there are significant differences between standardized scales in determining remission from depression.

about the incorporation of standardized measurement tools into their clinical practice, might object to the use of such tools if different scales yield markedly different results. Accordingly, in the present report from the Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS) project, we compared 3 self-report scales assessing the *DSM-IV* (and *DSM-5*) symptom criteria for MDD in the identification of remission in depressed patients treated in routine practice.

## METHODS

The Rhode Island MIDAS project represents an integration of research methodology into a community-based outpatient practice affiliated with an academic medical center.<sup>24</sup> Patients with a variety of psychiatric problems receive a comprehensive diagnostic evaluation at presentation for treatment. Not all patients are evaluated in this manner; however, there were no differences in demographic characteristics between patients who received a semistructured interview and those who received a routine clinical evaluation.<sup>25</sup> 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 Rhode Island Hospital institutional review committee approved the research protocol, and all patients provided informed, written consent.

Not all available patients participated in the study due to the lack of availability of raters. The sample therefore reflected a sample of convenience rather than a consecutive series of patients. Approximately half of the patients were diagnosed with MDD based on the Structured Clinical Interview for *DSM-IV*, Patient Edition (SCID-I/P),<sup>26</sup> whereas the other patients were diagnosed based on an unstructured clinical interview. The study was conducted from June 2011 to November 2012.

The patients completed the CUDOS, PHQ-9, and QIDS-SR at baseline and at follow-up and were evaluated with the 17-item HDRS by raters who were blinded to the scores on the self-report scales.

The QIDS-SR<sup>20</sup> uses 16 items to assess the *DSM-IV* symptom criteria. On the QIDS-SR, 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. The QIDS-SR score

**It is illegal to post this copyrighted PDF on any website.**

is derived by using the highest score 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 0–5 (no depression), 6–10 (mild depression), 11–15 (moderate depression), 16–20 (severe depression), and 21–27 (very severe depression).<sup>27</sup> Remission has been defined as a score that is within the “no depression” range.<sup>3</sup>

The PHQ-9<sup>21</sup> contains 9 items corresponding to the *DSM-IV* MDD criteria. It was designed and intended as a screening measure of MDD and has only secondarily been used as an outcome measure. Unlike the CUDOS and QIDS-SR, which assess symptoms over the past week, the time frame for the PHQ-9 is the past 2 weeks. Also unlike the CUDOS and the QIDS-SR, 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 ordinal scale indicating how often they have 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 0–4 (no depression), 5–9 (mild depression), 10–14 (moderate depression), 15–19 (moderately severe depression), and 20–27 (severe depression).<sup>21</sup> Remission has been defined as a score that is within the “no depression” range.<sup>28</sup>

The CUDOS<sup>19</sup> contains 18 items—16 symptom items assessing each of the *DSM-IV/DSM-5* inclusion criteria for MDD as well as 1 item assessing psychosocial impairment due to depression and 1 item assessing quality of life. The respondent is instructed to rate the symptom items on a 5-point ordinal 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). Unlike the PHQ-9, compound *DSM-IV* symptom criteria referring to more than 1 construct (eg, problems concentrating or making decisions, insomnia or hypersomnia, increased or decreased appetite) were subdivided into their respective components, and a CUDOS item corresponds to each component. Total scores range from 0 to 64. In the original study<sup>19</sup> of the scale’s validity, score ranges were empirically derived corresponding to depression severity categories: no depression, 0–10; minimal depression, 11–20; mild depression, 21–30; moderate depression, 31–45; and severe depression, 46 and above. In a separate analysis, the cutoff score for remission on the CUDOS (ie, <20) was empirically derived to maximize agreement with the HDRS definition of remission.<sup>29</sup> However, deriving a cutoff score for defining remission that corresponds to the cutoff on the HDRS allows for the presence of residual symptoms.<sup>30</sup> In contrast, the definition of remission in the PHQ-9 and QIDS-SR was equated with the nondepressed symptom severity range. To be comparable to the PHQ-9 and QIDS-SR

analyses, the cutoff score of 10 on the CUDOS, which corresponded to the nondepressed symptom severity range, was also examined.

The HDRS is the most commonly used clinician-rated outcome scale in depression treatment studies.<sup>31</sup> The original rating form included 21 items, although Hamilton<sup>7</sup> indicated that only the first 17 items should contribute to the total scale score because 1 of the last 4 items represented depressive type rather than depression severity (diurnal mood variation), and 3 other items did not occur with sufficient frequency (derealization, paranoia, and obsessional symptoms). Nine of the 17 items are rated from 0 to 4, whereas 8 items are rated 0 to 2; thus, the maximum score is 52. Remission on the HDRS was defined as a score of 7 or less.<sup>32</sup> While there has been some disagreement as to what cutoff score should be used to define remission on the HDRS, including research from our own group suggesting a lower cutoff score is more valid than a cutoff of 7, a cutoff score of 7 remains the most frequently used cutoff to define remission.<sup>30,33</sup> The HDRS was administered by highly trained research assistants with at least 2 years of experience administering psychiatric instruments.

### Statistical Analysis

We used the McNemar test to compare the percentage of patients classified as being in remission on the self-report measures. We examined the sensitivity and specificity of the self-report scales in identifying remission based on the HDRS. The  $\kappa$  statistic was used to determine the level of agreement between the scales in identifying remission.

## RESULTS

Patients diagnosed with *DSM-IV* MDD (N = 153) who presented for treatment to the Rhode Island Hospital Department of Psychiatry outpatient practice (n = 78) or who were in ongoing treatment and had their medication changed due to lack of efficacy (n = 75) were evaluated at baseline and at 4-month follow-up. The mean (SD) interval between the baseline and follow-up evaluations was 16.4 (4.2) weeks. The sample included 42 men (27.5%) and 111 women (72.5%) who ranged in age from 18 to 79 years (mean [SD] = 43.7 [13.6]).

On each scale, the patients showed significant mean  $\pm$  SD levels of improvement from baseline to follow-up (HDRS: 19.6  $\pm$  5.6 vs 11.8  $\pm$  8.3, paired  $t$  = 13.4,  $P$  < .001; CUDOS: 34.7  $\pm$  11.0.6 vs 20.4  $\pm$  14.0, paired  $t$  = 12.8,  $P$  < .001; PHQ-9: 17.0  $\pm$  5.6 vs 9.7  $\pm$  7.2, paired  $t$  = 11.8,  $P$  < .001; QIDS-SR: 15.8  $\pm$  4.4 vs 10.2  $\pm$  5.8, paired  $t$  = 12.5,  $P$  < .001). A large effect size was found for each scale, with little variability among the scales (HDRS, 1.1; CUDOS, 1.1; PHQ-9, 1.0; QIDS-SR, 1.0).

The data in Table 1 show the number of patients considered to be in remission at 4 months according to the different scales. Significantly more patients were classified as being in remission on the CUDOS when a cutoff score of 19 was used compared to the PHQ-9 (McNemar,  $P$  < .001) and QIDS-SR (McNemar,  $P$  < .001). Significantly more



**It is illegal to post this copyrighted PDF on any website.**

**Table 1. Percentage of Patients in Remission at 4-Month Follow-Up According to Different Measures of Depression**

Scale	Remitted		Nonremitted	
	n	% (95% CI)	n	% (95% CI)
CUDOS score ≤ 19	72	47.1 (39.2–55.0)	81	52.9 (45.0–60.8)
CUDOS score ≤ 10	52	34.0 (26.5–41.5)	101	66.0 (58.5–73.5)
PHQ-9 score ≤ 4	47	30.7 (23.4–38.0)	106	69.3 (62.0–76.6)
QIDS-SR score ≤ 5	35	22.9 (16.2–29.6)	118	77.1 (70.4–81.8)
17-item HDRS score ≤ 7	54	35.3 (27.7–42.9)	99	64.7 (57.1–72.3)

Abbreviations: CUDOS = Clinically Useful Depression Outcome Scale, HDRS = 17-item Hamilton Depression Rating Scale, PHQ-9 = Patient Health Questionnaire, QIDS-SR = Quick Inventory of Depressive Symptomatology–Self-Report.

**Table 2. Sensitivity and Specificity of Self-Report Scales in Identifying Remission at 4-Month Follow-Up According to the Hamilton Depression Rating Scale**

Scale	Sensitivity, % (95% CI)	Specificity, % (95% CI)
CUDOS	87.0 (78.0–96.0)	74.7 (76.1–83.3)
CUDOS-10	72.2 (60.2–84.2)	86.9 (80.3–93.5)
PHQ-9	70.4 (58.2–82.6)	90.9 (85.2–96.6)
QIDS-SR	59.3 (46.2–72.4)	97.0 (93.6–100.0)

Abbreviations: CUDOS = Clinically Useful Depression Outcome Scale, CUDOS-10 = CUDOS cutoff of 10 to define remission, PHQ-9 = Patient Health Questionnaire, QIDS-SR = Quick Inventory of Depressive Symptomatology–Self-Report.

patients were in remission according to the PHQ-9 than the QIDS-SR (McNemar,  $P = .031$ ). When a cutoff score of 10 was applied to the CUDOS, the rate of remission was still significantly higher than the rate on the QIDS-SR (McNemar,  $P < .001$ ) but was not significantly different from the rate on the PHQ-9 (McNemar,  $P = .30$ ). The rate of remission was significantly higher on the HDRS compared to the QIDS-SR (McNemar,  $P < .001$ ), significantly lower than the rate on the CUDOS based on a cutoff of 19 (McNemar,  $P < .001$ ), and not significantly different from the rate on the PHQ-9 and on the CUDOS when a cutoff score of 10 was used.

When the HDRS was used as the gold standard definition of remission, the CUDOS had the highest sensitivity for detecting remission and the QIDS-SR had the highest specificity (Table 2). Overall, though, the level of agreement between the 3 self-report scales and the HDRS in determining remission was approximately the same (range of agreement level, 79%–84%; range of  $\kappa$  coefficients, 0.57–0.63) (Table 3).

## DISCUSSION

Remission is the desired goal in the treatment of depression.<sup>4</sup> How to determine remission status is not settled. Different scales measure the severity of depression in different ways. Unlike the use of a thermometer to measure body temperature or a sphygmomanometer to measure blood pressure, in which all such instruments use the same metric, the scores on depression scales are not calibrated similarly, and the thresholds to determine remission on these scales have been derived in different ways. It is therefore not surprising that there was more than a 2-fold difference in the range of remission rates according to the scales used in the present study. At the most extreme, the remission rate

on the QIDS-SR was less than half the remission rate on the CUDOS (based on the cutoff score  $< 20$ ). On the other hand, the remission rates on the HDRS, PHQ-9, and CUDOS (based on the cutoff score of 10) were similar.

Standardized scales are typically not used in clinical practice.<sup>1,2</sup> In the past few years, there have been increasing calls for the utilization of such measures,<sup>4–6</sup> and self-report scales are more likely to be used than clinician-rated scales such as the HDRS. Efforts to compare health plans, institutions, clinical practices, or clinicians will be hampered when outcomes are measured with different instruments. In the present study, we found significant differences between 3 self-administered depression scales that presumably measure the same construct (ie, the symptom criteria of DSM-IV MDD) in the number of patients considered to be in remission. The scales differ somewhat in how they are scored, with the CUDOS and PHQ-9 assessing severity in terms of symptom frequency, whereas the QIDS-SR assesses severity in terms of both symptom frequency and symptom intensity. This difference might partly explain why the sensitivity to remission is the lowest with the QIDS-SR as compared to the others, despite the fact that the item content of the scales is largely the same.

A more likely factor that might account for the marked differences between the 3 scales of similar content in the determination of remission is the different methods employed for deriving the cutoff scores to define remission. The cutoffs on the QIDS-SR and the CUDOS were derived to correspond to the definition of remission on the 17-item HDRS.<sup>20,29</sup> The cutoff scores for severity ranges on the PHQ-9 were chosen for the pragmatic reason of making them easier for clinicians to recall.<sup>21</sup> The authors also noted that alternative cutoffs did not increase the association between increasing PHQ-9 severity and indices of construct validity. While the developers of the PHQ-9 did not discuss the use of the scale to define remission, other researchers have used the cutoff score on the PHQ-9 that identifies the nondepressed range as a measure for identifying the group that is in remission.<sup>34–36</sup> In contrast to the PHQ-9, the severity ranges on the CUDOS were derived from empirical study.<sup>19</sup> A large sample of psychiatric patients completed the scale and were rated on the Clinical Global Impressions–Severity of Illness scale (CGI-S).<sup>37</sup> The mean and standard deviation of CUDOS scores was 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. Thus, the cutoff score for the nondepressed range, which was used as an alternative definition of remission on the CUDOS in the current study, was empirically derived.

The cutoff score used to define remission on a depression severity measure will greatly influence how many patients meet the definition. To be sure, even on the HDRS, used for decades to evaluate outcome in treatment studies of depression, there is still debate and disagreement about which cutoff score should be used to define remission.<sup>16,30,33</sup>

The duration of the period of symptom resolution to define remission has also been a source of debate and

**Table 3. Concordance Among Depression Measures in Identifying Remission From Depression at 4-Month Follow-Up**

Scale	HDRS		CUDOS		PHQ-9		QIDS-SR	
	$\kappa$	% Agreement	$\kappa$	% Agreement	$\kappa$	% Agreement	$\kappa$	% Agreement
CUDOS	0.57	79.0	...	...	...	...	...	...
PHQ-9	0.63	83.6	0.64	82.3	...	...	...	...
QIDS-SR	0.61	84.3	0.45	73.2	0.57	83.0	...	...
CUDOS-10	0.60	81.7	0.73	86.9	0.78	90.2	0.60	83.6

Abbreviations: CUDOS=Clinically Useful Depression Outcome Scale, CUDOS-10=CUDOS cutoff of 10 to define remission, PHQ-9=Patient Health Questionnaire, QIDS-SR=Quick Inventory of Depressive Symptomatology–Self-Report.

confusion. In most treatment studies, remission is defined cross-sectionally, at a single point in time. In contrast, the *DSM-5* definition of remission requires the absence (or near complete absence) of symptoms to persist for at least 2 months. A recent analysis from the Collaborative Depression Study found that a 4-week period of symptom resolution was as valid a predictor of future course as an 8-week period.<sup>38</sup> It is important to realize that one of the purposes of defining remission is to identify a relatively homogeneous group of patients with regard to future morbidity.<sup>30</sup> In fact, this was the original conceptual basis for subdividing treatment responders into remitted and nonremitted patients. That is, it was established that treatment responders were a heterogeneous group, and subdividing them into remitters and nonremitters according to cutoff scores on symptom rating scales identified subgroups that differed in their risk of relapse.<sup>14,15</sup> However, measurement in routine clinical practice, without the benefit of a costly research infrastructure, is not as controlled as it is in research studies. In clinical practice, patients' appointments depend on the vagaries of clinicians' and patients' schedules. Appointments do not occur on the same proscribed time points for all patients; therefore, assessments are not completed at well-defined intervals as they are in research treatment studies. In routine clinical practice, it will be more difficult to define remission over a standardized, sustained interval. Thus, we anticipate that initial reports of remission status in routine clinical practice will be based on the type of cross-sectional assessments that are the focus of the present study and have been the standard in antidepressant treatment trials.

Not only is the cutoff score and time frame to define remission unsettled, but it is also unresolved whether remission from depression should be conceptualized more broadly than symptom reduction alone to include other constructs such as normalization of functioning,<sup>39,40</sup> and whether the distribution of symptom scale scores in healthy control samples should be considered in remission definitions.<sup>41,42</sup> Nonetheless, despite these unanswered questions in defining remission, as the measurement-based care approach toward treatment achieves greater acceptance in clinical practice, the tools to evaluate outcome are likely to be self-administered scales, and it is therefore important to appreciate that scale and cutoff score selection will impact one's results.

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. In the present study, remission was defined cross-sectionally at a single time point according to scores on symptom severity measures. Other constructs and time frames to define remission could be considered.<sup>43–45</sup> Information on longitudinal course following the assessment of remission status could help clarify the best thresholds to define remission status.<sup>38,46</sup> The study was limited to 4 scales—the HDRS and 3 self-report scales. Future studies of the comparability of measures in determining remission should also include the Clinical Global Index,<sup>47</sup> a simple, widely used global measure of severity that probably most closely corresponds to how clinicians determine whether patients are in remission, at least informally, in their practice. We did not establish the reliability of the HDRS ratings in the current study, though in prior studies in our clinical research laboratory, the reliability for administering the HDRS was high.<sup>48</sup>

Sometimes the self-report scales were completed first and sometimes the HDRS was done first. We did not systematically track this information and were therefore unable to examine the impact of an order effect. Finally, not all potential subjects were recruited into the study because of the lack of availability of raters. While no systematic selection bias impacted the composition of the sample, it is possible that the sample was unrepresentative of the larger sample. While this may have biased the findings with regard to the overall effect size of the instruments, it is less likely to have influenced the difference between instruments.

In conclusion, the disparity between standardized scales in determining whether depressed outpatients are in remission gives one pause. While we agree with recommendations to use quantitative measures of depression in clinical practice, we also caution against the use of these scales to compare outcome across clinical settings. It is important to empirically derive thresholds corresponding to severity levels of depression because of how these cutoff scores might be used to define important constructs such as remission.<sup>45,49–52</sup> Recommended cutoff scores on symptom severity scales to define remission, even when empirically derived, should not be reified but subject to repeated evaluation before they are adopted in a widespread manner.

**Submitted:** January 6, 2016; accepted September 27, 2016.

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

**Financial disclosure:** Dr Zimmerman has been on advisory boards for F. Hoffmann-La Roche, Genentech, and Lundbeck; has received research support from Azevan and Eli Lilly; has prepared educational material for Otsuka; and is the author of the Clinically Useful Depression Outcome Scale. Drs Friedman, Boerescu, and Attiullah and Ms Walsh have no personal affiliations or financial relationships with any commercial interest to disclose relative to the article.

**Funding/support:** None.

## REFERENCES

- Gilbody SM, House AO, Sheldon TA. Psychiatrists in the UK do not use outcomes measures: national survey. *Br J Psychiatry*. 2002;180:101–103.
- 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.
- Trivedi MH, Rush AJ, Wisniewski SR, et al; STAR\*D Study Team. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR\*D: implications for clinical practice. *Am J Psychiatry*. 2006;163(1):28–40.
- Gelenberg AJ, Freeman MP, Markowitz JC, et al; Work Group on Major Depressive Disorder. *Practice Guideline for the Treatment of Patients With Major Depressive Disorder*. 3rd ed. Arlington, VA: American Psychiatric Association; 2010.
- 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.
- Morris DW, Trivedi MH. Measurement-based care for unipolar depression. *Curr Psychiatry Rep*. 2011;13(6):446–458.
- Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960;23:56–62.
- Reilly TJ, MacGillivray SA, Reid IC, et al. Psychometric properties of the 16-item Quick Inventory of Depressive Symptomatology: a systematic review and meta-analysis. *J Psychiatr Res*. 2015;60:132–140.
- Keller MB. Past, present, and future directions for defining optimal treatment outcome in depression: remission and beyond. *JAMA*. 2003;289(23):3152–3160.
- Nierenberg AA, Wright EC. Evolution of remission as the new standard in the treatment of depression. *J Clin Psychiatry*. 1999;60(suppl 22):7–11.
- Rush AJ, Crismon ML, Toprac MG, et al. Consensus guidelines in the treatment of major depressive disorder. *J Clin Psychiatry*. 1998;59(suppl 20):73–84.
- Rush A, Trivedi M. Treating depression to remission. *Psychiatr Ann*. 1995;25:704–709.
- Stahl SM. Why settle for silver, when you can go for gold? response vs recovery as the goal of antidepressant therapy. *J Clin Psychiatry*. 1999;60(4):213–214.
- Paykel ES, Ramana R, Cooper Z, et al. Residual symptoms after partial remission: an important outcome in depression. *Psychol Med*. 1995;25(6):1171–1180.
- Thase ME, Simons AD, McGeary J, et al. Relapse after cognitive behavior therapy of depression: potential implications for longer courses of treatment. *Am J Psychiatry*. 1992;149(8):1046–1052.
- Rush AJ, Kraemer HC, Sackeim HA, et al; ACNP Task Force. Report by the ACNP Task Force on response and remission in major depressive disorder. *Neuropsychopharmacology*. 2006;31(9):1841–1853.
- 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.
- Zimmerman M, McGlinchey JB, Chelminski I. An Inadequate Community Standard of Care: Lack of Measurement of Outcome When Treating Depression in Clinical Practice. Primary Psychiatry Web site. June 1, 2008. <http://primarypsychiatry.com/an-inadequate-community-standard-of-care-lack-of-measurement-of-outcome-when-treating-depression-in-clinical-practice/>
- Zimmerman M, Chelminski I, McGlinchey JB, et al. A Clinically Useful Depression Outcome Scale. *Compr Psychiatry*. 2008;49(2):131–140.
- 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.
- 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.
- Patel MM, Brown JD, Croake S, et al. The current state of behavioral health quality measures: where are the gaps? *Psychiatr Serv*. 2015;66(8):865–871.
- Behavioral Health Endorsement Maintenance 2014 Final Report—Phase 3. National Quality Forum Web site. [http://www.qualityforum.org/Publications/2015/05/Behavioral\\_Health\\_Endorsement\\_Maintenance\\_2014\\_Final\\_Report\\_-\\_Phase\\_3.aspx](http://www.qualityforum.org/Publications/2015/05/Behavioral_Health_Endorsement_Maintenance_2014_Final_Report_-_Phase_3.aspx). May 2015.
- Zimmerman M. A review of 20 years of research on overdiagnosis and underdiagnosis in the Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS) Project. *Can J Psychiatry*. 2016;61(2):71–79.
- Zimmerman M, Mattia JJ. Psychiatric diagnosis in clinical practice: is comorbidity being missed? *Compr Psychiatry*. 1999;40(3):182–191.
- First MB, Spitzer RL, Gibbon M, et al. *Structured Clinical Interview for DSM-IV Axis I Disorders—Patient edition (SCID-I/P)*. New York, NY: Biometrics Research Department, New York State Psychiatric Institute; 1995.
- 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.
- Yeung AS, Jing Y, Brennenman SK, et al. Clinical Outcomes in Measurement-based Treatment (COMET): a trial of depression monitoring and feedback to primary care physicians. *Depress Anxiety*. 2012;29(10):865–873.
- Zimmerman M, Posternak MA, Chelminski I. Using a self-report depression scale to identify remission in depressed outpatients. *Am J Psychiatry*. 2004;161(10):1911–1913.
- 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.
- Zimmerman M, Clark HL, Multach MD, et al. Have treatment studies of depression become even less generalizable? a review of the inclusion and exclusion criteria used in placebo controlled antidepressant efficacy trials published during the past 20 years. *Mayo Clin Proc*. 2015;90(9):1180–1186.
- 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.
- 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.
- Katzelnick DJ, Duffy FF, Chung H, et al. Depression outcomes in psychiatric clinical practice: using a self-rated measure of depression severity. *Psychiatr Serv*. 2011;62(8):929–935.
- Menchetti M, Sighinolfi C, Di Michele V, et al. Effectiveness of collaborative care for depression in Italy: a randomized controlled trial. *Gen Hosp Psychiatry*. 2013;35(6):579–586.
- Watzke B, Heddaeus D, Steinmann M, et al. Effectiveness and cost-effectiveness of a guideline-based stepped care model for patients with depression: study protocol of a cluster-randomized controlled trial in routine care. *BMC Psychiatry*. 2014;14:230.
- Guy W. *ECDEU Assessment Manual for Psychopharmacology*. Revised Edition. Washington, DC: US Department of Health, Education, and Welfare; 1976.
- Judd LL, Schettler PJ, Rush AJ, et al. A new empirical definition of major depressive episode recovery and its positive impact on future course of illness. *J Clin Psychiatry*. 2016;77(8):1065–1073.
- 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(1):148–150.
- Zimmerman M, Galione JN, Attiullah N, et al. Depressed patients' perspectives of 2 measures of outcome: the Quick Inventory of Depressive Symptomatology (QIDS) and the Remission from Depression Questionnaire (RDQ). *Ann Clin Psychiatry*. 2011;23(3):208–212.
- Zimmerman M, Chelminski I, Posternak M. A review of studies of the Hamilton depression rating scale in healthy controls: implications for the definition of remission in treatment studies of depression. *J Nerv Ment Dis*. 2004;192(9):595–601.
- 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):1–7.
- Bech P, Carrozzino D, Austin SF, et al. Measuring euthymia within the Neuroticism Scale from the NEO Personality Inventory: a Mokken analysis of the Norwegian general population study for scalability. *J Affect Disord*. 2016;193:99–102.
- Nease DE Jr, Aikens JE, Klinkman MS, et al. Toward a more comprehensive assessment of depression remission: the Remission Evaluation and Mood Inventory Tool (REMIT). *Gen Hosp Psychiatry*. 2011;33(3):279–286.
- Zimmerman M, Martinez JH, Attiullah N, et al. The remission from depression questionnaire as an outcome measure in the treatment of depression. *Depress Anxiety*.

- 2014;31(6):533–538.
46. 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(2):131–142.
47. Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. *Psychiatry (Edgmont).* 2007;4(7):28–37.
48. Zimmerman M, Posternak MA, Chelminski I. Heterogeneity among depressed outpatients considered to be in remission. *Compr Psychiatry.* 2007;48(2):113–117.
49. Ballesteros J, Bobes J, Bulbena A, et al. Sensitivity to change, discriminative performance, and cutoff criteria to define remission for embedded short scales of the Hamilton Depression Rating Scale (HAM-D). *J Affect Disord.* 2007;102(1-3):93–99.
50. Hung CI, Liu CY, Wang SJ, et al. The cut-off points of the Depression and Somatic Symptoms Scale and the Hospital Anxiety and Depression Scale in detecting non-full remission and a current major depressive episode. *Int J Psychiatry Clin Pract.* 2012;16(1):33–40.
51. Østergaard SD, Rothschild AJ, Flint AJ, et al. Establishing the cut-off score for remission and severity-ranges on the Psychotic Depression Assessment Scale (PDAS). *J Affect Disord.* 2016;190:111–114.
52. Zimmerman M, Posternak MA, Chelminski I. Implications of using different cut-offs on symptom severity scales to define remission from depression. *Int Clin Psychopharmacol.* 2004;19(4):215–220.



## POSTTEST

To obtain credit, go to [PSYCHIATRIST.COM](http://PSYCHIATRIST.COM) (Keyword: February)

to take this Posttest and complete the Evaluation. A nominal processing fee is required.

1. In this comparison of 3 self-report depression scales, which finding about remission rates at 4-month follow-up is accurate?
  - a. Nearly equivalent rates of remission were found using the 3 scales
  - b. A two-fold difference was found between the lowest and highest remission rates
  - c. All rates of remission were above 50%
2. In this comparison of 3 self-report depression scales, which finding about effect sizes for improvement at 4-month follow-up is accurate?
  - a. Nearly equal effect sizes were found using the 3 scales
  - b. A two-fold difference was found between the lowest and highest effect sizes
  - c. All effect sizes were above 1
3. Mr P has been receiving treatment for depression for 6 months. His psychiatrist retires, and you receive a transfer summary noting that, at his last visit, Mr P scored 8 on a depression scale and is in remission. However, the note does not indicate which scale he completed. On which of the following scales could Mr P be considered to be in remission?
  - a. Clinically Useful Depression Outcome Scale (CUDOS)
  - b. Patient Health Questionnaire (PHQ-9)
  - c. Quick Inventory of Depressive Symptomatology–Self-Report (QIDS-SR)
  - d. Hamilton Depression Rating Scale (HDRS)