Published in final edited form as:

J Nerv Ment Dis. 2017 August ; 205(8): 656–664. doi:10.1097/NMD.0000000000000666.

Psychometric Properties of the Reconstructed Hamilton Depression and Anxiety Scales

Eliora Porter, M.A.^a, Dianne L. Chambless, Ph.D.^a, Kevin S. McCarthy, Ph.D.^b, Robert J. DeRubeis, Ph.D.^a, Brian A. Sharpless, Ph.D.^c, Marna S. Barrett, Ph.D.^d, Barbara Milrod, M.D.^e, Steven D. Hollon, Ph.D.^f, and Jacques P. Barber, Ph.D.^g

^aDepartment of Psychology, University of Pennsylvania, 425 S. University Avenue, Philadelphia, PA 19104

^bDepartment of Psychology, Chestnut Hill College, St. Joseph Hall, 4th Floor, North Wing, 9601 Germantown Avenue, Philadelphia, PA, 19118

^cAmerican School of Professional Psychology, Argosy University, 1550 Wilson Boulevard, Suite 763, Arlington, VA 22209

^dDepartment of Psychiatry, University of Pennsylvania, 3535 Market Street, Suite 670, Philadelphia, PA 19104

^eDepartment of Psychiatry, Weill Cornell Medical College, 525 East 68th Street, New York, NY 10065

^fDepartment of Psychology, Vanderbilt University, 2301 Vanderbilt Place, Nashville, TN, 37240

^gDerner Institute of Advanced Psychological Studies, Adelphi University, Hy Weinberg Center, 1 South Avenue, Garden City, NY 11530

Abstract

Although widely used, the Hamilton Rating Scale for Depression (HRSD) and Hamilton Anxiety Rating Scale (HARS) discriminate poorly between depression and anxiety. To address this problem, Riskind et al. (1987) created the Reconstructed Hamilton Scales by reconfiguring HRSD and HARS items into modified scales. To further analysis of the reconstructed scales, we examined their factor structure and criterion-related validity in a sample of patients with major depressive disorder and no comorbid anxiety disorders ($n=215$) or with panic disorder and no comorbid mood disorders ($n=149$). Factor analysis results were largely consistent with those of Riskind et al. The correlation between the new reconstructed scales was small. Compared to the original scales, the new reconstructed scales correlated more strongly with diagnosis in the expected direction. The findings recommend the use of the reconstructed HRSD over the original HRSD but highlight problems with the criterion-related validity of the original and reconstructed HARS.

Correspondence concerning this paper should be addressed to Eliora Porter, Department of Psychology, University of Pennsylvania, 425 S. University Avenue, Philadelphia, PA 19104-6081. Phone: 215-898-7376. eliora@sas.upenn.edu.

Conflicts of Interest: The authors have no conflicts of interest to declare.

Keywords

Hamilton Rating Scale for Depression; Hamilton Anxiety Rating Scale; anxiety; depression; psychometrics

The Hamilton Rating Scale for Depression (HRSD; Hamilton, 1960) and the Hamilton Anxiety Rating Scale (HARS; Hamilton, 1959) are widely used measures of depression and anxiety. The HRSD is considered to be the gold standard measure of depression and is frequently used as an outcome measure in depression treatment trials. Despite their wide usage, the HRSD and the HARS have been criticized on a number of grounds. Specifically, researchers have questioned the content validity of these scales: HRSD items capture many symptoms thought to be characteristic of anxiety but not depression, and HARS items capture symptoms thought to be characteristic of depression but not anxiety (Bagby et al., 2004; Maier et al., 1988). Further, researchers have noted that the HRSD and HARS fail to capture important aspects of depression and anxiety, respectively. For example, the widely used 17-item version of the HRSD does not assess worthlessness or concentration difficulties, even though these symptoms are included in the diagnostic criteria for major depressive disorder (MDD; Bagby et al., 2004), and the HARS has been criticized for not adequately measuring worry, a key feature of anxiety (Koerner et al., 2010). Empirical research has also raised questions about the criterion-related validity of the scales: One study found that the HARS failed to effectively distinguish between anxious and depressed individuals (Mountjoy & Roth, 1982).

In response to some of these criticisms, Riskind et al. (1987) reconfigured the existing HRSD and HARS items to improve the scales' discrimination between depression and anxiety. By and large, the Reconstructed Hamilton Scales (HRSD-R and HARS-R) were empirically derived based on the results of discriminant analyses and principal components analysis performed on a sample of 60 patients with *DSM-III* MDD (American Psychiatric Association [APA], 1980) and no comorbid anxiety disorder diagnoses and 60 patients with generalized anxiety disorder (GAD) and no comorbid depressive disorder diagnoses. The reconstructed scales proved equivalent or superior to the original scales in internal consistency ($\alpha = .73$ for both the original and revised HRSD and $\alpha = .83$ for the HARS-R versus $\alpha = .78$ for the HARS) and superior to the original scales in criterion-related validity, as measured by point-biserial correlations with diagnosis and discriminant function analyses. The latter results were replicated in a cross-validation sample of 38 patients with MDD and 33 patients with GAD. Further, the correlation between the reconstructed scales ($r = .15$) was much lower than the correlation between the original scales ($r = .62$), although the authors did not statistically test the difference in correlations.

A number of investigators have utilized the Reconstructed Hamilton Scales in their research: A search conducted using PsycInfo indicated that Riskind et al.'s (1987) paper had been cited 154 times. Unfortunately, efforts to replicate Riskind et al.'s findings regarding the superior psychometric properties of the Reconstructed Hamilton Scales have been few and have met with mixed success. Moras et al. (1992) examined the psychometric properties of the original and reconstructed scales in a sample of 412 outpatients with primary *DSM III-R*

anxiety disorder diagnoses (APA, 1987). Of these, 285 had a comorbid diagnosis of MDD. Although the correlation between the reconstructed scales was lower than that of the original scales, suggesting improved discriminant validity, the correlation between the HRSD-R and HARS-R ($r=.62$) was substantially larger than that found by Riskind et al. in their cross-validation sample. Further, the reconstructed scales were not superior to the original scales in distinguishing anxious individuals with MDD from anxious individuals without a comorbid MDD diagnosis. Similarly, Diefenbach et al. (2001) examined the properties of the original and reconstructed scales in 82 older adults who met *DSM-IV* criteria for GAD (APA, 1994), 36 of whom also met criteria for a comorbid depressive disorder diagnosis. Like Moras et al. (1992), Diefenbach and colleagues found that the correlation between the revised scales ($r=.58$) was lower than that of the original scales, but remained much larger than the correlation found by Riskind et al. (1987) in their cross-validation sample. Discriminant analysis indicated only a modest improvement for the reconstructed over the original scales in correct classification of the GAD-only vs. comorbid groups.

Several limitations of the present literature on the psychometric properties of the Reconstructed Hamilton Scales are readily apparent. First, apart from Riskind et al. (1987), no one has examined the extent to which the reconstructed scales differentiate between individuals with depression and no comorbid anxiety disorder diagnoses and those with an anxiety disorder and no comorbid mood disorder diagnoses. Rather, Moras et al. (1992) and Diefenbach et al. (2001) both tested the extent to which the HARS-R and HRSD-R distinguished between individuals with anxiety disorders with no comorbid depressive disorders versus those with anxiety disorders and a depressive disorder. This is an important issue to ultimately investigate, but it is first important to test the replicability of Riskind et al.'s basic finding that the reconstructed scales successfully distinguish between depressive disorder without anxiety disorder and anxiety disorder without depressive disorder. Although depression and anxiety are overlapping constructs (Watson et al., 1988) and often co-occur (Kessler et al., 2005), optimal measures of these constructs should capture the extent to which a given patient is depressed, independent of that patient's anxiety level, to the extent possible, and vice versa. Such measures will be most useful for researchers and clinicians alike. It is to be expected that differentiating between anxiety with and without depression may be quite a bit more challenging than distinguishing between cases without comorbidity given that the diagnostic picture may be less clear in comorbid populations. Second, the sample population was limited to older adults in one of these two studies. Previous research has suggested that depression may manifest differently in older adults (Blazer et al., 1986) and that it may be more difficult to distinguish anxiety from depression in this population (Parmelee et al., 1993). Finally, although Riskind et al.'s factor analyses were conducted on a relatively small sample of participants, no research to date has investigated their obtained factor structure of the Hamilton scales in an independent sample. This is particularly important because item scores on the Hamilton scales are ordinal, not continuous, and traditional methods of factor analysis may yield biased estimates when indicators are not continuous (Kline, 2013). More appropriate factor analytic methods have been developed for ordinal data since the publication of Riskind et al.'s paper.

Given continued widespread use of the Hamilton scales, the reconstructed scales offer the potential for improving the validity of research in depression and anxiety. Thus, further

examination of the psychometric properties of the reconstructed scales is important. In the present study, we test the factor structure and discriminant and criterion-related validity of the reconstructed scales in a sample of individuals with MDD and no comorbid anxiety disorder diagnoses and individuals with panic disorder (PD) and no comorbid mood disorder diagnoses. Participants included in the present sample were adults spanning a wide age range. A weighted least squares mean and variance (WLSMV) adjusted estimator, which is appropriate for ordinal level data (Wirth & Edwards, 2007), was utilized in factor analyses.

Methods

Participants

The present sample consisted of 215 patients with MDD and 149 patients with PD, as defined by *DSM-IV* criteria (APA, 1994). This study was conducted with the approval of the Institutional Review Board of the University of Pennsylvania.

MDD sample—MDD patients completed baseline measures for one of two trials comparing psychotherapy, pharmacotherapy, and placebo (Barber et al., 2012; DeRubeis et al., 2005). In addition to a current primary diagnosis of MDD, patients were required to have a 17-item HRSD score >13 in the Barber et al. trial ($n=70$) and >19 in the DeRubeis et al. trial ($n=145$). Both trials excluded individuals with serious medical conditions and those with a history of schizophrenia or psychosis, as well as those deemed to be at high risk for suicide. These studies were approved by the Institutional Review Boards of the University of Pennsylvania and of Vanderbilt University. MDD patients with a current comorbid anxiety disorder diagnosis were included in the parent studies but excluded from this psychometric study; 55% of patients in the Barber et al. trial and 40% of patients in the DeRubeis et al. trial were excluded from the present sample due to comorbidity.

PD sample—PD patients completed baseline measures for a study comparing the efficacy of three forms of psychotherapy for PD (Milrod et al., 2016). To meet criteria for the PD trial, patients were required to have a current primary diagnosis of PD with or without agoraphobia and to have had at least one panic attack per week in the month prior to study entry, unless they were avoiding most panic-inducing situations due to agoraphobia. Actively suicidal patients, as well as those with a history of psychosis or substance dependence within the past six months, were excluded from the trial. This study was approved by the Institutional Review Boards of the University of Pennsylvania and of Weill Cornell Medical College. PD patients with a current comorbid mood disorder diagnosis were included in the parent study but excluded from this psychometric study; 26% of patients in the PD trial were excluded from the present sample due to comorbid mood disorders. However, the present sample includes PD patients with comorbid *DSM-IV* anxiety disorder diagnoses (61.7% of the sample had 1 comorbid anxiety diagnosis).

Measures

Participants provided basic demographic information (see Table 1) and gave written informed consent for the parent studies. All measures relevant to the present study were administered at intake by master's and doctoral level psychologists.

Structured Clinical Interview for DSM-IV Axis I Disorders (SCID)—The SCID (First et al., 1995) was used to assess Axis I diagnoses in the MDD samples. Interrater reliability was excellent (Cicchetti & Sparrow, 1981) in both the Barber et al. (2012) and DeRubeis et al. (2005) samples (for MDD diagnosis, $r_1=.97$ in the Barber et al. trial and $\kappa=.80$ in the DeRubeis et al. trial).

Anxiety Disorders Interview Schedule for DSM-IV: Adult Version (ADIS)—The ADIS (Brown et al., 1994) was used to assess Axis I diagnoses in the PD sample. Interrater reliability was excellent ($\kappa=.79$ for PD, agoraphobia, and mood disorder diagnoses).

Hamilton Rating Scale for Depression (HRSD)—The HRSD was originally designed as 17-item, semi-structured interview (Hamilton, 1960); it included an additional four items that were not included in the total score. Subsequently, Guy (1976) developed a 24-item version of the HRSD, which includes all 21 items from the original HRSD rating form, as well as 3 additional items assessing helplessness, hopelessness, and worthlessness. The Guy version of the HRSD was employed by Riskind et al. (1987) in their development of the Reconstructed Hamilton Scales. Subsequently, Williams (1988) developed a structured interview guide for the 24-item version of the HRSD to standardize administration and improve reliability. In all three studies comprising the present data set, the HRSD was administered in accordance with Williams' (1988) guide with some minor revisions to item 13 (described below). It should be noted that the content of HRSD items 12 and 13 differs slightly between the Guy and Williams versions of the scale. Item 12 (called *somatic: gastrointestinal* on Guy's version of the scale and *appetite* on Williams' version of the scale) includes questions about feelings of heaviness in the abdomen and constipation on the Guy, but not the Williams, version of the scale. Additionally, Guy's item 13 (called *somatic: general* on his version of the scale and *energy* on the version of the HRSD used in the current study) includes questions about head-, back-, and muscle aches that were removed from the version of the HRSD employed in the present study. Apart from these changes, the two versions of the HRSD assess the same content.

Each HRSD item is rated on a 3-point or 5-point ordinal scale. Both 17-item and 24-item HRSD totals were calculated for the present study. Possible scores range from 0 to 52 on the 17-item HRSD and from 0 to 76 on the 24-item HRSD. Interrater reliability was excellent in all three study samples, $r_1=.80-.96$.

Hamilton Anxiety Rating Scale (HARS)—The HARS is a 14-item measure originally designed to be administered as a semi-structured interview (Hamilton, 1959). The DeRubeis et al. (2005) MDD study utilized the original, semi-structured HARS, whereas the PD study and the Barber et al. (2012) MDD study utilized a structured interview version of the HARS, which has been shown to have superior interrater reliability compared to the original measure (Bruss et al., 1994). On both versions of the HARS, each symptom is rated on a 0 to 4 ordinal scale, with total possible HARS scores ranging from 0 to 56. Interrater reliability on the HARS was not calculated for either MDD study, as this was not a primary outcome measure. For the PD study, interrater reliability was excellent ($r_1=.96$).

Analytical Approach

We first used exploratory factor analysis (EFA) to examine the factor structure of the Hamilton scales in our sample. We employed this approach rather than confirmatory factor analysis (CFA) because of slight differences in item content between the version of the Hamilton scales we utilized and the version used by Riskind et al. (1987), and because CFA may yield poor model fit when item-level data are used as indicators (Floyd & Widaman, 1995). Because parallel analysis does not work well for categorical data, we examined scree plots to determine the number of factors to retain. Based on the results of the EFA, we made slight modifications to the Reconstructed Hamilton Scales. Because we lacked the sample size to replicate our factor structure in an independent sample, we adopted a conservative approach to modifying the reconstructed scales: We sought to remain as consistent with Riskind et al. as possible without discounting our empirical findings. Next, we examined the internal consistency of our modified scales (entitled the HRSD-R-II and HARS-R-II), as well as the correlation between the two scales and the point-biserial correlations between the respective scales and diagnosis (MDD or PD). We used Cohen's (1988) guidelines to interpret the magnitude of these correlations. We also compared the size of the correlations between the two scales and between each scale and diagnosis for the original HRSD and HARS and for the HRSD-R-II and HARS-R-II to determine whether our reconstructed scales demonstrated superior discriminant and criterion-related validity compared to the original scales. Finally, we examined the receiver operating characteristic (ROC) curves for the original and revised scales and compared the areas under the curve (AUCs) for the original and revised depression and anxiety scales, respectively, as a further test of whether our reconstructed scales demonstrated superior discriminant and criterion-related validity compared to the original scales.

Results and Discussion

See Table 1 for demographic variables and median HRSD and HARS scores.

Factor Analysis

We first conducted an exploratory item-level factor analysis with all HARS and 24-item HRSD items as indicators to determine whether the two-factor solution generated would produce depression and anxiety factors roughly corresponding to those found by Riskind et al. (1987). This analysis was conducted in MPlus using the categorical option and a WLSMV adjusted estimator. WLSMV makes use of the polychoric correlation matrix and assumes that categorical ordered responses reflect an underlying, continuous, multivariate normal response distribution (Wirth & Edwards, 2007). This analysis was conducted using a Geomin (oblique) rotation and allowing for one- to five-factor solutions, to explore other factor solutions that might provide a better fit for the data. Eigenvalues for the five factors were as follows: 8.22, 4.34, 2.32, 1.95, and 1.49. The two-factor solution closely matched the factor structure found by Riskind et al., producing factors corresponding to depression and anxiety, which were correlated at $r=.05$. An examination of the scree plot indicated that a three-factor solution provided the best fit to the data. This factor solution produced an anxiety factor identical to that found in the two-factor solution, but further subdivided depression items into two factors representing vegetative and cognitive aspects of

depression, respectively, with most depression items loading strongly on both depression factors but few anxiety items with secondary loadings on the depression factors and vice versa. There was a medium-sized and significant positive correlation between the two depression factors ($r=.32$, $p<.05$), and a small negative correlation between the anxiety factor and both depression factors ($r=-.20$ with both depression factors, both $p<.05$). Combining the items from the two depression factors in the three-factor solution yielded a larger depression component virtually identical to that generated in the two-factor solution. Given that our goal was to validate reconstructed versions of the Hamilton scales that best distinguished between depression and anxiety, rather than to examine subfactors of depression, we chose to combine the two depression subfactors into a larger depression component for the purpose of assigning items to reconstructed depression or anxiety scales. Factor loadings for the three-factor solution, as well as item medians and ranges, are shown in Table 2.

Although the three-factor solution generated with the present data was similar to the factor solution of Riskind et al. (1987) when the two depression subfactors were combined, several key differences emerged. Some of these appeared to reflect genuine differences in the response characteristics of our sample compared to those of Riskind et al.'s sample, whereas others appeared to be better attributed to the slight variations in item content between the version of the HRSD used in our study and the version used in Riskind et al.'s study. We detected three main differences in factor loadings for items whose content was unchanged between the different versions of the HRSD. First, Riskind et al. found that HARS item 12 (*genitourinary*) loaded comparably on depression and anxiety, and for this and other reasons, they excluded this item from the reconstructed scales. In contrast, we found that HARS item 12 loaded more strongly on the vegetative depression subfactor. Second, Riskind et al. found that HRSD item 9 (*agitation*) loaded more strongly on the anxiety factor. In our sample this item failed to load strongly on any factor, although the loading was somewhat higher on the cognitive depression subfactor than the anxiety factor. Agitation was rarely observed in our sample, restricting the range on this variable for psychometric analysis. For this reason, we opted to stay consistent with Riskind et al. and kept agitation on the anxiety factor. Third, Riskind et al. found the HRSA item 14 (*behavior at interview*) loaded more strongly on the anxiety factor, whereas we found that this item had comparable loadings on the anxiety factor and cognitive depression subfactor. To be conservative, we chose to remain consistent with Riskind et al. and to keep this item on the anxiety scale. Finally, Riskind et al. included HRSD items 4 (*insomnia early*) and 16 (*weight loss*) on the HRSD-R, despite the fact that they found that the former loaded more strongly on anxiety, and the latter did not load strongly on either factor; in contrast, we found both items to load strongly on the depression factors, and not the anxiety factor.

Two main differences between our factor solution and the factor structure found by Riskind et al. (1987) appeared to be due directly to differences in HRSD item content. First, Riskind et al. found that HRSD item 12 (called *somatic: gastrointestinal* on their version of the scale) loaded on the anxiety factor; in contrast, we found that this item, called *appetite* on our version of the scale, loaded on the vegetative depression subfactor. Second, Riskind et al. found that HRSD item 13 (*somatic: general* on Guy's HRSD) loaded on the anxiety factor; in contrast, we found this item (*energy* on our version of the HRSD) to load on the

depression factors. Riskind et al. chose to exclude this item from the reconstructed scales, despite its high loading on anxiety, because “it confounds headaches and backaches (aspects of anxiety) with heaviness in the back and limbs (usually an aspect of depression)” (p. 477). As noted, item content differed slightly for items 12 and 13 between our version of the HRSD and the version used by Riskind et al. such that questions about head-, back-, and muscle aches were removed from item 13 in the version of the HRSD used in the present study.

To summarize, when the two depression subfactors were combined, the three-factor solution generated by our EFA closely matched the factor structure reported by Riskind et al. (1987), with a few minor differences which appeared to be by and large the result of slight differences in item wording between our version of the HRSD and the version employed by Riskind et al. The similarity is remarkable given differences in the two study populations and in the factor analytic methods employed.

Final revised scales (HRSD-R-II and HARS-R-II)—Because we were unable to cross-validate the results of our exploratory factor analysis in an independent sample, we were wary of capitalizing on random variance that may have affected our results, and therefore chose to largely retain Riskind et al.’s (1987) reconstructed scales in the remaining analyses, altering the configuration of the Reconstructed Hamilton Scales only when there was a very strong empirical or theoretical rationale for doing so. Given that HRSD items 12 and 13 differed both in item content and factor loading in our analyses compared to Riskind et al.’s analyses, we chose to remove HRSD item 12 from the reconstructed HARS and to include both HRSD items 12 and 13 on the reconstructed HRSD. Additionally, we chose to include HRSD item 14 (*genital* on Guy’s [1976] scale, *libido* on our version of the scale) on the reconstructed HRSD. Both we and Riskind et al. found this item to load on the depression factor, but Riskind et al. excluded this item from the reconstructed scales without providing a theoretical or empirical rationale for this decision. Scale item assignments are displayed in Table 2. We refer to our scales as the HRSD-R-II and HARS-R-II to distinguish them from Riskind et al.’s HRSD-R and HARS-R.

Psychometric Evaluation of the HRSD-R-II and HARS-R-II

We then evaluated the psychometric properties of the HRSD-R-II and HARS-R-II and compared these estimates to estimates obtained using the original HRSD 17- and 24-item versions and the original HARS.

Internal consistency—We assessed the internal consistency (Cronbach’s α) of the HRSD-R-II and HARS-R-II and compared estimates for the reconstructed scales and the original scales. For the HRSD-R-II, $\alpha=.85$ compared to $\alpha=.77$ for the 17-item HRSD and $\alpha=.81$ for the 24-item version. For the HARS-R-II, $\alpha=.75$ compared to $\alpha=.68$ for the original HARS. Thus, internal consistency of the reconstructed scales was acceptable, consistent with previous research (Diefenbach et al., 2001; Moras et al., 1992; Riskind et al., 1987).

Correlation between the depression and anxiety scales—To test whether the HRSD-R-II and HARS-R-II distinguished the constructs of depression and anxiety, we correlated the two scores and compared these correlations to the correlations between the original versions of the scales. The correlation between HRSD-R-II and HARS-R-II scores was small and non-significant ($r=.03$, $p=.60$), whereas there was a large correlation between scores on the original scales ($r=.49$ and $r=.50$ for HARS with 17-item HRSD and 24-item HRSD, respectively, both $p<.001$). Implementation of the Raghunathan et al. (1996) procedure for comparing non-overlapping correlated correlation coefficients indicated that the correlation between the revised scales was significantly smaller than the correlation between the HARS and both the 17-item HRSD ($ZPF=13.03$, $p<.001$) and the 24-item HRSD ($ZPF=15.10$, $p<.001$). This finding is consistent with results reported by other researchers (Diefenbach et al., 2001; Moras et al., 1992; Riskind et al., 1987). However, the very small magnitude of the correlation between the revised scales is consistent with Riskind et al., but diverges from Diefenbach et al. and Moras et al.'s findings. Perhaps the reason for this discrepancy relates to differences in study populations: Both we and Riskind et al. examined a sample of individuals with depression without comorbid anxiety and individuals with anxiety without comorbid depression, whereas other authors examined the properties of the reconstructed scales in comorbid samples (Diefenbach et al., 2001; Moras et al., 1992). We would expect the distinction between depression and anxiety to be most evident when examining these constructs in a sample without overlap in conditions, and indeed this is what the results show.

Nonetheless, both we and Riskind et al. (1987) found the correlation between scales to be lower than expected given the well-documented common features of depression and anxiety (e.g., negative affectivity; Watson et al., 1988). The correlation between the reconstructed scales may be artificially low in our sample because factor analyses conducted on this sample guided the development of the HRSD-R-II and HARS-R-II, but this explanation cannot account for the low correlation between the reconstructed scales found in Riskind et al.'s cross-validation sample. More likely, the low correlation between scales may reflect problems with content validity. The Hamilton scales lack items that specifically address perseverative thought processes (e.g., rumination, worry), which are important shared features of depression and anxiety. Inclusion of such items on the depression and anxiety scales, respectively, may lead to more complete coverage of the constructs of depression and anxiety, as well as an increase in the correlation between scales.

Relationship between diagnosis and depression and anxiety scale scores—

We next examined whether HRSD-R-II and HARS-R-II scores distinguished between patients with a PD vs. an MDD diagnosis (coded as 0=PD and 1=MDD) using point biserial correlations, and compared these correlations for the newly revised scales to those found for the original scales. For the depression scales, the correlation should be strongly positive if scale scores distinguish patients diagnosed with PD vs. MDD. The correlation between the HRSD-R-II and a diagnosis of depression was very large and significant ($r_{pb}=.83$, $p<.001$), as were the correlations between the 17-item and 24-item HRSD and MDD diagnosis ($r_{pb}=.80$ and $r_{pb}=.77$, respectively, both $p<.001$). Implementation of the Meng et al. (1992) procedure for comparing correlated correlation coefficients indicated that the coefficient for

the HRSD-R-II was significantly larger than that of both the 17-item HRSD ($z=2.07$, $p=.04$) and the 24-item HRSD ($z=6.79$, $p<.001$).

For the anxiety scales, the correlation should be strongly negative if scale scores distinguish patients diagnosed with PD vs. MDD. There was a small negative effect size ($r_{pb}=-.19$, $p<.001$) for the HARS-R-II, suggesting that HARS-R-II scores were significantly lower in the MDD than PD group, albeit to a modest degree. On the original HARS, contrary to the intent of the scale, patients with MDD scored higher than those with PD, although the effect size was small ($r_{pb}=.19$, $p<.001$). The difference between the correlations for revised and original scales was statistically significant ($z=-12.47$, $p<.001$).

To further examine the relationship between HRSD-R-II and HARS-R-II scores and diagnosis, we used ROC analysis to compare the areas under the curve for the original and revised scales, which constituted a test of the relative sensitivity and specificity of the original and revised scales. Analyses were conducted in the “pROC” package in R (Robin et al., 2011) using the DeLong et al. (1988) methodology for comparing the areas under correlated ROC curves. For the HRSD-R-II, $AUC=.98$, $SE=.008$, $p<.001$, as compared to $AUC=.97$, $SE=.01$, $p<.001$ for the 17-item HRSD and $AUC=.95$, $SE=.01$, $p<.001$ for the 24-item HRSD. The AUC for the HRSD-R-II was significantly greater than that of the 24-item HRSD ($z=3.74$, $p<.001$) but not that of the 17-item HRSD ($z=1.37$, $p=.17$). For the HARS-R-II, $AUC=.61$, $SE=.03$, $p=.001$, as compared to $AUC=.39$, $SE=.03$, $p<.001$ for the HARS. Thus, the HARS-R-II performed significantly better than chance in predicting a diagnosis of PD, whereas the original HARS performed significantly worse than chance (an AUC of .50 indicates chance agreement). The AUC for the HARS-R-II was significantly greater than that of the HARS ($z=-13.20$, $p<.001$). See Tables 3 and 4 for sensitivity, specificity, positive predictive power, and negative predictive power at each value on the HRSD-R-II and the HARS-R-II.

Thus, mirroring Riskind et al.'s (1987) results we found the HRSD-R-II and HARS-R-II to correlate more strongly with diagnosis in the expected direction relative to the original HRSD and HARS, thus supporting the superior criterion-related validity of the reconstructed scales. However, our results indicated that while scores on the HRSD-R-II were closely related to diagnosis, scores on the HARS-R-II showed only a small relationship with diagnosis. Similarly, we found that the HRSD-R-II and HARS-R-II outperformed the 24-item HRSD and HARS in terms of sensitivity and specificity as measured by comparing the areas under the curve, but the HRSD-R-II demonstrated nearly optimal sensitivity and specificity, whereas the HARS-R-II was only slightly better than chance. The superior performance of the depression scales relative to the anxiety scales as indexed by their correlations with diagnosis and the areas under the curve may be artifactual: MDD patients in the present sample were selected on the basis of receiving a certain minimum score on the HRSD, whereas PD patients were not required to achieve any particular cutoff on the HARS. Indeed, when we restricted the PD sample to those who scored >14 on the HARS (the empirical cutoff score for moderate generalized anxiety disorder; Matza et al., 2010), we found higher correlations in the expected direction between anxiety scale scores and diagnosis ($r_{pb}=-.29$ for diagnosis and HARS, $p<.001$; $r_{pb}=-.58$ for diagnosis and HARS-R-II, $p<.001$) and greater AUCs for the anxiety scales (for HARS, $AUC=.72$, $SE=.03$, $p<.001$;

for HARS-R-II, $AUC=.90$, $SE=.02$, $p<.001$) than those found in the full sample. However, in this subsample the correlations between the depression scales and diagnosis remained substantially larger than those between the anxiety scales and diagnosis and the AUCs for the depression scales remained somewhat larger than those of the anxiety scales, suggesting that the use of cutoff scores for the MDD sample only cannot fully explain the superior performance of the depression scales in our sample. Details of these analyses are available on request.

Alternatively, these findings might reflect better content validity for the HRSD as compared to the HARS: The HARS has been criticized for its lack of items capturing worry and other core cognitive features of anxiety (Koerner et al., 2010). In contrast, the HRSD captures the core depressive symptoms of depressed mood and anhedonia, and it has been modified to include core cognitive features of depression such as helplessness, hopelessness, and worthlessness. Thus, the 24-item HRSD seems to more fully capture the construct of depression than the HARS captures the construct of anxiety.

Conclusions

We examined the psychometric properties of the Reconstructed Hamilton Scales in a substantial clinical sample of individuals with MDD and no comorbid anxiety disorder diagnoses and individuals with PD and no comorbid mood disorder diagnoses. To our knowledge, no published study apart from Riskind et al. (1987) has undertaken this endeavor in a sample of individuals with depression without comorbid anxiety and anxiety without comorbid depression. Furthermore, unlike Riskind et al., in our study we employed factor analytic methods appropriate to the analysis of ordinal data. Overall, our results largely mirror those of Riskind et al. and favor the use of the HRSD-R-II over the original HRSD in future clinical and research endeavors. Our results also highlight problems with the criterion-related validity of the anxiety scales. Although the HARS-R-II outperformed in original HARS in terms of its correlation with diagnosis in the expected direction, this correlation remained quite modest, suggesting that even in its reconfigured form, the HARS-R-II remains far from optimal as a measure of anxiety symptoms.

The present study has limitations. First, interrater reliability on the HARS was not calculated for the trials from which our MDD sample was drawn. Thus, we cannot be certain raters were being consistent in their scoring of the HARS for MDD patients. Second, given our sample size, we were unable to confirm whether the factor structure generated by our exploratory factor analysis could be replicated in an independent sample. Such replication is highly desirable because we made several minor changes to Riskind et al.'s (1987) reconstructed scales. Future research demonstrating that this factor structure can be replicated in an independent sample would increase our confidence in these results. Finally, our anxious sample included participants with a primary diagnosis of only one type of anxiety disorder, panic disorder. Given the somatic nature of panic disorder and the fact that the HARS tends to emphasize the somatic, rather than the cognitive, aspects of anxiety, it is possible that the HARS items may have greater content validity for panic disorder than for other anxiety disorder diagnoses. Thus, one concern is that the HARS might perform even worse in a sample with different anxiety disorder diagnoses. However, our concern is

lessened for two reasons: First, Riskind et al. found results similar to ours in a sample of patients with GAD, and second, the majority of PD patients in our sample (61.7%) met criteria for at least one additional anxiety disorder diagnosis.

In summary, together with Riskind et al.'s (1987) findings, the present study provides support for the superior psychometric properties of the Reconstructed Hamilton Scales relative to the original scales. Although there is a long tradition of using the original HRSD and HARS to capture depression and anxiety, respectively, the results of the present study suggest that this decision warrants reconsideration. Specifically, researchers and clinicians might consider using the HRSD-R-II in place of the original HRSD to obtain a purer measure of depression. In contrast, although the HARS-R-II constitutes an improvement over the original HARS, our results do not recommend the use of this or the original scale given the low correlation between both versions of the reconstructed HARS and PD diagnosis. The HARS has also been criticized for not including items capturing key aspects of anxiety such as worry (Koerner et al., 2010), and this criticism also applies to the HARS-R-II, because we did not add additional items to the scale. In contrast, we are less concerned about the content validity of the HRSD-R-II, because the 24-item HRSD (from which most HRSD-R-II items were drawn) appears to more fully capture the core symptoms of depression. It should be noted that although the HRSD-R-II appears to be quite successful in distinguishing between cases of depression without comorbid anxiety and cases of anxiety without comorbid depression, it may be considerably more difficult to make the distinction between depression and anxiety in comorbid samples, as was the case for Moras et al. (1992) and Diefenbach et al. (2001). Nevertheless, if the Hamilton scales are to be used, our findings argue for selection of the HRSD-R-II and HARS-R-II over the original scales.

Acknowledgments

The authors thank John Riskind for providing additional details regarding the methods and results of his original paper on the Reconstructed Hamilton Scales.

Source of Funding: This research was supported in part by the National Institute of Mental Health [grant numbers R01-MH070664, R01-MH070918, R01-MH061410, R10-MH055877, R10-MH055875, and K02-MH001697]. The sponsor played no role in the design or conduct of this research.

References

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 3rd. Washington, DC: Author; 1980.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 3rd ed., text rev. Washington, DC: Author; 1987.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th. Washington, DC: Author; 1994.
- Bagby RM, Ryder AG, Schuller DR, Marshall MB. The Hamilton Depression Rating Scale: Has the gold standard become a lead weight? *American Journal of Psychiatry*. 2004; 161:2163–2177. [PubMed: 15569884]
- Barber JP, Barrett MS, Gallop R, Rynn MA, Rickles K. Short-term dynamic psychotherapy versus pharmacotherapy for major depressive disorder: A randomized, placebo-controlled trial. *Journal of Clinical Psychiatry*. 2012; 73:66–73. [PubMed: 22152401]

- Blazer, D., George, L., Landerman, R. The phenomenology of late-life depression. In: Bebbington, PE., Jacoby, R., editors. *Psychiatric disorders in the elderly*. London: Mental Health Foundation; 1986. p. 143-151.
- Brown, TA., Di Nardo, PA., Barlow, DH. *Anxiety Disorders Interview Schedule for DSM-IV (ADIS-IV)*. San Antonio, TX: Psychological Corporation/Graywind Publications Incorporated; 1994.
- Bruss GS, Gruenberg AM, Goldstein RD, Barber JP. Hamilton Anxiety Rating Scale Interview guide: Joint interview and test-retest methods for interrater reliability. *Psychiatry Research*. 1994; 53:191–202. [PubMed: 7824679]
- Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*. 1981; 86:127–137. [PubMed: 7315877]
- Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd. Hillsdale, NJ: Lawrence Erlbaum; 1988.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometric*. 1988; 44:837–845.
- DeRubeis RJ, Hollon SD, Amsterdam JD, Shelton RC, Young PR, Salomon RM, Gallop R. Cognitive therapy vs medications in the treatment of moderate to severe depression. *Archives of General Psychiatry*. 2005; 62:409–416. [PubMed: 15809408]
- Diefenbach GJ, Stanley MA, Beck JG, Novy DM, Averill PM, Swann AC. Examination of the Hamilton scales in assessment of anxious older adults: A replication and extension. *Journal of Psychopathology and Behavioral Assessment*. 2001; 23:117–124.
- First, M., Spitzer, R., Gibbon, M., Williams, J. *Structured Clinical Interview for Axis I DSM-IV Disorders - Patient Edition (SCID-I/P)*. New York: Biometrics Research Department, NY State Psychiatric Institute; 1995.
- Floyd FJ, Widaman KF. Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*. 1995; 7:286–299.
- Guy, W. *NCDEU assessment manual for psychopharmacology* (DHHS Publication No. ADM 91-338). Washington, DC: U.S. Department of Health, Education, and Welfare; 1976.
- Hamilton M. The assessment of anxiety states by rating. *British Journal of Medical Psychology*. 1959; 32:50–55. [PubMed: 13638508]
- Hamilton M. A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*. 1960; 23:56–62.
- Kessler RC, Chiu WT, Demler O, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*. 2005; 62:617–627. [PubMed: 15939839]
- Kline, RB. Exploratory and confirmatory factor analysis. In: Petscher, Y., Schatsschneider, C., editors. *Applied quantitative analysis in the social sciences*. New York: Routledge; 2013. p. 171-207.
- Koerner N, Antony MM, Dugas MJ. Limitations of the Hamilton Anxiety Rating Scale as a primary outcome measure in randomized, controlled trials of treatments for generalized anxiety disorder. *American Journal of Psychiatry*. 2010; 167:103–104.
- Maier W, Buller R, Philipp M, Heuser I. The Hamilton Anxiety Scale: Reliability, validity, and sensitivity to change in anxiety and depressive disorders. *Journal of Affective Disorders*. 1988; 14:61–68. [PubMed: 2963053]
- Matza LS, Morlock R, Sexton C, Malley K, Feltner D. Identifying HAM-A cutoffs for mild, moderate, and severe generalized anxiety disorder. *International Journal of Methods in Psychiatric Research*. 2010; 19(4):223–232. [PubMed: 20718076]
- Meng X, Rosenthal R, Rubin DB. Comparing correlated correlation coefficients. *Psychological Bulletin*. 1992; 111:172–175.
- Milrod BL, Chambless DL, Gallop R, Busch FN, Schwalberg M, McCarthy KS, Barber JP. Psychotherapies for panic disorder: A tale of two sites. *Journal of Clinical Psychiatry*. 2016; 77:927–935. [PubMed: 27464313]
- Moras K, Di Nardo PA, Barlow DH. Distinguishing anxiety and depression: Reexamination of the Reconstructed Hamilton Scales. *Psychological Assessment*. 1992; 4:224–227.

- Mountjoy CQ, Roth M. Studies in the relationship between depressive disorders and anxiety states: II. Clinical items. *Journal of Affective Disorders*. 1982; 4:149–161. [PubMed: 6213692]
- Parmelee PA, Katz IR, Lawton MP. Anxiety and its association with depression among institutionalized elderly. *American Journal of Geriatric Psychiatry*. 1993; 1:46–58. [PubMed: 28530946]
- Raghunathan TE, Rosenthal R, Rubin DB. Comparing correlated but nonoverlapping correlations. *Psychological Methods*. 1996; 1:178–183.
- Riskind JH, Beck AT, Brown G, Steer RA. Taking the measure of anxiety and depression: Validity of the Reconstructed Hamilton Scales. *Journal of Nervous and Mental Disease*. 1987; 175:474–479. [PubMed: 3625186]
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J, Muller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011; 12
- Watson D, Clark LA, Carey G. Positive and negative affectivity and their relation to anxiety and depressive disorders. *Journal of Abnormal Psychology*. 1988; 97:346–353. [PubMed: 3192830]
- Williams JB. A structured interview guide for the Hamilton Depression Rating Scale. *Archives of General Psychiatry*. 1988; 45:742–747. [PubMed: 3395203]
- Wirth RJ, Edwards MC. Item factor analysis: Current approaches and future directions. *Psychological Methods*. 2007; 12:58–79. [PubMed: 17402812]

Table 1

Demographics and Hamilton Scores for MDD and PD Samples

	MDD (n=215)		PD (n=149)		
	n	%	n	%	
Sex	Female	115	53.5%	101	67.8%
Race	White	150	69.8%	113	75.8%
	Black/African American	47	21.9%	23	15.4%
Ethnicity	Asian	5	2.3%	5	3.4%
	Native American/Alaska Native	4	1.9%	2	1.3%
	Other	9	4.2%	6	4.0%
	Hispanic	8	3.7%	22	14.8%
	Non-Hispanic	179	83.3%	127	85.2%
	Unknown	28	13.0%	0	0.0%
Marital status	Married or Cohabitating	59	27.4%	70	47.0%
	Not Married or Cohabitating	155	72.1%	79	53.0%
Education	Unknown	1	0.5%	0	0.0%
	High School Diploma or Less	67	31.2%	13	8.7%
Some College or 2 Year College Degree		55	25.6%	39	26.2%
	4 Year College Degree	46	21.4%	44	29.5%
	Any Graduate School	28	15.8%	52	34.9%
Comorbidity	Unknown	13	6.0%	1	0.7%
	Social Phobia	0	0%	32	21.5%
	Generalized Anxiety Disorder	0	0%	59	39.6%
	Obsessive-Compulsive Disorder	0	0%	4	2.7%
	Specific Phobia	0	0%	29	19.5%
	Posttraumatic Stress Disorder	0	0%	8	5.4%
Any Comorbid Anxiety Disorder	0	0%	92	61.7%	
Age (years)	Mean	SD	Mean	SD	
	39.7	12.31	38.3	13.42	

	MDD (<i>n</i> =215)		PD (<i>n</i> =149)	
	<i>n</i>	%	<i>n</i>	%
	Median	Range	Median	Range
<i>17-item HRSD</i>	20	13 – 30	10	1 – 22
<i>24-item HRSD</i>	26	14 – 40	12	1 – 30
<i>HRSD-R-II</i>	21	12 – 33	7	0 – 22
<i>HARS</i>	15	3 – 31	13	2 – 33
<i>HARS-R-II</i>	13	0 – 28	15	1 – 33

Note. HRSD=Hamilton Rating Scale for Depression; HARS=Hamilton Anxiety Rating Scale; MDD=Major Depressive Disorder; PD=Panic Disorder; HRSD-R-II=Hamilton Rating Scale for Depression – Reconstructed – II; HARS-R-II=Hamilton Anxiety Rating Scale – Reconstructed – II.

HRSD and HARS Item Geomin-rotated Factor Loadings, Assignments to Revised Scales, Medians, and Ranges

Table 2

Item	MDD sample				PD sample			
	Depression: Vegetative loading	Depression: Cognitive loading	Anxiety loading	Revised scale assignment	Median	Range	Median	Range
HRSD								
1. Depressed mood	.50	.56	-.14	Depression	3	1-4	1	0-3
2. Guilt	.14	.62	-.01	Depression	2	0-3	0	0-3
3. Suicide	.32	.50	-.07	Depression	1	0-3	0	0-2
4. Insomnia early	.58	.03	.15	Depression	1	0-2	0	0-2
5. Insomnia middle ^a	.63	-.09	.12		1	0-2	0	0-2
6. Insomnia late	.46	.03	.19	Depression	0	0-2	0	0-2
7. Work and activities	.54	.52	-.11	Depression	3	1-4	0	0-3
8. Retardation	.40	.29	-.21	Depression	1	0-3	0	0-2
9. Agitation ^b	-.09	.35	.18	Anxiety	0	0-2	0	0-2
10. Anxiety psychic	.00	.38	.57	Anxiety	2	0-3	2	0-3
11. Anxiety somatic	.16	-.10	.62	Anxiety	2	0-3	2	0-4
12. Appetite	.41	.22	.00	Depression	0	0-2	0	0-2
13. Energy	.50	.48	-.02	Depression	2	0-2	0	0-2
14. Libido	.61	.14	.03	Depression	1	0-2	0	0-2
15. Hypochondriasis	-.03	-.08	.40	Anxiety	0	0-3	0	0-3
16. Weight Loss	.33	.25	-.09	Depression	0	0-2	0	0-2
17. Insight	-.07	.14	-.08		0	0-1	0	0-1
18. Diurnal variation	.11	.15	.10	Depression	1	0-2	0	0-2
19. Depersonalization and derealization	-.21	.25	.30	Anxiety	0	0-3	0	0-2
20. Paranoia ^a	-.01	.33	.13		0	0-2	0	0-2
21. Obsessions and compulsions ^c	-.18	.17	.21		0	0-2	0	0-2
22. Helplessness	.00	.72	.21	Depression	1	0-3	0	0-3
23. Hopelessness	.02	.69	.06	Depression	2	0-3	0	0-3

Item	Depression: Vegetative loading	Depression: Cognitive loading	Anxiety loading	Revised scale assignment	MDD sample			PD sample		
					Median	Range	Median	Median	Range	Range
24. Worthlessness	-.05	.81	.06	Depression	2	0 - 4	0	0	0 - 3	
HARS										
1. Anxious mood	-.02	.47	.65	Anxiety	2	0 - 4	2	2	0 - 4	
2. Tension	.14	.43	.58	Anxiety	2	0 - 4	2	2	0 - 4	
3. Fears	-.14	.04	.27	Anxiety	0	0 - 3	0	0	0 - 4	
4. Insomnia ^d	.79	-.04	.30		2	0 - 4	1	1	0 - 4	
5. Intellectual	.19	.42	.22	Depression	2	0 - 4	1	1	0 - 3	
6. Depressed mood ^d	.53	.57	.05		3	0 - 4	1	1	0 - 3	
7. Somatic muscular	-.01	.04	.43	Anxiety	1	0 - 4	1	1	0 - 3	
8. Somatic sensory	.03	.10	.53	Anxiety	0	0 - 3	1	1	0 - 3	
9. Cardiovascular	-.17	-.17	.62	Anxiety	0	0 - 3	1	1	0 - 3	
10. Respiratory	-.14	-.02	.57	Anxiety	0	0 - 3	1	1	0 - 4	
11. Gastrointestinal	.16	.04	.44	Anxiety	1	0 - 3	1	1	0 - 3	
12. Genitourinary ^d	.64	.11	.17		1	0 - 4	0	0	0 - 3	
13. Autonomic	.25	-.02	.53	Anxiety	1	0 - 3	1	1	0 - 3	
14. Behavior at interview ^e	-.10	.29	.29	Anxiety	0	0 - 3	0	0	0 - 2	

Note: Positive loadings that are significant at $p < .05$ are listed in bold. HRSD=Hamilton Rating Scale for Depression; HARS=Hamilton Anxiety Rating Scale; MDD=Major Depressive Disorder; PD=Panic Disorder.

^aDespite a significant loading on the depression factor, this item was excluded from the HRSD-R-II in an effort to remain consistent with Riskind et al.'s (1987) findings.

^bAgitation was rarely observed in our sample, restricting the range on this variable for psychometric analysis. Therefore, despite a slightly higher loading on the depression factor as compared to the anxiety factor, this item was included in the HARS-R-II in an effort to remain consistent with Riskind et al.'s (1987) findings.

^cDespite a significant loading on the anxiety factor, this item was excluded from the HARS-R-II in an effort to remain consistent with Riskind et al.'s (1987) findings.

^dDespite a significant loading on the depression factor, this item was excluded from the HRSD-R-II due to redundancy with items already included on the HRSD-R-II.

^eDespite a significant loading on the depression factor as well as the anxiety factor, this item was included in the HARS-R-II in an effort to remain consistent with Riskind et al.'s (1987) findings.

Table 3

Receiver Operating Characteristic Curve Analyses for HRSD-R-II Scores with Diagnosis of MDD: Sensitivity, Specificity, and Positive and Negative Predictive Power

HRSD-R-II score >	Sensitivity	Specificity	Positive Predictive Power	Negative Predictive Power
1	1.00	0.09	0.60	1.00
2	1.00	0.17	0.63	1.00
3	1.00	0.21	0.63	1.00
4	1.00	0.30	0.66	1.00
5	1.00	0.40	0.69	1.00
6	1.00	0.47	0.72	1.00
7	1.00	0.56	0.76	1.00
8	1.00	0.62	0.78	1.00
9	1.00	0.68	0.81	1.00
10	1.00	0.73	0.84	1.00
11	1.00	0.81	0.88	1.00
12	0.99	0.84	0.89	0.98
13	0.97	0.86	0.90	0.95
14	0.95	0.89	0.92	0.93
15	0.91	0.94	0.95	0.88
16	0.86	0.95	0.96	0.83
17	0.79	0.97	0.97	0.77
18	0.70	0.98	0.98	0.71
19	0.64	0.98	0.98	0.67
20	0.56	0.98	0.97	0.62
21	0.49	0.98	0.97	0.58
22	0.39	1.00	1.00	0.54
23	0.32	1.00	1.00	0.52
24	0.21	1.00	1.00	0.48
25	0.16	1.00	1.00	0.46
26	0.11	1.00	1.00	0.45
27	0.07	1.00	1.00	0.44
28	0.04	1.00	1.00	0.43
29	0.02	1.00	1.00	0.43
30	0.01	1.00	1.00	0.42
31	0.01	1.00	1.00	0.42

Note. HRSD-R-II=Hamilton Rating Scale for Depression – Reconstructed – II; MDD=Major Depressive Disorder.

Table 4

Receiver Operating Characteristic Curve Analyses for HARS-R-II Scores with Diagnosis of PD: Sensitivity, Specificity, and Positive and Negative Predictive Power

HARS-R-II score >	Sensitivity	Specificity	Positive Predictive Power	Negative Predictive Power
1	0.99	0.00	0.42	0.50
2	0.99	0.01	0.42	0.67
3	0.97	0.01	0.42	0.29
4	0.96	0.02	0.41	0.45
5	0.94	0.04	0.41	0.47
6	0.92	0.09	0.41	0.47
7	0.88	0.13	0.42	0.60
8	0.82	0.19	0.42	0.60
9	0.76	0.25	0.42	0.59
10	0.73	0.34	0.44	0.60
11	0.69	0.39	0.45	0.64
12	0.65	0.47	0.47	0.63
13	0.61	0.54	0.49	0.65
14	0.56	0.66	0.54	0.66
15	0.50	0.72	0.56	0.67
16	0.42	0.79	0.59	0.66
17	0.37	0.85	0.65	0.66
18	0.30	0.87	0.63	0.63
19	0.24	0.89	0.61	0.62
20	0.20	0.93	0.66	0.62
21	0.16	0.94	0.67	0.61
22	0.13	0.97	0.76	0.61
23	0.09	0.99	0.81	0.60
24	0.07	0.99	0.83	0.60
25	0.05	0.99	0.80	0.59
26	0.04	1.00	0.86	0.59
27	0.03	1.00	0.80	0.59
28	0.03	1.00	1.00	0.59
29	0.02	1.00	1.00	0.59
30	0.02	1.00	1.00	0.59
31	0.01	1.00	1.00	0.59
32	0.01	1.00	1.00	0.59

Note. HARS-R-II=Hamilton Anxiety Rating Scale – Reconstructed – II; PD=Panic Disorder.