Published in final edited form as:

Eur Neuropsychopharmacol. 2006 December ; 16(8): 601–611.

The Montgomery Åsberg and the Hamilton Ratings of Depression:

A Comparison of Measures

Thomas Carmody, Ph.D.¹, A. John Rush, M.D.¹, Ira Bernstein, Ph.D.², Diane Warden, Ph.D., M.B.A.¹, Stephen Brannan, M.D.³, Daniel Burnham, Ph.D.⁴, Ada Woo, M.A.², and Madhukar Trivedi, M.D.¹

¹Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, TX

²University of Texas at Arlington, King of Prussia, PA

³Cyberonics, Inc., King of Prussia, PA

⁴GlaxoSmithKline, King of Prussia, PA

Abstract

The 17-item Hamilton Rating Scale for Depression (HRSD₁₇) and the Montgomery Åsberg Depression Rating Scale (MADRS) are two widely used clinician-rated symptom scales. A 6-item version of the HRSD (HRSD₆) was created by Bech to address the psychometric limitations of the HRSD₁₇. The psychometric properties of these measures were compared using classical test theory (CTT) and item response theory (IRT) methods. IRT methods were used to equate total scores on any two scales. Data from two distinctly different outpatient studies of nonpsychotic major depression: a 12-month study of highly treatment-resistant patients (n=233) and an 8-week acute phase drug treatment trial (n=985) were used for robustness of results.

MADRS and HRSD₆ items generally contributed more to the measurement of depression than HRSD₁₇ items as shown by higher item-total correlations and higher IRT slope parameters. The MADRS and HRSD₆ were unifactorial while the HRSD₁₇ contained 2 factors. The MADRS showed about twice the precision in estimating depression as either the HRSD₁₇ or HRSD₆ for average severity of depression. An HRSD₁₇ of 7 corresponded to an 8 or 9 on the MADRS and 4 on the HRSD₆.

The MADRS would be superior to the HRSD₁₇ in the conduct of clinical trials.

Keywords

MADRS; HRSD; item response theory; classical test theory; psychometrics

INTRODUCTION

The measurement of depressive symptom severity is important not only for the conduct of efficacy and effectiveness trials but increasingly also for the proper implementation of treatment guideline recommendations for major depressive and other mood disorders (Crismon et al., 1999; Depression Guideline Panel, 1993; Rush et al., 2003; Trivedi et al., 2004). A number of self-reports (e.g., Carroll Rating Scale, Beck Depression Inventory, Zung Self-Rating Scale, and Inventory of Depressive Symptomatology - Self Report) and clinician ratings (e.g., Hamilton Rating Scale for Depression, Montgomery Åsberg Depression Rating Scale,

and Inventory of Depressive Symptomatology - Clinician-rated) are available. Perhaps the two most popular clinical ratings are the Hamilton Rating Scale for Depression (HRSD), which comes in several versions (e.g., 17, 21, 24, 28, and 31 items) (Hamilton, 1960; Hamilton, 1967) and the 10-item Montgomery Åsberg Depression Rating Scale (MADRS) (Montgomery and Åsberg, 1979). The MADRS is used frequently in European registration and other clinical trials, while the Hamilton continues to be more widely used in the United States, though recent reports (Bagby et al., 2004; Zimmerman et al., 2005) have highlighted significant shortcomings in the HRSD.

The MADRS has been reported as equivalent to or more sensitive to change in symptoms over time than the HRSD₁₇ (Mulder et al., 2003; Rivera et al., 2000; Senra, 1996) and equivalent to the HRSD₁₇ in detecting drug/placebo differences (Khan et al., 2002). The MADRS has been reported to be unifactorial after treatment (Galinowski and Lehert, 1995; Rocca et al., 2002), although more than one factor has been found using ratings with more limited ranges in total score (i.e., prior to treatment) (Corruble et al., 1999; Craighead and Evans, 1996; Galinowski and Lehert, 1995; Hammond, 1998; Rocca et al., 2002). A meta-analysis (Faries et al., 2000), however, found that the superiority of the MADRS or the HRSD in detecting differences between drug and placebo depended on the class of the medication, and the specific effects and side effects of the medication.

The HRSD₁₇ has been found consistently to be multidimensional (Bech et al., 1981; Gibbons et al., 1993; Hamilton, 1967; Maier et al., 1988), which may reduce its sensitivity to detecting changes in depression severity or in differentiating between two treatments. Prior analyses of the HRSD₁₇ have identified specific problematic items in terms of response characteristics (Bagby et al., 2004; Santor and Coyne, 2001). Several briefer versions of the HRSD have been developed to improve upon the HRSD₁₇ by creating a more unifactorial measure of depression that, consequently, should be more sensitive to detecting changes in depression or to detecting drug/placebo differences than the HRSD₁₇. The most commonly used brief HRSD may be that developed by Bech (Bech et al., 1975) — a 6-item scale that includes the following items: depressed mood, guilt, work and activities, retardation, psychic anxiety, and somatic symptoms general. In fact, the HRSD₆ has been found to be more clearly unidimensional (Bagby et al., 2004; Bech et al., 1992; Bech et al., 1997; Bech et al., 1984; Bech et al., 1975), more sensitive to change than the HRSD₁₇ (de Montigny et al., 1981; O'Sullivan et al., 1997), and equivalent to (Hooper and Bakish, 2000) or more sensitive to detecting drug/placebo or drug/drug differences than the HRSD₁₇ (Bech et al., 2000; Faries et al., 2000). This briefer version appears to have less psychometric bias as a result of side effects (Moller, 2001).

Item response theory (IRT) models (Embretson and Reise, 2000; Hambleton and Swaminathan, 1985; Hulin et al., 1983; Nunnally and Bernstein, 1994) represent an important and increasingly sophisticated framework for examination of the psychometric properties of rating scale total scores and individual items. The Rasch model has been used to examine the psychometric properties of the HRSD₁₇ (Bech et al., 1981) and recently the more complex Samejima's graded IRT model has been used to further examine the HRSD₁₇ (Rush et al., 2005a). Unlike classical test theory (CTT) methods, IRT methods can be used to create conversion tables that allow a reliable crosswalk between total scale scores (Orlando et al., 2000). Such conversion tables allowing a crosswalk between the HRSD₁₇, MADRS, and Bech's HRSD₆ would greatly facilitate extrapolating findings from published reports that use one scale to allow reliable estimates of the same study results using the alternative scale (e.g., MADRS to HRSD₁₇). In addition, such information would clarify what total symptom scores are comparable in defining remission, as well as mild, moderate, and severe symptom levels. Ideally, the validity of such conversion tables would be higher if they were developed from a large diverse study sample or samples.

This report provides both CTT and IRT results on two distinctly different depressed outpatient samples, each of which were developed for different research purposes. These results provide an empirical basis for converting one scale total score into another scale total score. Further, since the HRSD₁₇ and the MADRS were collected at the same time on each of two samples, these two clinician ratings, as well as the HRSD₆, which was extracted from the HRSD₁₇, can be compared on the basis of item response and other psychometric features.

EXPERIMENTAL PROCEDURES

Two datasets were analyzed for this report. The first (Study 1) (n=233) was generated from a 12-month uncontrolled, long-term study of adult outpatients (18-75 years old) with highly treatment-resistant, nonpsychotic major depressive episodes (MDEs) who participated in a study of adjunctive vagus nerve stimulation added onto ongoing diverse medication regimens (Rush et al., 2005b). Diagnoses were rendered with the Structured Clinical Interview for DSM-IV (SCID) (First et al., 1994). This population included 208 (89.3%) patients with major depressive disorder and 25 (10.7%) in a depressed phase of bipolar I (n=12) or bipolar II (n=13) disorder. The baseline features of this sample included 62.2% female with an average age 47.2 (SD=8.9) (range: 24 to 72), 96.6% Caucasian with an average baseline HRSD₁₇ total score of 21.9 (SD=4.4) (range: 13 to 37), an average baseline HRSD₆ total score of 12.4 (SD=2.5) (range: 6 to 19), and an average baseline MADRS total score of 31.9 (SD=6.7) (range: 14 to 50).

This patient group had not responded adequately to 2-6 trials of known effective treatments delivered at adequate doses and durations in the current MDE as assessed by the Antidepressant Treatment History Form (Oquendo et al., 1999; Prudic et al., 1996; Prudic et al., 1990; Sackeim et al., 1990; Sackeim et al., 2000; Sackeim, 2001). When counting all clinical treatments received, patients had on average received over 12 different medications in the current MDE.

Raters were not blind to treatment when the Study 1 data analyzed in this report were collected. Data at study exit (or the date closest to 12 months following study initiation) were subjected to analysis. Patients with no post-baseline data were excluded. The HRSD₂₈ was collected using a structured interview modeled after Williams et al. (Williams, 1988). The HRSD₁₇ was extracted from the HRSD₂₈ for these analyses. Study 1 data were supplied by Cyberonics, Inc.

The second sample (Study 2) (n=985) included only outpatients with nonpsychotic major depressive disorder (MDD) defined by DSM-IV. A complete psychiatric evaluation following APA guidelines (American Psychiatric Association, 2000) was performed by a psychiatrist after a screening interview using the Mini International Neuropsychiatric Interview (MINI) (Sheehan et al., 1998). These subjects were randomized to one of three treatment cells (placebo, standard antidepressant, experimental antidepressant) for an 8-week double-blind treatment phase. A total score of 20 on the HRSD₁₇ and an item 1 (sad mood) score of 2 were required for study entry. Treatment-resistant patients were excluded. Altogether, 59.9% were female with an average age 39.8 (SD=11.6) (range: 18-65); 81.5% were Caucasian. The average baseline HRSD₁₇ total score was 23.6 (SD=2.9) (range: 20 to 35); the average baseline HRSD₆ total score was 12.8 (SD=1.7) (range: 7 to 18). The average baseline MADRS total score was 28.6 (SD=5.1) (range: 10 to 45). Raters were blind to treatment assignment. Raters were trained in the conventional way by completing the HRSD₁₇ while watching videotaped interviews. For Study 2, the HRSD₁₇ was obtained without a structured interview. For these analyses, data were supplied by GlaxoSmithKline.

These two data sets were chosen to maximize the generalizability of the findings in this report. Exit ratings were chosen to provide a maximum range in symptom scores. In both data sets,

the HRSD₁₇ and MADRS were collected by the same evaluator in the same interview. The HRSD₆ was extracted and totaled from the HRSD₁₇ items.

Statistical Analyses

Classical test theory (CTT) measures of consistency: Cronbach's alpha (α) (Cronbach, 1951) and item-total correlations (not corrected for item/total overlap), were computed for the HRSD₁₇, MADRS, and HRSD₆ at study exit for each study. Also, effect sizes were computed for each total score and item for each measure within each study. The effect size was computed as the exit item or total score minus the baseline item or total score divided by the standard deviation of the change in item or total score. Let $D = X_{Base} - X_{Exit}$, where X_{Base} and X_{Exit} are baseline and exit scores and D , is their difference for a given item. The effect size (E) is then simply D/S_d , where D is the mean of $\underline{D}$ and S_d is its standard deviation.

IRT methods were used to equate total scores for each pair of scales. Samejima's graded IRT model (Samejima, 1997) item parameters were estimated for each item of each measure. These parameters were used according to the procedure of Orlando et al. (Orlando et al., 2000) (and associated software) to generate an IRT score for each possible total score on the HRSD₁₇, MADRS, and HRSD₆. The IRT score, usually called theta, is a unitless measure of depression estimated from the IRT procedure commonly scaled to a mean of 0 and a standard deviation of 1. The total scores for each pair of scales were equated by matching the IRT scores for the two corresponding scales (Orlando et al., 2000). When an IRT score did not match exactly, best judgment was used to equate the scales taking into account the matching of total scores immediately above and below the total score in question.

The graded IRT model was also used to compute the test information function (TIF) (Birnbaum, 1968) for each scale in each study. The "information" provided by a scale is defined to be the inverse of the standard error of the total score of the scale. Thus, a total score that provides a precise estimate of symptom severity contains more "information" than a total score that provides an imprecise estimate. The TIF allows one to see at which levels of symptom severity any given measure's total scores provide the most precise estimates and also to compare the precision of two or more measures across all levels of symptom severity.

An assumption of the IRT approach is that the measures assess only depression (i.e., are unidimensional, see the above IRT references). Therefore, a principal components factor analysis was conducted to assess the dimensionality of each measure in each study. Parallel analysis (Horn, 1965; Humphreys and Ilgen, 1969; Humphreys and Montanelli, 1975; Montanelli and Humphreys, 1976) was used to infer how many "real" factors (dimensions) were present in the data. This approach avoids some of the limitations of the traditional Kaiser-Guttman eigenvalue-greater-than-1 rule. Parallel analysis involves comparing the eigenvalues from a principal components analysis of the real data to eigenvalues that might be expected to arise by chance alone. To determine how large the latter eigenvalues might be, we generated a series of simulated datasets consisting of random numbers (where correlations between all variables are zero) using the same number of observations and variables (items) as the real data. Eigenvalues of the principal components for each simulated dataset are computed and averaged over replications. The largest real data eigenvalue is compared to the largest simulated data eigenvalue, then the second largest real and simulated eigenvalues are compared, and so on until we find a real eigenvalue smaller than the corresponding simulated eigenvalue. The number of principal components for which the real eigenvalues exceed the simulated eigenvalues defines the dimensionality.

These analyses were applied to (1) the 17-item Hamilton Rating Scale for Depression (HRSD₁₇), (2) the 10-item Montgomery Åsberg Depression Rating Scale (MADRS), and (3)

the 6-item Hamilton Rating Scale for Depression (HRSD₆) defined by Bech et al. (Bech et al., 1975).

RESULTS

Classical Test Theory Analyses

Correlations—All three measures were highly correlated with each other at exit in both studies. In Study 1, the correlation between the HRSD₁₇ and HRSD₆ total scores was 0.89; between the HRSD₁₇ and MADRS, the correlation was 0.88, and between the HRSD₆ and MADRS, the correlation was 0.86. In Study 2, all the correlations were slightly higher: HRSD₁₇ vs. HRSD₆ was 0.94, HRSD₁₇ vs. MADRS was 0.92, and HRSD₆ vs. MADRS was 0.91.

Internal Consistency—Cronbach's alpha showed highly acceptable internal consistency for all measures using study exit data. For the HRSD₁₇, the values were 0.81 (Study 1) and 0.88 (Study 2). For the MADRS, values were slightly higher: 0.90 (Study 1) and 0.92 (Study 2). Finally, for the HRSD₆, the values were 0.78 (Study 1) and 0.86 (Study 2).

Item Total Correlations—Table 1 summarizes the item total correlations for each test in each study. Most items on the MADRS correlated with the total score at ≥ 0.60 (both studies). For the HRSD₁₇, only 4 of 17 items (Study 1) and only 6 of 17 items (Study 2) correlated at ≥ 0.60 with the total score. In addition, median item-total correlations were 0.75 (Study 1) and 0.78 (Study 2) for the MADRS. HRSD₆ item-total correlations were about the same magnitude as for the MADRS. For the HRSD₁₇ median item total correlations were lower (0.50 for Study 1 and 0.56 for Study 2).

Item Response Theory Analyses

Item Performance—Table 2 summarizes the item performance for the HRSD₁₇ for each study. Tables 3 and 4 provide similar information for the MADRS and HRSD₆, respectively. Items with numerically higher slopes ("a" values) contribute more to the measurement of depression because these items are better able to discriminate among different levels of depression. For the HRSD₁₇, only 7/17 items had a value ≥ 1.0 (Study 1), while for Study 2, 13/17 items were ≥ 1.0 . For the MADRS, 9/10 items in each study had slopes of ≥ 1.0 . For the HRSD₆, all items in both studies had slopes ≥ 1.0 .

Dimensionality—Principal components factor analyses were conducted on each measure for each study. For the HRSD₁₇ in Study 1, two factors were identified using parallel analysis to determine the number of factors. The average of the first three eigenvalues from the simulated datasets were 1.50, 1.39, and 1.31, which were compared to the first 3 real eigenvalues of 4.33, 1.73, and 1.19. Two factors were chosen because the first two real data eigenvalues were larger than the first two simulated data eigenvalues. After oblique rotation, factor 1 included items for depressed mood (1), suicide (3), work and activities (7), retardation (8), psychic anxiety (10), somatic symptoms general (13), and libido (14). These items had loadings greater than 0.3 on the first factor and less than 0.1 on the second factor. Factor 2 included initial insomnia (4), middle insomnia (5), delayed insomnia (6), and reduced appetite/gastrointestinal symptoms (12). The other 6 items did not clearly belong to either factor. These items had loadings of similar magnitude on both factors.

The HRSD₁₇ in Study 2 also revealed two factors based on the comparison of the first 3 simulated data eigenvalues of 1.23, 1.19, and 1.15 to real data eigenvalues of 5.77, 1.30, and 1.11. After oblique rotation, factor 1 included depressed mood (1), guilt (2), suicide (3), work and activities (7), psychic anxiety (10), somatic symptoms general (13), and hypochondriasis

(15). Factor 2 items included initial insomnia (4), middle insomnia (5), and delayed insomnia (6). The other 7 items did not clearly belong to either factor.

For the MADRS, only one factor was identified for Study 1 because the first real eigenvalue of 5.41 was much larger than the first simulated eigenvalue of 1.33, while the second real eigenvalue of 1.06 was smaller than the second simulated eigenvalue of 1.23. One MADRS factor was also identified for Study 2 based on simulated eigenvalues of 1.15 and 1.11 compared with real eigenvalues of 6.00 and 0.91.

Comparison of simulated versus real data eigenvalues showed the HRSD₆ to be unifactorial in both Study 1 and Study 2. In Study 1, the first 2 eigenvalues were 1.21 and 1.11 (simulated) versus 2.92 and 0.92 (real). In Study 2, the first 2 eigenvalues were 1.10 and 1.06 (simulated) versus 3.57 and 0.73 (real).

Conversion Tables—Table 5 summarizes the IRT conversions for each rating scale total score for Study 1 and Study 2 combined. In the combined sample, an HRSD₁₇ of 7 was comparable to a MADRS total score of 8 or 9, and an HRSD₆ total score of 4.

Relative Precision—Figures 1 and 2 show the test information functions (TIFs) for each scale in Study 1 and Study 2, respectively. In these figures theta represents a unitless measure of depression estimated from the IRT procedure scaled to a mean of 0 and a standard deviation of 1. In Study 1, the TIF for the MADRS was over twice that of the HRSD₁₇ or HRSD₆ for thetas from -1 to +2 indicating that for patients with average levels of depression, the MADRS was about 2 times as precise as the HRSD₁₇. For very high or low levels of depression, the differences between scales were not as great. Similar findings were seen for Study 2. The small improvement in precision from the HRSD₆ to the HRSD₁₇ showed that most of the precision of the HRSD₁₇ came from the 6 items included in the HRSD₆. The remaining 11 items did not add substantially to the precision of the HRSD₁₇.

Item Effect Sizes—Table 6 displays the item and total score effect sizes (baseline to exit) for each scale in each study. Not surprisingly, the more treatment-resistant sample (Study 1) had lower overall item and total score effect sizes with each of the three measures. The items on the MADRS with the highest effect size (in both studies) were reported sadness, apparent sadness, concentration, lassitude, inability to feel, and pessimistic thoughts. For the HRSD₁₇, the items with the largest effect sizes included depressed mood, work and activities, guilt, retardation, psychic anxiety, and somatic symptoms general.

DISCUSSION

When the HRSD₁₇, MADRS, and HRSD₆ were compared across two different populations, the total scores obtained at study exit on all three measures were highly correlated. The MADRS consistently had the highest Cronbach alpha levels. Both the MADRS and the HRSD₆ were unifactorial. The HRSD₁₇ had two factors, with 6-7 items not loading on either factor or loading on both factors. This is consistent with prior reports of the multidimensionality of the HRSD₁₇ and unidimensionality of the MADRS and the HRSD₆ (Bagby et al., 2004; Bech et al., 1992; Bech et al., 1997; Bech et al., 1984; Bech et al., 1981; Bech et al., 1975; Galinowski and Leher, 1995; Gibbons et al., 1993; Hamilton, 1967; Maier et al., 1988; Rocca et al., 2002).

Given the unifactorial nature of the MADRS and the HRSD₆, it is not surprising that more MADRS items and more HRSD₆ items evidenced high (≥ 0.60) item total correlations as compared to items on the HRSD₁₇. IRT analyses revealed that for the HRSD₁₇, depressed mood, work and activities, somatic symptoms general, suicide, psychic anxiety, and guilt were

most highly reflective of the core concept of depression (both studies) as shown by larger values of the slope (a) parameter. All items on the HRSD₆ related highly to core depression. For the MADRS, IRT analyses revealed that most (8/10) items related highly to the core concept of depression (both studies). Reduced sleep and reduced appetite on the MADRS also related to overall depression, but less so than the other 8 items.

For the HRSD₁₇, IRT analyses revealed several items that were minimally related to the core concept of depression (e.g., hypochondriasis, loss of insight, weight loss, agitation). For the MADRS, only reduced sleep in Study 1 and reduced appetite in Study 2 showed a similarly weak relationship to the core concept of depression, and no HRSD₆ items showed such a weak relationship to core depression.

A majority of HRSD₁₇, MADRS, and HRSD₆ items rarely exhibited the full range of response options (where b0, b1, b2, etc. were >2.0 or 7 score units), which is understandable given the outpatient nature of the samples. However, for 7 HRSD₁₇ items, 2 MADRS items, and 2 HRSD₆ items, the most severe range of responses were endorsed so infrequently that the “b” parameters were either very large (20+) or unestimable (e.g., HRSD₁₇: guilt, suicide, retardation, agitation, hypochondriasis, loss of insight, and weight loss, MADRS: pessimism and suicide, HRSD₆: guilt and retardation).

Conversion tables were generated by combining data from both studies. An HRSD₁₇ of 7 approximated a MADRS of 8 or 9 (i.e., 8.5). An HRSD₁₇ total of 20 approximated a MADRS of 27. In general, results were comparable to those reported by Hawley et al. (Hawley et al., 1998) who recommended that $MADRS = 1.3 HRSD_{17} + 0.7$ based on a regression analysis.

All MADRS items had acceptable effect sizes, and were therefore sensitive to change over time, although the sleep and appetite items were less sensitive to change. The MADRS had a consistent order to the effect size for each item in both studies, which revealed the consistency of the MADRS items in measuring the degree to which symptoms improved in two very different depressed samples. The items with the greatest effect size for change over time in the HRSD₁₇ were the six items assessed by the HRSD₆ (e.g., work and activities, sad mood, psychic anxiety, etc.). HRSD₁₇ items with the lowest effect size (both studies) included loss of insight, weight loss, hypochondriasis, libido, appetite, agitation, and the three insomnia items. The inclusion of items that are relatively less sensitive to change over time risks decreased power to detect change over time. These results support the conclusion that the MADRS is preferred over the HRSD₁₇ in measuring depression severity and change in depression severity over time given its unifactorial structure, the high and consistent relationship between items and the measured concept of depression (by IRT) or to total score (by CTT), and its greater precision (Figures 1 and 2). This report, however, did not compare the scales in terms of drug vs. placebo effect sizes.

A variety of additional factors also recommend the MADRS over the HRSD₁₇. These factors include: (1) each MADRS item measures a single symptom whereas some HRSD₁₇ items measure two related but different concepts, such as work and activities, making it difficult to relate the item score to a specific symptom; (2) the even weighting of items by the MADRS such that each symptom item may contribute equally to the total score; (3) the use of a broader range of responses to individual items with the MADRS as opposed to the HRSD₁₇, suggesting increased sensitivity of the items to change, and (4) the 10-item MADRS typically takes 15 minutes to complete compared with 15 to 20 minutes for the 17-item HRSD (Rush et al., 2000). In addition, the HRSD₁₇ is known to perform better with a structured interview and trained raters, while there is no need for a structured interview with the MADRS. In fact, the MADRS outperformed the HRSD₁₇ in Study 1, even when a structured interview was used. From the point of view of enhancing the cost efficiency of trials, the MADRS should be the

first choice, since it typically takes no longer to complete than the HRSD₁₇, yet it provides substantially better precision in assessing symptom severity.

These conclusions are strengthened by the consistency of results across two very different patient populations and by the fact that in both studies the HRSD₁₇ and MADRS were administered at the same visit thus eliminating as many extraneous influences as possible. However, the administration of both measures at the same interview by the same interviewer may have lead to the subtle influence of responses to one interview affecting the other. Since the order of the two measures was not strictly randomized, we cannot estimate the effect of this potential confound.

Neither the HRSD₁₇ nor the MADRS, however, assess all of the core criterion symptoms used in DSM IV TR to diagnose a major depressive episode. In that regard, neither is fully adequate to define severity of depression or remission - assuming remission refers to the virtual absence of core criterion symptoms of MDD. The HRSD₁₇ lacks ratings of oversleeping, overeating, and concentration. The MADRS lacks ratings of oversleeping and overeating, as well as interest (though it assesses inability to feel), energy (though it assesses lassitude), self criticism (guilt), and psychomotor changes.

Conclusion

These data provide a strong basis for selecting the MADRS rather than the HRSD₁₇ in the conduct of efficacy and effectiveness trials, given its unifactorial nature and other preferable psychometric properties, and the avoidance of a structured interview and training (for experienced clinicians). How other clinical or self-report depression measures compare with the MADRS deserves study.

Acknowledgements

This project was funded in part by the National Institute of Mental Health (NIMH), National Institutes of Health (MH-68851 to the University of Texas Southwestern Medical Center at Dallas, A. John Rush, M.D., PI, and by MH-68852 to the University of Texas at Arlington, Ira H. Bernstein, Ph.D., PI).

The authors wish to express their appreciation to both Cyberonics Inc. and GlaxoSmithKline for providing the datasets used in this report. The authors personally conducted these analyses and received no support for conducting these analyses or preparing this report. A. John Rush, M.D. has received payments as a consultant and speaker for both GlaxoSmithKline and Cyberonics Inc. Madhukar Trivedi, M.D. has received payments as a consultant to Cyberonics. Daniel Burnham, M.D. is an employee and stockholder of GlaxoSmithKline. Steve Brannan, M.D. is an employee and stockholder of Cyberonics, Inc.

REFERENCES

- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed, Text Revision. American Psychiatric Press; Washington DC: 2000.
- Bagby RM, Ryder AG, Schuller DR, Marshall MB. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *Am. J. Psychiatry* 2004;161(12):2163–2177. [PubMed: 15569884]
- Bech P, Allerup P, Gram LF, Reisby N, Rosenberg R, Jacobsen O, Nagy A. The Hamilton Depression Scale. Evaluation of objectivity using logistic models. *Acta Psychiatr. Scand* 1981;63:290–299. [PubMed: 7015793]
- Bech P, Allerup P, Maier W, Albus M, Lavori P, Ayuso JL. The Hamilton scales and the Hopkins Symptom Checklist (SCL-90). A cross-national validity study in patients with panic disorders. *Br. J. Psychiatry* 1992;160:206–211. [PubMed: 1540760]
- Bech P, Allerup P, Reisby N, Gram LF. Assessment of symptom change from improvement curves on the Hamilton depression scale in trials with antidepressants. *Psychopharmacology* 1984;84:276–281. [PubMed: 6438690]

- Bech P, Cialdella P, Haugh MC, Birkett MA, Hours A, Boissel JP, Tollefson GD. Meta-analysis of randomised controlled trials of fluoxetine v. placebo and tricyclic antidepressants in the short-term treatment of major depression. *Br. J. Psychiatry* 2000;176:421–428. [PubMed: 10912216]
- Bech P, Gram LF, Dein E, Jacobsen O, Vitger J, Bolwig TG. Quantitative rating of depressive states. *Acta Psychiatr. Scand* 1975;51(3):161–170. [PubMed: 1136841]
- Bech P, Stage KB, Nair NP, Larsen JK, Kragh-Sorensen P, Gjerris A. The Major Depression Rating Scale (MDS). Inter-rater reliability and validity across different settings in randomized moclobemide trials. Danish University Antidepressant Group. *J. Affect. Disord* 1997;42(1):39–48. [PubMed: 9089057]
- Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In: Lord, FM.; Novick, MR., editors. *Statistical Theories of Mental Test Scores*. Addison-Wesley; Reading, MA: 1968.
- Corruble E, Legrand JM, Duret C, Charles G, Guelfi JD. IDS-C and IDS-SR: psychometric properties in depressed in-patients. *J. Affect. Disord* 1999;56(23):95–101. [PubMed: 10701466]
- Craighead WE, Evans DD. Factor analysis of the Montgomery-Asberg Depression Rating Scale. *Depression* 1996;4(1):31–33. [PubMed: 9160651]
- Crismon ML, Trivedi M, Pigott TA, Rush AJ, Hirschfeld RM, Kahn DA, DeBattista C, Nelson JC, Nierenberg AA, Sackeim HA, Thase ME. The Texas Medication Algorithm Project: report of the Texas Consensus Conference Panel on Medication Treatment of Major Depressive Disorder. *J. Clin. Psychiatry* 1999;60(3):142–156. [PubMed: 10192589]
- Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297–334.
- de Montigny C, Grunberg F, Mayer A, Deschenes JP. Lithium induces rapid relief of depression in tricyclic antidepressant drug non-responders. *Br. J. Psychiatry* 1981;138:252–256. [PubMed: 7272619]
- Depression Guideline Panel. *Treatment of Major Depression*. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research; Rockville, MD: 1993. Clinical Practice Guideline, Number 5: Depression in Primary Care: Volume 2.
- Embretson, SE.; Reise, SP. *Item Response Theory for Psychologists*. Lawrence E. Erlbaum Associates; Mahwah, NJ: 2000.
- Faries D, Herrera J, Rayamajhi J, DeBrotta D, Demitrack M, Potter WZ. The responsiveness of the Hamilton Depression Rating Scale. *J. Psychiatr. Res* 2000;34(1):3–10. [PubMed: 10696827]
- First, MB.; Spitzer, R.; Gibbon, M.; Williams, J. *Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I), Patient Edition*. NY State Psychiatric Institute Biometrics Research Department; New York: 1994.
- Galinowski A, Leher P. Structural validity of MADRS during antidepressant treatment. *Int. Clin. Psychopharmacol* 1995;10(3):157–161. [PubMed: 8675968]
- Gibbons RD, Clark DC, Kupfer DJ. Exactly what does the Hamilton Depression Rating Scale measure? *J. Psychiatr. Res* 1993;27:259–273. [PubMed: 8295158]
- Hambleton, RK.; Swaminathan, H. *Item Response Theory*. Kluwer-Nijhoff; Boston: 1985.
- Hamilton M. A rating scale for depression. *J. Neurol. Neurosurg. Psychiatry* 1960;23:56–62. [PubMed: 14399272]
- Hamilton M. Development of a rating scale for primary depressive illness. *Br. J. Soc. Clin. Psychol* 1967;6(4):278–296. [PubMed: 6080235]
- Hammond MF. Rating depression severity in the elderly physically ill patient: reliability and factor structure of the Hamilton and the Montgomery-Asberg Depression Rating Scales. *Int. J. Geriatr. Psychiatry* 1998;13(4):257–261. [PubMed: 9646154]
- Hawley CJ, Gale TM, Smith VRH, Sen P. Depression rating scales can be related to each other by simple equations. *Int. J. Psychiatry Clin. Pract* 1998;2:215–219.
- Hooper CL, Bakish D. An examination of the sensitivity of the six-item Hamilton Rating Scale for Depression in a sample of patients suffering from major depressive disorder. *J. Psychiatry Neurosci* 2000;25(2):178–184. [PubMed: 10740991]
- Horn JL. An empirical comparison of various methods for estimating common factor scores. *Educ. Psychol. Meas* 1965;25:313–322.

- Hulin, CL.; Dragow, F.; Parsons, CK. Item-Response Theory: Applications to Psychological Measurement. Dow Jones Irwin; Homewood, IL: 1983.
- Humphreys LG, Ilgen D. Note on a criterion for the number of common factors. *Educ. Psychol. Meas* 1969;29:571–578.
- Humphreys LG, Montanelli RG Jr. An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behav. Res* 1975;10:193–206.
- Khan A, Khan SR, Shankles EB, Polissar NL. Relative sensitivity of the Montgomery-Asberg Depression Rating Scale, the Hamilton Depression rating scale and the Clinical Global Impressions rating scale in antidepressant clinical trials. *Int. Clin. Psychopharmacol* 2002;17(6):281–285. [PubMed: 12409681]
- Maier W, Philipp M, Heuser I, Schlegel S, Buller R, Wetzel H. Improving depression severity assessment--I. Reliability, internal validity and sensitivity to change of three observer depression scales. *J. Psychiatr. Res* 1988;22:3–12. [PubMed: 3397908]
- Moller HJ. Methodological aspects in the assessment of severity of depression by the Hamilton Depression Scale. *Eur. Arch. Psychiatry Clin. Neurosci* 2001;251(Suppl 2):1113–1120.
- Montanelli RG Jr. Humphreys LG. Latent roots of random data correlation matrices with squared multiple correlations on the diagonal: a Monte Carlo study. *Psychometrika* 1976;41:341–348.
- Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. *Br. J. Psychiatry* 1979;134:382–389. [PubMed: 444788]
- Mulder RT, Joyce PR, Frampton C. Relationships among measures of treatment outcome in depressed patients. *J. Affect. Disord* 2003;76(13):127–135. [PubMed: 12943942]
- Nunnally, JC.; Bernstein, IH. *Psychometric Theory*. 3rd edn.. McGraw-Hill; New York: 1994.
- O'Sullivan RL, Fava M, Agustin C, Baer L, Rosenbaum JF. Sensitivity of the six-item Hamilton Depression Rating Scale. *Acta Psychiatr. Scand* 1997;95:379–384. [PubMed: 9197901]
- Oquendo MA, Malone KM, Ellis SP, Sackeim HA, Mann JJ. Inadequacy of antidepressant treatment for patients with major depression who are at risk for suicidal behavior. *Am. J. Psychiatry* 1999;156(2):190–194. [PubMed: 9989553]
- Orlando M, Sherbourne CD, Thissen D. Summed-score linking using item response theory: application to depression measurement. *Psychol. Assess* 2000;12(3):354–359. [PubMed: 11021160]
- Prudic J, Haskett RF, Mulsant B, Malone KM, Pettinati HM, Stephens S, Greenberg R, Rifas SL, Sackeim HA. Resistance to antidepressant medications and short-term clinical response to ECT. *Am. J. Psychiatry* 1996;153:985–992. [PubMed: 8678194]
- Prudic J, Sackeim HA, Devanand DP. Medication resistance and clinical response to electroconvulsive therapy. *Psychiatry Res* 1990;31:287–296. [PubMed: 1970656]
- Rivera CS, Perez CR, Cao ES. Use of three depression scales for evaluation of pretreatment severity and of improvement after treatment. *Psychol. Rep* 2000;87:389–394. [PubMed: 11086582]
- Rocca P, Fonzo V, Ravizza L, Rocca G, Scotta M, Zanalda E, Bogetto F. A comparison of paroxetine and amisulpride in the treatment of dysthymic disorder. *J. Affect. Disord* 2002;70(3):313–317. [PubMed: 12128243]
- Rush AJ, Bernstein IH, Trivedi MH, Carmody TJ, Wisniewski SM, Munt J, C Shores-Wilson K, Biggs MM, Nierenberg A, Fava M. An evaluation of the Quick Inventory of Depressive Symptomatology and the Hamilton Rating Scale for Depression: a STAR*D report. *Biol. Psychiatry* 2005a Epub ahead of print, doi:10.1016/j.biopsych.2005.08.022
- Rush AJ, Crismon ML, Kashner TM, Toprac MG, Carmody TJ, Trivedi MH, Suppes T, Miller AL, Biggs MM, Shores-Wilson K, Witte BP, Shon SP, Rago WV, Altshuler KZ. Texas Medication Algorithm Project, phase 3 (TMAP-3): rationale and study design. *J. Clin. Psychiatry* 2003;64(4):357–369. [PubMed: 12716235]
- Rush, AJ.; Pincus, HA.; First, MB.; Blacker, D.; Endicott, J.; Keith, SJ.; Phillips, KA.; Ryan, ND.; Smith, GR.; Tsuang, MT.; Widiger, TA.; Zarin, DA. *Handbook of Psychiatric Measures*. American Psychiatric Association; Washington, DC: 2000.
- Rush AJ, Sackeim HA, Marangell LB, George MS, Brannan SK, Davis SM, Lavori P, Howland R, Kling MA, Rittberg B, Carpenter L, Ninan P, Moreno F, Schwartz T, Conway C, Burke M, Barry JJ. Effects of 12 months of vagus nerve stimulation in treatment-resistant depression: a naturalistic study. *Biol. Psychiatry* 2005b;58(5):355–363. [PubMed: 16139581]

- Sackeim HA. The definition and meaning of treatment-resistant depression. *J. Clin. Psychiatry* 2001;62 (Suppl 16):10–17. [PubMed: 11480879]
- Sackeim HA, Prudic J, Devanand DP, Decina P, Kerr B, Malitz S. The impact of medication resistance and continuation pharmacotherapy on relapse following response to electroconvulsive therapy in major depression. *J. Clin. Psychopharmacol* 1990;10:96–104. [PubMed: 2341598]
- Sackeim HA, Prudic J, Devanand DP, Nobler MS, Lisanby SH, Peyser S, Fitzsimons L, Moody BJ, Clark J. A prospective, randomized, double-blind comparison of bilateral and right unilateral electroconvulsive therapy at different stimulus intensities. *Arch. Gen. Psychiatry* 2000;57:425–434. [PubMed: 10807482]
- Samejima, F. Graded response model. In: van Linden, W.; Hambleton, RK., editors. *Handbook of Modern Item Response Theory*. Springer-Verlag; New York: 1997. p. 85-100.
- Santor DA, Coyne JC. Examining symptom expression as a function of symptom severity: item performance on the Hamilton Rating Scale for Depression. *Psychol. Assess* 2001;13(1):127–139. [PubMed: 11281034]
- Senra C. Evaluation and monitoring of symptom severity and change in depressed outpatients. *J. Clin. Psychol* 1996;52(3):317–324. [PubMed: 8835694]
- Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatry* 1998;59 (Suppl 20):22–33. [PubMed: 9881538]
- Trivedi MH, Rush AJ, Crismon ML, Kashner TM, Toprac MG, Carmody TJ, Key T, Biggs MM, Shores-Wilson K, Witte B, Suppes T, Miller AL, Altshuler KZ, Shon SP. Clinical Results for Patients With Major Depressive Disorder in the Texas Medication Algorithm Project. *Arch. Gen. Psychiatry* 2004;61(7):669–680. [PubMed: 15237079]
- Williams JB. A structured interview guide for the Hamilton Depression Rating Scale. *Arch. Gen. Psychiatry* 1988;45(8):742–747. [PubMed: 3395203]
- 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. [PubMed: 15729106]

Figure 1.
Test Information Functions for Study 1 (n=233)

Figure 2.
Test Information Functions for Study 2 (n=985)

Table 1

Item Total Correlations for Each Measure in Each Study

Item	HRSD ₁₇		MADRS		HRSD ₆	
	(Study 1)	(Study 2)	(Study 1)	(Study 2)	(Study 1)	(Study 2)
1.	0.79	0.85	0.87	0.88	0.86	0.88
2.	0.50	0.66	0.89	0.90	0.62	0.72
3.	0.61	0.57	0.65	0.77	0.81	0.87
4.	0.51	0.56	0.47	0.68	0.52	0.60
5.	0.48	0.58	0.54	0.44	0.64	0.77
6.	0.42	0.55	0.68	0.79	0.67	0.75
7.	0.75	0.82	0.78	0.84		
8.	0.41	0.52	0.83	0.87		
9.	0.24	0.45	0.76	0.77		
10.	0.58	0.73	0.73	0.66		
11.	0.50	0.64				
12.	0.51	0.41				
13.	0.64	0.72				
14.	0.46	0.53				
15.	0.27	0.48				
16.	0.05	0.06				
17.	0.24	0.17				
Median	0.50	0.56	0.75	0.78	0.66	0.76

For the HRSD₁₇, the items are: 1. depressed mood; 2. guilt; 3. suicide; 4. initial insomnia; 5. middle insomnia; 6. delayed insomnia; 7. work and activities; 8. retardation; 9. agitation; 10. psychic anxiety; 11. somatic anxiety; 12. reduced appetite GI symptoms; 13. somatic symptoms general; 14. libido/genital symptoms; 15. hypochondriasis; 16. loss of insight; 17. weight loss. For the MADRS, the items are: 1. apparent sadness; 2. reported sadness; 3. inner tension; 4. reduced sleep; 5. reduced appetite; 6. concentration difficulties; 7. lassitude; 8. inability to feel; 9. pessimistic thoughts; 10. suicidal thoughts. For the HRSD₆, the items are: 1. depressed mood; 2. guilt; 3. work and activities; 4. retardation; 5. psychic anxiety; and 6. somatic symptoms general.

Table 2
Item Response Analyses for the HRSD₁₇ for Study 1 (n=233) and Study 2 (n=985)

HRSD ₁₇ Item	Study	a	b ₀	b ₁	b ₂	b ₃
1. depressed mood	1	2.89	-1.36	-0.32	0.37	1.24
	2	3.61	-0.81	0.18	0.88	2.07
2. guilt	1	1.02	-1.07	0.36	3.17	19.53
	2	1.75	-0.39	0.84	2.88	4.71
3. suicide	1	1.43	-0.09	0.84	2.48	21.38
	2	2.00	1.00	1.84	3.00	6.47
4. initial insomnia	1	0.95	0.24	1.10		
	2	1.20	0.29	1.30		
5. middle insomnia	1	0.80	-0.18	1.38		
	2	1.21	-0.15	1.15		
6. delayed insomnia	1	0.65	0.98	2.56		
	2	1.18	0.26	1.69		
7. work and activities	1	2.44	-1.56	-0.54	0.35	1.95
	2	2.93	-0.78	0.18	0.85	2.14
8. retardation	1	0.87	0.59	2.23	5.39	24.61
	2	1.28	0.42	2.31	5.40	5.95
9. agitation *	1	0.24	4.03	14.14	63.30	*
	2	0.84	0.02	2.49	4.83	6.91
10. psychic anxiety	1	1.04	-1.55	0.51	1.87	4.64
	2	2.04	-1.04	0.24	1.59	3.45
11. somatic anxiety	1	0.91	-2.38	0.03	2.87	5.63
	2	1.42	-0.40	1.04	3.26	5.04
12. appetite	1	1.11	0.70	1.75		
	2	0.96	1.49	4.15		
13. somatic symptoms general	1	1.90	-1.45	-0.08		
	2	2.13	-0.61	0.48		
14. libido	1	0.85	-1.69	-0.32		
	2	1.10	-0.24	1.14		
15. hypochondriasis	1	0.44	2.56	5.44	9.41	12.59
	2	1.08	0.79	2.58	5.04	20.55
16. loss of insight	1	0.26	15.93	21.47		
	2	0.38	5.32	10.74		
17. weight loss	1	0.60	3.52	6.31		
	2	0.22	11.19	22.37		

* insufficient data to estimate this parameter

Table 3
Item Response Analyses for the MADRS for Study 1 (n=233) and Study 2 (n=985)

MADRS Item	Study	a	b ₀	b ₁	b ₂	b ₃	b ₄	b ₅
1. apparent sadness	1	3.99	-1.33	-0.76	-0.04	0.46	1.22	2.65
	2	4.26	-0.83	-0.23	0.45	1.01	1.94	3.09
2. reported sadness	1	4.83	-1.29	-0.86	-0.26	0.30	1.00	1.90
	2	5.24	-0.81	-0.35	0.29	0.86	1.89	2.80
3. inner tension	1	1.41	-1.63	-1.09	0.08	1.40	3.21	8.62
	2	2.26	-1.21	-0.67	0.49	1.71	3.22	5.92
4. reduced sleep	1	0.68	-0.80	-0.41	0.99	1.94	4.05	7.27
	2	1.43	-0.90	-0.49	0.40	1.16	2.61	4.04
5. reduced appetite	1	1.10	0.71	0.84	1.45	2.54	3.57	10.08
	2	0.83	1.27	1.69	3.20	4.79	6.89	8.57
6. concentration difficulties	1	1.55	-1.84	-1.03	-0.15	0.71	2.37	3.74
	2	2.31	-0.82	-0.37	0.54	1.21	2.86	5.85
7. lassitude	1	2.26	-1.63	-0.97	-0.11	0.30	1.72	5.44
	2	2.90	-0.86	-0.30	0.38	0.88	2.50	5.76
8. inability to feel	1	2.82	-1.24	-0.81	-0.05	0.59	1.69	2.38
	2	3.40	-0.68	-0.18	0.50	1.02	2.09	2.93
9. pessimistic thoughts	1	2.29	-1.35	-0.74	-0.08	0.77	2.06	5.24
	2	2.31	-0.74	-0.16	0.97	1.78	6.29	5.24
10. suicidal thoughts	1	2.04	-0.52	0.30	1.01	1.62	2.44	6.18
	2	1.94	0.26	1.33	2.04	2.66	6.74	6.18

* insufficient data to estimate this parameter

Table 4Item Response Analyses for the HRSD₆ for Study 1 (n=233) and Study 2 (n=985)

HRSD ₆ Item	Study	a	b ₀	b ₁	b ₂	b ₃
1. depressed mood	1	3.13	-1.32	-0.28	0.39	1.22
	2	3.59	-0.79	0.20	0.90	2.09
2. guilt	1	1.09	-1.01	0.36	3.02	24.69
	2	1.77	-0.37	0.85	2.87	4.66
3. work and activities	1	2.42	-1.56	-0.54	0.35	1.94
	2	3.16	-0.75	0.19	0.85	2.12
4. retardation	1	1.01	0.53	1.99	4.73	26.03
	2	1.41	0.42	2.19	5.05	5.55
5. psychic anxiety	1	1.02	-1.56	0.53	1.90	4.69
	2	1.99	-1.04	0.26	1.63	3.50
6. somatic symptoms general	1	1.82	-1.47	-0.08		
	2	2.08	-0.60	0.50		

Table 5IRT Conversion Table for the HRSD₁₇, MADRS, and HRSD₆ (Studies 1 and 2 combined) (n=1218)

Conversion HRSD ₁₇ - MADRS		Conversion HRSD ₁₇ - HRSD ₆		Conversion HRSD ₆ - MADRS	
0 or 1	0	0 or 1	0	0	0 or 1
2	1	2 or 3	1	1	2 or 3
3	2 or 3	4 or 5	2	2	4 to 6
4	4 or 5	6	3	3	7 or 8
5	6	7 or 8	4	4	9 or 10
6	7	9 or 10	5	5	11 to 13
7	8 or 9	11	6	6	14 or 15
8	10	12 or 13	7	7	16 to 18
9	11 or 12	14 or 15	8	8	19 or 20
10	13	16	9	9	21 or 22
11	14 or 15	17 or 18	10	10	23 to 25
12	16	19 or 20	11	11	26 or 27
13	17 or 18	21 or 22	12	12	28 to 30
14	19	23 or 24	13	13	31 to 33
15	20	25 or 26	14	14	34 or 35
16	21 or 22	27 or 28	15	15	36 to 38
17	23	29 to 31	16	16	39 to 41
18	24 or 25	32 or 33	17	17	42 or 43
19	26	34 or 35	18	18	44 to 46
20	27	36 or 37	19	19	47 to 49
21	28 or 29	38 or 39	20	20	50 or 51
22	30	40	21	21	52
23	31	41 to 52	22	22	53 to 60
24	32 or 33				
25	34				
26	35				
27	36				
28	37				
29	38 or 39				
30	40				
31	41				
32	42 or 43				
33	44				
34	45				
35	46 or 47				
36	48				
37	49				
38	50				
39	51				
40	52				
41	53				
42 or 43	54				
44	55				
45 to 47	56				
48 to 50	57				
51 or 52	58 to 60				

Table 6

Effect Sizes for Study 1 (n=232) and Study 2 (n=985)

HRSD ₁₇		MADRS		HRSD ₆	
Item	Study 1	Study 2	Item	Study 1	Study 2
1. depressed mood	0.81 (1)*	1.23 (1)	1. apparent sadness	0.76 (3)	1.10 (2)
2. guilt	0.58 (3)	0.98 (3)	2. reported sadness	0.82 (2)	1.16 (1)
3. suicide	0.44 (7)	0.60 (12)	3. inner tension	0.46 (8)	0.80 (7)
4. initial insomnia	0.33 (8)	0.69 (8)	4. reduced sleep	0.32 (9)	0.78 (8)
5. middle insomnia	0.29 (9)	0.66 (10)	5. reduced appetite	0.28 (10)	0.51 (10)
6. delayed insomnia	0.16 (15)	0.55 (13)	6. concentration difficulties	0.66 (5)	0.85 (6)
7. work and activities	0.66 (2)	1.19 (2)	7. lassitude	0.73 (4)	0.92 (4)
8. retardation	0.54 (5)	0.75 (6)	8. inability to feel	0.84 (1)	1.04 (3)
9. agitation	0.29 (10)	0.66 (9)	9. pessimistic thoughts	0.55 (6)	0.87 (5)
10. psychic anxiety	0.50 (6)	0.93 (4)	10. suicidal thoughts	0.52 (7)	0.64 (9)
11. somatic anxiety	0.24 (12)	0.72 (7)	Total Score	0.90	1.26
12. appetite	0.24 (13)	0.50 (15)			
13. somatic symptoms	0.57 (4)	0.87 (5)			
general					
14. libido	0.25 (11)	0.54 (14)			
15. hypochondriasis	0.17 (14)	0.64 (11)			
16. loss of insight	0.12 (16)	0.25 (16)			
17. weight loss	0.06 (17)	0.19 (17)			
Total Score	0.87	1.49			

* Number in parenthesis indicates the rank order of the effect size from largest to smallest