



Published in final edited form as:

J Psychiatr Res. 2015 February ; 61: 40–45. doi:10.1016/j.jpsychires.2014.12.015.

Rating Depression Over Brief Time Intervals with the Hamilton Depression Rating Scale: Standard vs. Abbreviated Scales

David A Luckenbaugh^a, Rezvan Ameli^a, Nancy E Brutsche^a, and Carlos A Zarate Jr^a

David A Luckenbaugh: dave.luckenbaugh@nih.gov; Rezvan Ameli: rezvan.ameli@nih.gov; Nancy E Brutsche: nancy.brutsche@nih.gov; Carlos A Zarate: carlos.zarate@nih.gov

^aExperimental Therapeutics and Pathophysiology Branch, Intramural Research Program, National Institute of Mental Health, National Institutes of Health, and Department of Health and Human Services, Bethesda, Maryland, USA

Abstract

Although antidepressant trials typically use weekly ratings to examine changes in symptoms over six to 12 weeks, antidepressant treatments may improve symptoms more quickly. Thus, rating scales must be adapted to capture changes over shorter intervals. We examined the use of the 17-item Hamilton Depression Rating Scale (HDRS) to evaluate more rapid changes. Data were examined from 58 patients with major depressive disorder or bipolar disorder enrolled in double-blind, placebo-controlled, crossover studies who received a single infusion of ketamine (0.5 mg/kg) or placebo over 40 minutes then crossed over to the other condition. HDRS subscales, a single HDRS Depressed mood item, and a visual analog scale were used at baseline, after a brief interval (230 minutes), and one week post-infusion. Effect sizes for the ketamine-placebo difference were moderate ($d > 0.50$), but one and two-item HDRS subscales had the smallest effects. Response rates on active drug were lowest for the complete HDRS (43%); the remaining scales had higher response rates to active drug, but the shortest subscales had higher response rates to placebo. Correlations between the changes from baseline to 230 minutes post-ketamine across scores were similar for most subscales ($r = .82-.97$), but correlations using the single items were lower ($r < .74$). Overall, effect sizes for drug-placebo differences and correlations between changes were lower for one- and two-item measures. Response rates were lower with the full HDRS scale. The data suggest that, to best identify rapid antidepressant effects, a scale should have more than two items, but fewer items than a full scale.

Keywords

antidepressant; bipolar disorder; depression; HDRS; ketamine; ratings

Corresponding Author: Carlos A. Zarate Jr. M.D., National Institutes of Health/National Institute of Mental Health, Experimental Therapeutics & Pathophysiology Branch, Building 10/Clinical Research Center (CRC), 10 Center Dr., Room 7-5342, Bethesda, MD 20892, Phone: (301)-451-0861, Fax: (301)-480-8792, zaratec@mail.nih.gov.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

For decades, clinical trials in depression—major depressive disorder (MDD) and bipolar disorder (BD)—have typically used weekly ratings to examine changes after treatment. Some recent experimental treatments, however, appear to exert antidepressant effects within hours or days, underscoring the need for rating instruments capable of detecting much more rapid antidepressant effects (Machado-Vieira et al., 2008). Current common scales used to assess depressive symptoms, such as the Hamilton Depression Rating Scale (HDRS) (Hamilton, 1960) and the Montgomery-Asberg Depression Rating Scale (MADRS) (Montgomery and Asberg, 1979) were designed to obtain measurements at weekly intervals. In addition, these scales measure certain symptoms that cannot be evaluated over a short time frame (e.g., changes in sleep or weight).

Various research groups have suggested different ways to use the original HDRS items to assess depressive symptoms. In a series of clinical trials, Santen and colleagues (2008) examined the individual items of the HDRS to determine those that were most sensitive to changes in depressive symptoms. Santor and colleagues (2008) took an item response theory approach in an attempt to differentiate patients with higher and lower degrees of symptom severity.

Because these approaches led to different subscales of the HDRS, additional studies examined how well various subscales performed compared to total HDRS score. While some studies showed that shorter subscales improved the rate of response to the outcome measure (e.g. Bech et al., 2010; Faries et al., 2000; Mallinckrodt et al., 2011; Revicki et al., 2010; Santen et al., 2009; Silverstone et al., 2002), others found no noticeable difference (e.g. Ballesteros et al., 2007; McIntyre et al., 2005; Revicki et al., 2010; Ruhe et al., 2005). Boessen and colleagues (2013) pointed out that some of the differences across studies were likely due to the type of studies used to evaluate the scales. For instance, they noted that total HDRS score appeared to work better when evaluating tricyclic antidepressants (TCAs), but HDRS subscales often worked better with selective serotonin reuptake inhibitors (SSRIs). Santen et al. (2009) found that a specific subscale of the HDRS worked best for both of these drug classes.

Within the scientific community, considerable interest exists in developing or adapting instruments capable of evaluating improvement over much shorter periods of time. Several groups have pointed out that antidepressants appear to improve symptoms long before the commonly reported value of three to six weeks (e.g. Papakostas et al., 2006; Posternak et al., 2005; Stassen et al., 1998) and, notably, some current trials examine symptoms within the course of a single day (e.g. Zarate et al., 2006). Because some depression severity scales include items that cannot change over short intervals (e.g., insomnia), the need exists to understand which scales, if any, will be sensitive to changes within these very brief time frames but nevertheless remain relevant over longer periods.

The present study used the HDRS to illustrate change in depressive symptoms over the course of a brief treatment trial with the N-methyl-D-aspartate (NMDA) antagonist ketamine. This scale was chosen as representative of typical clinical trial scales because 1) it

is widely used; 2) it was the primary outcome measure for our initial ketamine study; and 3) we had the most available data with it. The standard HDRS scale and HDRS subscales were used to determine better approaches for handling data in studies involving rapid changes in depressive symptoms. The analysis examined depressive symptoms assessed at baseline with changes assessed at 230 minutes post-ketamine infusion to examine extremely short time frames; the symptoms were also assessed at seven days post-ketamine infusion in order to reflect a more standard antidepressant time point.

Material and Methods

Fifty-eight treatment resistant inpatients with either BD (n=36) or MDD (n=22)—as assessed via the Structured Clinical Interview for the DSM IV-R, clinical interviews, and patient history—were recruited to participate in double-blind, placebo-controlled, crossover studies of ketamine to reduce depressive symptomatology. All studies were approved by the Combined CNS IRB at the NIH. Patients provided informed consent prior to participation. The methodology and results of these studies have been published elsewhere (DiazGranados et al., 2010; Zarate et al., 2006, 2012), but additional patients recruited in the process of generating those manuscripts are included here. Briefly, patients were randomized to receive a single infusion of either placebo or 0.5 mg/kg of ketamine hydrochloride and then crossed over to the other condition after a week for MDD patients and two weeks for BD patients. MDD patients were medication free for at least two weeks prior to the study and BD patients were on stable doses of either lithium or valproic acid. Trained clinicians rated patients on the HDRS, a visual analog scale (VAS) for “depressed, sad, blue”, the MADRS, the Beck Depression Inventory (BDI) (Beck & Beamesderfer, 1974), and several other psychiatric rating scales at 60 minutes prior to infusion, then at 40, 80, 120, and 230 minutes post-infusion, and finally on days 1, 2, 3, and 7 post-infusion. Ratings were made relative to the most recent time period assessed.

The 17-item HDRS was examined as a whole and in several subscales designed to evaluate depressive symptoms in a clinical trial. Outcome measures included: 1) the total HDRS; 2) seven subscales of the HDRS drawn from the extant literature (see Table 1) (Bech et al., 1981; Evans et al., 2004; Gibbons et al., 1993; Maier et al., 1985; McIntyre et al., 2002; Santen et al., 2008; Silverstone et al., 2002); 3) a shortened version of the HDRS that eliminated items that would not change over brief time intervals (e.g., early, middle, and late insomnia and weight change) (Leibenluft et al., 1993); 4) the Depressed mood item from the HDRS, and 5) the VAS rating. Table 1 shows the items used for each of the subscales. The insight item was not included in the analyses because it was a constant zero throughout the dataset presented here; this is likely due to patients accepting that they have an illness in order to participate in research.

Statistics

To understand the coherence (internal consistency reliability) of the full scale HDRS as well as that of its subscales, Cronbach’s alpha was calculated with data from the 230-minute time point following ketamine infusion because that was the primary point of interest. Baseline measures were not used due to the restricted variance likely for a clinical trial where

participants would have to meet a certain severity criterion for entry into the study. Once the study began, subjects could have any score within the range of the scale in question, which would provide a more realistic assessment of the relationships among scale items.

To understand the degree of overlap among scales (criterion related validity), Pearson correlations were used to examine the inter-relationships among scores at 230 minutes and percent change in scores on ketamine from baseline to 230 minutes. Similar analyses were conducted substituting the seven-day scores for the 230-minute scores.

Finally, linear mixed models with a compound symmetry covariance structure and restricted maximum likelihood estimation were used to examine differences between ketamine and placebo over the course of the first week of the ketamine crossover trials which should provide a sense of the sensitivity to change, or responsiveness, of the scales. A factorial model with time and drug effects was used with the drug-specific baseline as covariate. *A priori* post-hoc comparisons were performed comparing the drug and placebo phases at 230 minutes as the primary comparison for understanding the degree of rapid improvement in depressive symptoms. Similar comparisons are reported at seven days. Significance was evaluated at $p < .05$, two-tailed. Based on the *a priori* comparisons, the effect size for the difference from the corresponding placebo point (Cohen's d) was calculated to standardize the degree of improvement. The formula was $d = 2t / \sqrt{df}$ (Rosenthal and Rosnow, 1991). Cohen's d is the number of standard deviations of change for the scale in question (Cohen, 1988). Greater differences are shown as larger effect sizes. The standard interpretation for this effect size is as follows: 0.2 is small, 0.5 is moderate, and 0.8 is large.

To evaluate a more clinically relevant outcome, response rates were calculated for each measure using 50% change from baseline as the criterion. McNemar tests were used to compare the response rates to drug and placebo for subjects who completed both phases for the time point in question. Further, the proportion of agreement between responders and non-responders was examined for each measure in relation to total HDRS and Leibenluft subscale. IBM SPSS Statistics for Windows, Version 21.0 (IBM Corp., Armonk, N.Y.) was used for all analyses.

Results

The inpatients included in this analysis had an average total HDRS score of 22.2 (SD=4.2) on the 17-item HDRS at study baseline. Their average age was 47.0 (SD=11.0), 60% (35/58) were female, and they had been ill on average 27.5 (SD=11.3) years.

Reliability: Internal Consistency

Cronbach's alpha was 0.75 for total HDRS at 230 minutes (Figure 1). The subscales had values in the range of 0.64 to 0.81. The two-item scale from Silverstone and colleagues (2002) had the lowest value, and the seven-item scale from Santen and colleagues (2008) had the highest value. These values suggest reasonably consistent relationships among the items included in each subscale. Given that scales with a larger number of items tend to have greater coherence in general, the smaller value associated with the shortest scale is not surprising.

Criterion-Related Validity

Correlations for raw scores with all items and subscales with the total HDRS at 230 minutes (Figure 2A) were high ($r = 0.73$); one smaller relationship was seen with the HDRS Depressed mood item ($r=0.60$) (Figure 2). The other two smallest of these relationships were with the VAS ($r=0.73$) and Silverstone subscales ($r=0.76$). Subscales with six or more items were all more highly associated with each other ($r = .94$). The shorter scales—those with just one or two items—did not seem to overlap with the longer ones as much, suggesting that they may have missed something captured by the other scales.

Percent change in total HDRS score was highly related to percent change in all of the other measures (230 Minutes: $r=0.72$ to 0.97 ; seven days: $r=0.73$ to 0.94) (Figure 2B). However, for total HDRS, the smallest relationships were with the scales with the fewest items: VAS (230 Minutes: $r=0.72$; seven days: $r=0.73$), HDRS mood item (230 Minutes: $r=0.73$; seven days: $r=0.75$), and the Silverstone scale (230 Minutes: $r=0.82$; 7 Days: $r=0.78$). This suggests that these very brief measures could be missing some of the variance in change captured by the longer scales. The VAS had the smallest correlations with HDRS subscales (230 Minutes: $r=0.73$ to 0.80 ; seven days: $r=0.69$ to 0.82). These relationships were somewhat higher with the HDRS Depressed mood item (230 Minutes: $r=0.75$ to 0.93 ; seven days: $r=0.80$ to 0.91) and the Silverstone subscale (230 Minutes: $r=0.83$ to 0.89 ; seven days: $r=0.85$ to 0.90). All of the other scales were more highly related, with correlations ranging from 0.94 and 1.00 at 230 minutes and 0.91 to 0.99 at seven days. Given the substantial overlap in items across scales, such high correlations would be expected. However, along with the similar overlap with total HDRS score, the relationships suggest that any of the longer subscales may offer a comparable evaluation of the depression construct as ratings change over time. Results were similar at 230 minutes and seven days, suggesting the stability of these findings at various time points, specifically a very brief interval and a longer, more common interval for clinical trials of depression.

Sensitivity to Change

Linear mixed models for all of the scales evaluated showed significant differences between drug and placebo at 230 minutes and at seven days (Figure 3A). At 230 minutes, the effect size for the total HDRS was moderate ($d=0.55$) and the effect sizes for the subscales were moderate as well. The d values fit into a tight range between 0.52 and 0.64 ; Silverstone and colleagues' (2002) scale had the smallest value. Thus, each of the measures was able to detect drug differences in a brief time interval. The VAS item and the HDRS Depressed mood item had effect sizes similar to the subscales. When examining seven-day data, the one- and two-item HDRS subscales had smaller effect sizes ($ds=0.16$) than the other scales. The longer subscales had small to moderate effects ($d=0.24$ to 0.27) that were all higher in number to the total HDRS ($d=0.23$) even if similar in magnitude.

Next, response rates to placebo and to ketamine at 230 minutes (Figure 3B) were examined using a McNemar test for each of the scales. Response rates were significantly higher on ketamine than on placebo for all of the measures ($p<.001$). Response rates on placebo were 4% to 7% on all of the scales except that of Silverstone and colleagues (2002), which had a 17% response rate. Response rates on ketamine ranged from 59% and 67% for all of the

scales except the two longest; total HDRS showed a 48% response rate—which was the lowest of all the measures—and the Leibenluft subscale showed a 56% response rate. These lower rates could be due to the inclusion of items such as sleep measures that would not be able to change over short intervals. Because some items could not change except by error, the increased stability in the standard measures would make achieving response criteria more difficult.

When the seven-day time point was examined, response rate to ketamine for the full HDRS (16%) was slightly lower than that of the longer subscales (18 to 30%). The one- and two-item measures showed similar response rates to active treatment as the other scales (Silverstone 22%; HDRS Depressed mood item 30%), but slightly higher response to placebo (Silverstone 5%; HDRS Depressed mood item 7%) that was not detected by the longer scales. This suggests that the shorter scales may have been somewhat less stable than the other measures. The difference between active and inactive drug was generally smaller with one- and two-item scales.

Because response rates alone do not show whether the same patients are categorized as responders for various measures, measures of agreement were examined with the active drug (Figure 4). Of those called non-responders by the total HDRS, the VAS and the two brief HDRS measures agreed between 57% and 67% of the time. In contrast, the longer subscales agreed between 71% and 86% of the time. For those called responders by the total HDRS, scores on the VAS concurred 77% of the time, while the HDRS Depressed mood item and the Silverstone subscale agreed 88% and 96% of the time, respectively. The subscales agreed on responder categorizations 100% of the time. This suggests that these brief measures did not categorize patients the same way as the total HDRS, especially when patients were called non-responders. Because this could be due to time frame, the seven-day data were used to determine whether categorizations differed in a time interval more appropriate for the total HDRS. The data indicated that the VAS and HDRS Depressed mood item continued to have lower agreement (83% and 81%, respectively) than the rest of the measures (>90%) with the exception of the Maier subscale. This suggests that single items may not accurately reflect the type of information gained from a larger subscale. The longest scales may miss a certain proportion of responders. If the Leibenluft subscale was used as the reference, agreement was higher, but the shorter time frame had less agreement than the longer one. Categorizations at the shorter time frame depended much more on the scale used than those made at a week.

Discussion

This study examined the ability of a standard depression scale and its subscales to detect changes in depressive symptoms over a very brief time frame. We found that total HDRS score appeared to be strongly related to both longer and shorter measures whether examining raw values at an individual time point or changes over an interval. Effect sizes indicating the degree of difference between drug and placebo were also comparable among subscales and individual items. The place where total HDRS score differed from other scales was primarily in the assessment of response rates. Response rates for HDRS total score were the smallest of all the scales, including those for single item measures. This is most likely due to

the fact that a few items are included in the total score that cannot change measurably over the brief time frame. With relatively unchangeable elements, the proportion of change is lower. This is an important point, as it suggests limitations associated with HDRS total score.

If researchers hope to examine clinically important improvements in depressive symptoms, total HDRS score may underestimate the proportion of patients who improve in response to a particular drug. In the present study, every scale showed a higher proportion of responders to ketamine than the total score. When items were removed that could not change over a very brief interval, response rate still remained the lowest of any of the subscales. This was true regardless of whether the 230-minute or seven-day interval was examined.

On the other hand, the one- and two-item scales showed higher placebo response rates than longer HDRS subscales. This could be due to the fact that the individual items detect more non-specific change in general. Or, they may just be more variable, so the higher placebo response rates in this study could be more random error. Because response rates to active drug were similar with subscales of various lengths, the higher placebo response rates with the shortest scales means the longer scales may be more sensitive to differences between drugs. This suggests that researchers would be more likely to detect drug differences using moderately sized subscales.

Further, the overall pattern of results suggests that extremely short one- or two-item scales may not capture the full extent of the variation that longer scales do. This is underscored by the relatively lower correlations of these scales with longer scales, regardless of whether we examined a static point or changes over time. One could argue that the higher correlations among longer subscales are due to the pure number of items overlapping in those scales. However, the effect sizes in the shorter scales tended to be on the lower end as well, which supports the notion that they may miss something important captured by the more extensive scales.

Results for moderately sized subscales were comparable in all of the analyses performed. Internal consistency, correlations of raw scores and changes, effect sizes indicating drug-placebo differences, and response rates were all similar. The subscale from Maier and colleagues (1985), however, showed somewhat less sensitivity to response relative to the total score than the other subscales. In addition, effect sizes were on the lower end of the longer subscales for both time points examined. This subscale is one of only two that includes the HDRS agitation item, whereas most of the other scales include the general somatic symptoms item. Thus, the other subscales may perform better. With the rest of the subscales, there is no clear basis to recommend one over another. Researchers could use any of the moderately sized subscales of the HDRS presented here to test changes in depressive symptoms. The use of one over the others likely depends on the expectations of what symptoms should change within a given clinical trial. Given the effects demonstrated, most abbreviated scales with more than two items appear sensitive to changes in severity of depressive symptoms as well as response rates at very brief intervals.

Several important limitations exist in the present analysis. First, the sample size of 58 was very small for determining the true relationship between measures and the true size of effects detected by specific scales. Second, the subjects in the sample were a highly treatment refractory group who may respond differently than other groups of patients. Third, the analysis examined data for only one specific drug—ketamine, an NMDA antagonist—whose effects may differ from those of other drugs. Indeed, Boessen and colleagues (2013) suggested that the HDRS may be more useful for evaluating changes in some classes of drugs than others (i.e. TCAs over SSRIs), though Santen, et al. (2009) found similarities with a TCA and an SSRI. Nelson, et al., (2006) found similarities in the symptoms affected by serotonergic and noradrenergic drugs. Finally, only two pre-chosen time frames were examined. While the results here may generalize to other intervals, the present study examined data only at 230 minutes and seven days. Results could differ at other time intervals.

Despite these limitations, the present results illustrate the ability of a standard depression scale and a variety of its subscales to detect changes over a very brief time frame. Future studies should examine changes over short time intervals with other drugs at different intervals to determine whether the present results generalize to a broader context. Research should include multiple intervals to determine whether the value of scales differs depending on the length of the study. Further, evaluation of individual items could help determine the most critical symptoms to examine in various types of studies. Future studies should also evaluate additional depression rating scales—such as the MADRS—to determine their ability to detect changes in depressive symptoms. Results across scales should be compared to determine an optimal scale or set of scales for clinical trials studying antidepressant effects over brief time intervals. Future work in this area should also attempt to include larger sample sizes to decrease the confidence intervals of the effect sizes and allow direct statistical comparisons of the changes on various subscales.

Acknowledgements

The authors thank Ioline Henter for her careful eye for detail in editing this manuscript. In addition, the authors thank the staff of the Mood and Anxiety Disorders Research Unit at the NIH Clinical Center for their extensive work in collecting the data presented here. Data was collected under protocol 04-M-0220 and registered under NCT#00088699.

REFERENCES

- Ballesteros J, Bobes J, Bulbena A, Luque A, Dal-Ré R, Ibarra N, et al. Sensitivity to change, discriminative performance, and cutoff criteria to define remission for embedded short scales of the Hamilton depression rating scale (HAMD). *J Affect Disord.* 2007; 102:93–99. [PubMed: 17258323]
- Bech P, Allerup P, Gram LF, Reisby N, Rosenberg R, Jacobsen O. The Hamilton depression scale: Evaluation of objectivity using logistic models. *Acta Psychiatr Scand.* 1981; 63:290–299. [PubMed: 7015793]
- Bech P, Boyer P, Germain JM, Padmanabhan K, Haudiquet V, Pitrosky B, et al. HAM-D17 and HAM-D6 sensitivity to change in relation to desvenlafaxine dose and baseline depression severity in major depressive disorder. *Pharmacopsychiatry.* 2010; 7:271–276. [PubMed: 20830664]
- Beck AT, Beamesderfer A. Assessment of depression: The depression inventory. *Mod Prob Pharmacopsychiatry.* 1974; 7:151–169.

- Boessen R, Groenwold RHH, Knol MJ, Grobbee DE. Comparing HAMD₁₇ and HAMD subscales on their ability to differentiate active treatment from placebo in randomized controlled trials. *J Affect Disord.* 2013; 145:363–369. [PubMed: 22959683]
- Cohen, J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Hillsdale, NJ: Erlbaum; 1988.
- DiazGranados N, Ibrahim L, Brutsche NE, Newberg A, Kronstein P, Khalife S, et al. A randomized add-on trial of an N-methyl-D-aspartate antagonist in treatment-resistant bipolar depression. *Arch Gen Psychiatry.* 2010; 67:793–802. [PubMed: 20679587]
- Entsuah R, Shaffer M, Zhang J. A critical examination of the sensitivity of unidimensional subscales derived from the Hamilton Depression Rating Scale to antidepressant drug effects. *J Psychiatr Res.* 2002; 36:437–448. [PubMed: 12393314]
- Evans KR, Sills T, DeBrotta DJ, Gelwicks S, Engelhart N, Santor D. An item response analysis of the Hamilton Depression Rating Scale using shared data from two pharmaceutical companies. *J Psychiatr Res.* 2004; 38:275–284. [PubMed: 15003433]
- Faries D, Herrera J, Rayamajhi J, DeBrotta D, Demitrack M, Potter WZ. The responsiveness of the Hamilton depression rating scale. *J Psychiatr Res.* 2000; 34:3–10. [PubMed: 10696827]
- Gibbons RD, Clark DC, Kupfer DJ. Exactly what does the Hamilton Depression Rating Scale measure? *J Psychiatr Res.* 1993; 27:259–273. [PubMed: 8295158]
- Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry.* 1960; 23:56–62. [PubMed: 14399272]
- Leibenluft E, Moul DE, Schwartz PJ, Madden PA, Wehr TA. A clinical trial of sleep deprivation in combination with antidepressant medication. *Psychiatry Res.* 1993; 46:213–227. [PubMed: 8493292]
- Mallinckrodt CH, Tamura RN, Tanaka Y. Recent developments in improving signal detection and reducing placebo response in psychiatric clinical trials. *J Psychiatr Res.* 2011; 45:1202–1207. [PubMed: 21453932]
- Machado-Vieira R, Salvadore G, Luckenbaugh DA, Manji HK, Zarate CA Jr. Rapid onset of antidepressant action: A new paradigm in the research and treatment of major depressive disorder. *J Clin Psychiatry.* 2008; 69:946–958. [PubMed: 18435563]
- Maier W, Philipp M. Improving the assessment of severity of depressive states: A reduction of the Hamilton depression scale. *Pharmacopsychiatry.* 1985; 18:114–115.
- McIntyre R, Kennedy S, Bagby RM, Bakish D. Assessing full remission. *J Psychiatry Neurosci.* 2002; 27:235–239. [PubMed: 12174732]
- McIntyre RS, Konarski JZ, Mancini DA, Fulton KA, Parikh SV, Grigoriadis S, et al. Measuring the severity of depression and remission in primary care: Validation of the HAMD-7 scale. *CMAJ.* 2005; 173:1327–1331. [PubMed: 16301700]
- Montgomery SA, Asberg MA. A new depression scale designed to be sensitive to change. *Br J Psychiatry.* 1979; 28:611–616.
- Nelson JG, Portera L, Leon AC. Assessment of outcome in depression. *J Psychopharm.* 2006; 20:47–53.
- Papakostas GI, Perlis RH, Scalia MJ, Petersen TJ, Fava M. A meta-analysis of early sustained response rates between antidepressants and placebo for the treatment of major depressive disorder. *J Clin Psychopharmacol.* 2006; 26:56–60. [PubMed: 16415707]
- Posternak MA, Zimmerman M. Is there a delay in the antidepressant effect? A meta-analysis. *J Clin Psychiatry.* 2005; 66:148–158. [PubMed: 15704999]
- Revicki DA, Chen W-H, Frank L, Feltner D, Morlock R. Development and analysis of item response theory-based short-form depression severity scales based on the HDRS and MADRS. *Health Outcomes Research in Medicine.* 2010; 1:e111–e122.
- Rosenthal, R.; Rosnow, R. *Essentials of Behavioral Research: Methods and Data Analysis.* 2nd ed. NY: McGraw-Hill; 1991.
- Ruhé HG, Dekker JJ, Peen J, Holman R, De Jonghe F. Clinical use of the Hamilton Depression Rating Scale: Is increased efficiency possible? A post hoc comparison of Hamilton Depression Rating Scale, Maier and Bech subscales, Clinical Global Impression, and Symptom Checklist-90 scores. *Compr Psychiatry.* 2005; 46:417–427. [PubMed: 16275208]

- Santen G, Danhof M, Pasqua OD. Sensitivity of the Montgomery Asberg Depression Rating Scale to response and its consequences for the assessment of efficacy. *J Psychiatr Res.* 2009; 43:1049–1056. [PubMed: 19344914]
- Santen G, Gomeni R, Danhof M, Pasqua OD. Sensitivity of the individual Hamilton depression rating scale to response and its consequences for the assessment of efficacy. *J Psychiatr Res.* 2008; 42:1000–1009. [PubMed: 18206909]
- Santor DA, Debrota D, Engelhardt N, Gelwicks S. Optimizing the ability of the Hamilton Depression Rating Scale to discriminate across levels of severity and between antidepressants and placebos. *Depress Anxiety.* 2008; 25:774–786. [PubMed: 17935212]
- Silverstone PH, Entsuah R, Hackett D. Two items on the Hamilton Depression rating scale are effective predictors of remission: Comparison of selective serotonin reuptake inhibitors with the combined serotonin/norepinephrine reuptake inhibitor, venlafaxine. *Int Clin Psychopharmacol.* 2002; 17:273–280. [PubMed: 12409680]
- Stassen HH, Angst J. Delayed onset of action antidepressants: Fact or fiction? *CNS Drugs.* 1988; 9:177–184.
- Zarate CA Jr, Brutsche NE, Ibrahim L, Chaves JF, Diazgranados N, Cravchik A, et al. Replication of ketamine's antidepressant efficacy in bipolar depression: A randomized controlled add-on trial. *Biol Psychiatry.* 2012; 71:939–946. [PubMed: 22297150]
- Zarate CA Jr, Singh JB, Carlson P, Brutsche N, Ameli R, Luckenbaugh DA, et al. A randomized trial of an NMDA antagonist in treatment resistant major depression. *Arch Gen Psychiatry.* 2006; 63:856–864. [PubMed: 16894061]

Highlights

1. Total HDRS was strongly related to subscales at one time or with changes over time
2. Effect sizes for drug and placebo difference were comparable for subscales & items
3. Scales with more than two items were sensitive to changes at very brief intervals
4. Response rates were lowest with HDRS total score
5. To best identify rapid effects, use more than two items but less than a full scale

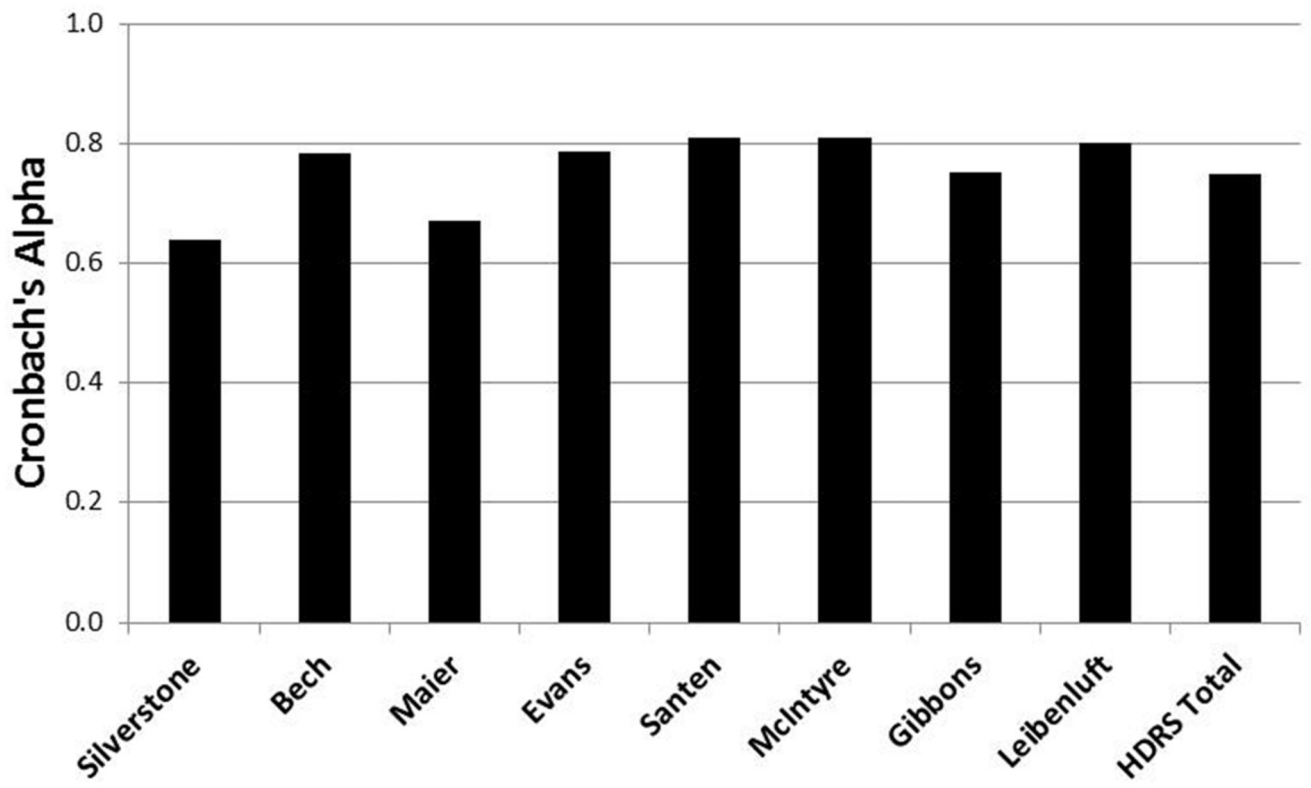
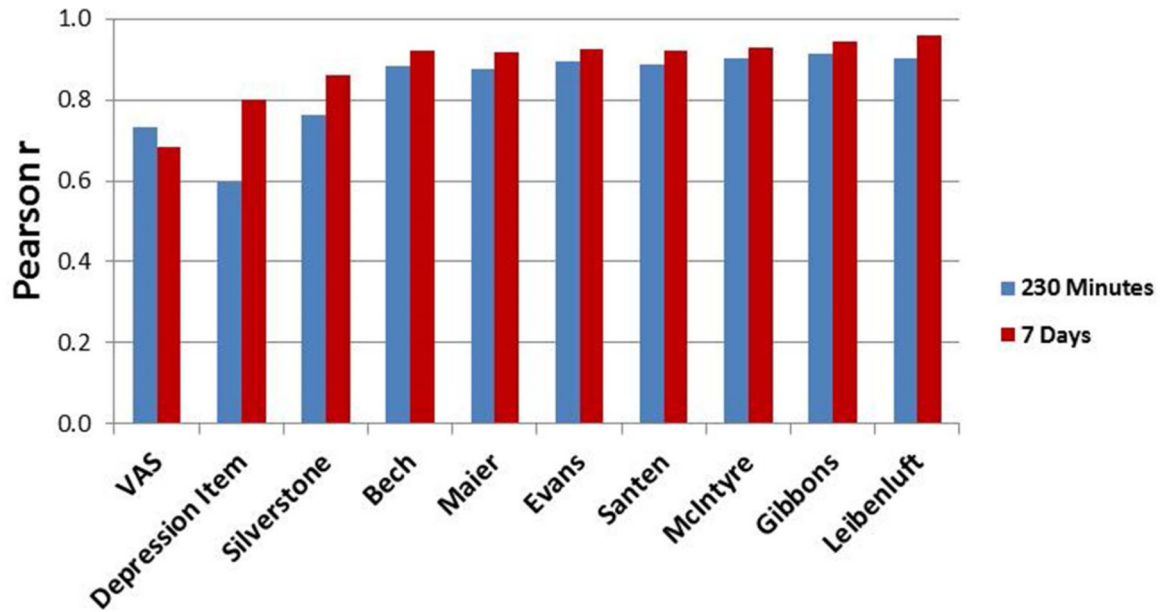
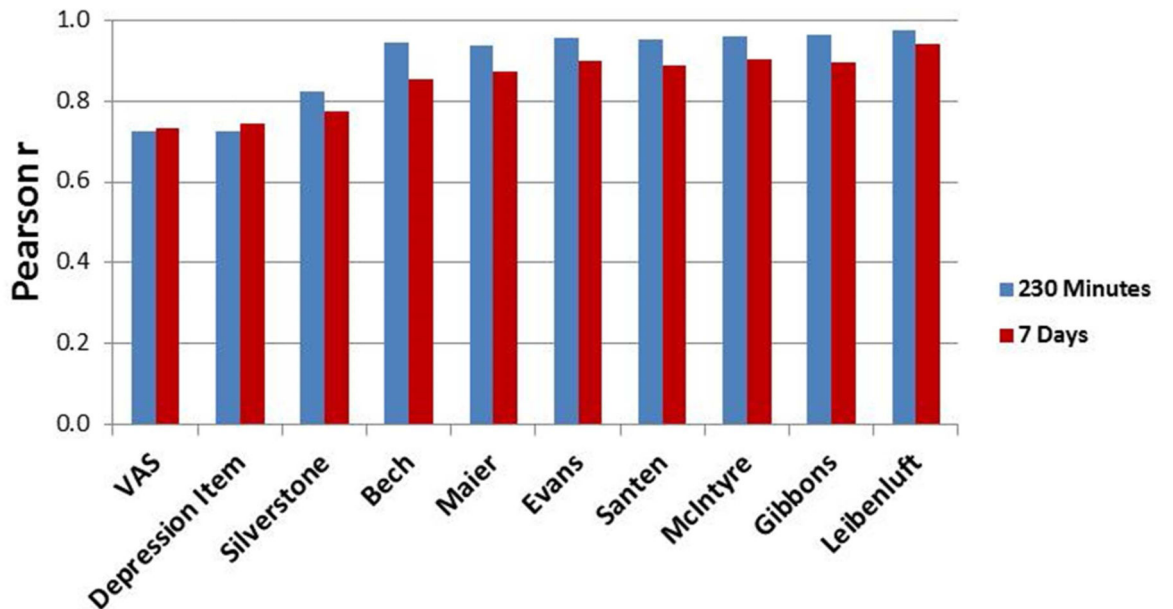
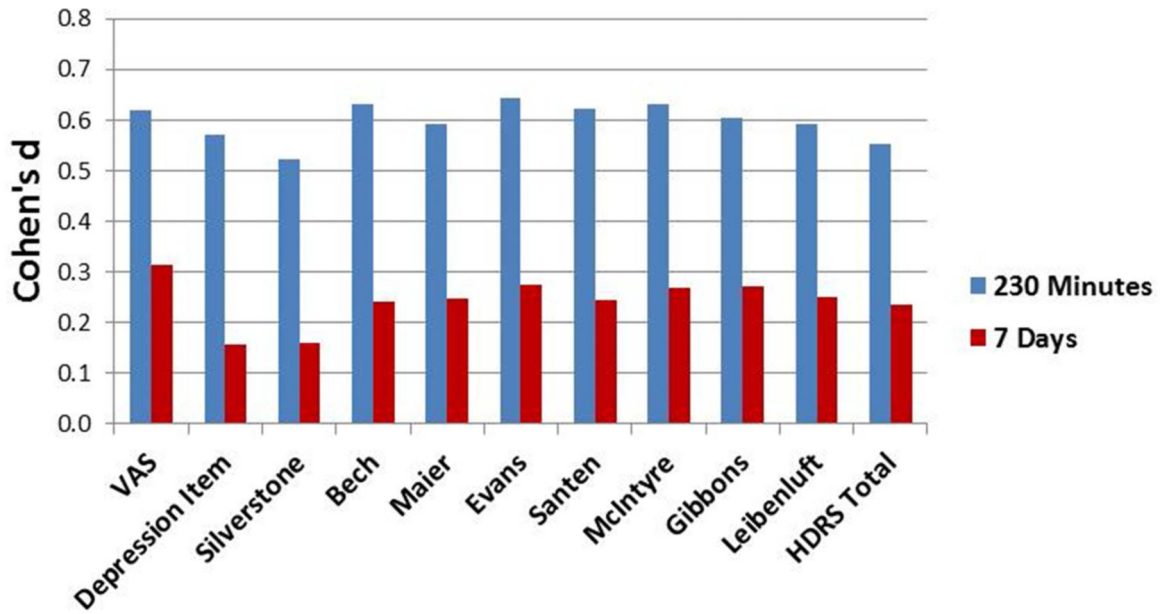


Figure 1. Cronbach's alpha for Hamilton Depression Rating Scale (HDRS) subscales using data at 230 minutes. G.I. = gastrointestinal.

A**B****Figure 2.**

(**A**) Pearson correlations between Hamilton Depression Rating Scale (HDRS) total score and HDRS subscores as well as Visual Analogue Scale (VAS). (**B**) Pearson correlations between percent changes from baseline HDRS total score and percent changes from baseline HDRS subscores as well as VAS.

A



B

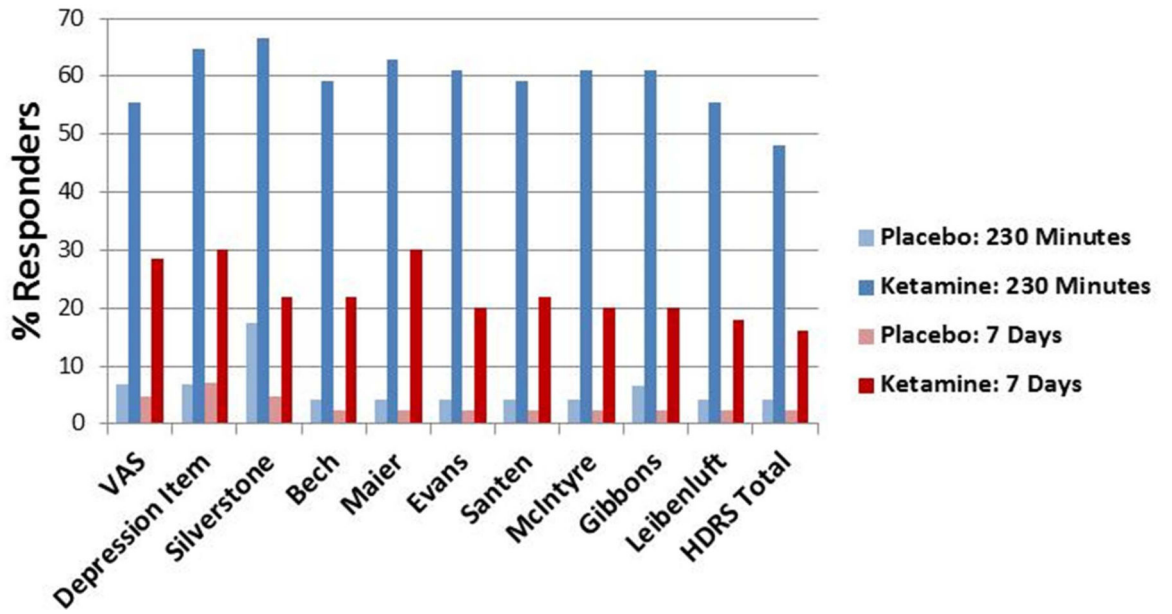


Figure 3. (A) Cohen's d effect sizes for comparisons of active drug versus placebo by scale. (B) Response rates for active drug versus placebo by scale. Response is defined as 50 percent improvement from baseline.

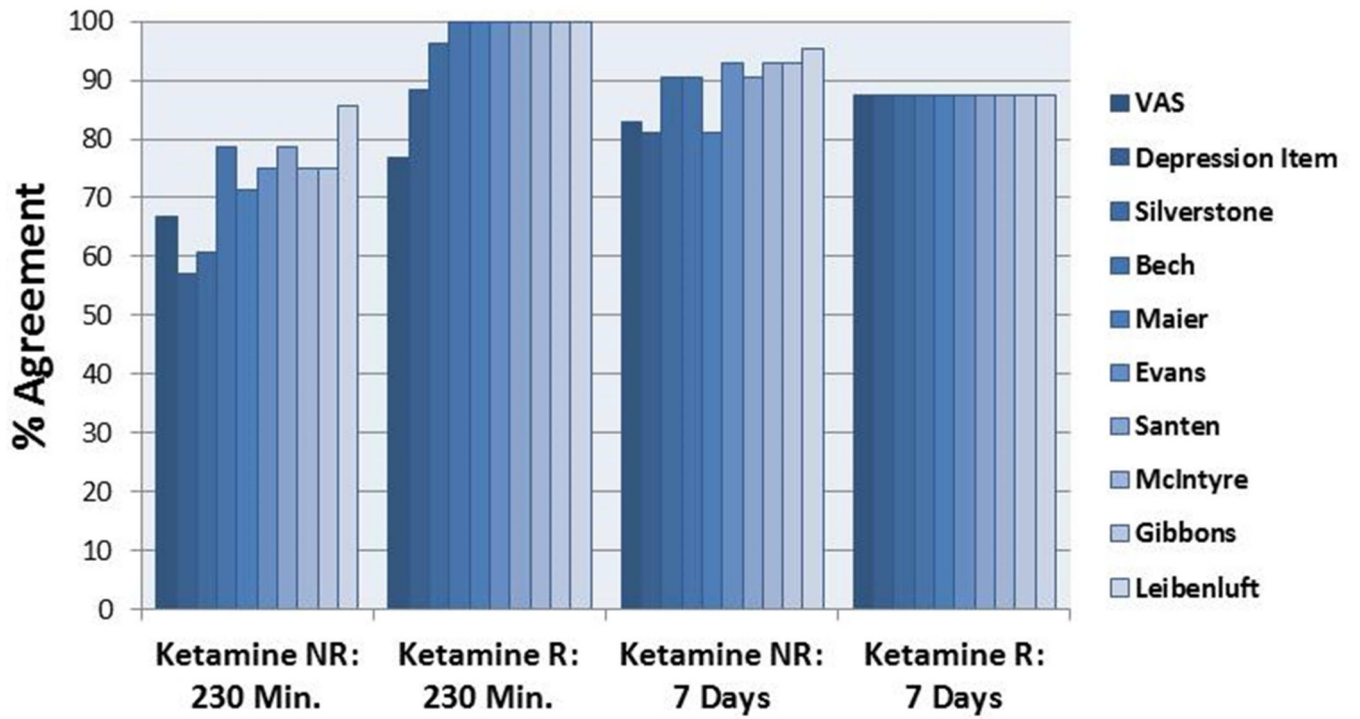


Figure 4. Percent agreement with Hamilton Depression Rating Scale (HDRS) total score on response categorization of responder or non-responder by subscale for active drug phase.

Table 1

Items Included in HDRS Subscales.

Item	Description	Silverstone	Maier	Bech	Evans	Santen	McIntyre	Gibbons	Shortened
1	Depressed Mood	•	•	•	•	•	•	•	•
2	Guilt		•	•	•	•	•	•	•
3	Suicide				•		•	•	•
4	Early Insomnia								
5	Middle Insomnia								
6	Late Insomnia								
7	Work & Activities		•	•	•	•	•	•	•
8	Motor Retardation		•	•		•			•
9	Agitation		•					•	•
10	Anxiety: Psychic	•	•	•	•	•	•	•	•
11	Anxiety: Somatic				•		•	•	•
12	Somatic Symptoms: G.I.								•
13	Somatic Symptoms: General		•	•	•	•	•	•	•
14	Genital Symptoms							•	•
15	Hypochondriasis								•
16	Weight Loss								
17	Lack of Insight								