Published in final edited form as:

Hum Psychopharmacol. 2016 May ; 31(3): 185–192. doi:10.1002/hup.2526.

Validation of the 17-item Hamilton Depression Rating Scale definition of response for adults with major depressive disorder using equipercentile linking to Clinical Global Impression scale ratings: analysis of Pharmacogenomic Research Network Antidepressant Medication Pharmacogenomic Study (PGRN-AMPS) data

William V. Bobo, MD, MPH^{*,a}, Gabriela C. Angleró^{a,b}, Gregory Jenkins, MS^c, Daniel K. Hall-Flavin, MD^a, Richard Weinshilboum, MD^d, and Joanna M. Biernacka, PhD^{a,c}

^aDepartment of Psychiatry and Psychology, Mayo Clinic, Rochester, MN, USA

^bSchool of Medicine, University of Puerto Rico, San Juan, Puerto Rico

^cDepartment of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

^dDepartment of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN, USA

Abstract

Objective—To define thresholds of clinically significant change in 17-item Hamilton Depression Rating Scale (HDRS-17) scores using the Clinical global Impression-Improvement (CGI-I) scale as a gold standard.

Methods—We conducted a secondary analysis of individual patient data from the Pharmacogenomic Research Network Antidepressant Medication Pharmacogenomic Study (PGRN-AMPS), an 8-week, single-arm clinical trial of citalopram or escitalopram treatment of adults with major depression. We used equipercentile linking to identify levels of absolute and percent change in HDRS-17 scores that equated with scores on the CGI-I at 4 and 8 weeks. Additional analyses equated changes in the 7-item HDRS and Bech 6 scale scores with CGI-I scores.

Results—A CGI-I score of 2 (much improved) corresponded to an absolute decrease (improvement) in HDRS-17 total score of 11 points, and a percent decrease of 50%–57%, from baseline values. Similar results were observed for percent change in HDRS-7 and Bech 6 scores. Larger absolute (but not percent) decreases in HDRS-17 scores equated with CGI-I scores of 2 in persons with higher baseline depression severity.

*Correspondence to: William V. Bobo, MD, MPH, 200 First Street SW, Generose 2A, Rochester, MN 55905, USA, William.v.bobo@mayo.edu, Telephone: 507-255-9412.

CONFLICT OF INTEREST

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

Conclusions—Our results support the consensus definition of response based on HDRS-17 scores (50% decrease from baseline). A similar definition of response may apply to the HDRS-7 and Bech 6.

Keywords

Hamilton Depression Rating Scale; response; major depressive disorder; citalopram; escitalopram; equipercentile linking

INTRODUCTION

For the last 50 years, the Hamilton Depression Rating Scale (HDRS) has been regarded as a gold standard measure of the severity of depressive symptoms, and has been widely employed in clinical trials of antidepressive treatments in persons with major depressive disorder (MDD) and other mood disorders (Hamilton, 1960;Williams, 2001). As with other classic depression rating instruments, the HDRS measures depressive symptoms on a continuous scale, but the degree to which given changes in a patient's HDRS score after treatment initiation translates to observable improvement by practicing clinicians is not always clear (Bech, 2006).

To improve the clinical translation of changes in depression rating scale scores, consensus definitions of response (a clinically significant degree of depressive symptom improvement after treatment initiation) and remission (the virtual absence of depressive symptoms) have been developed (Rush et al., 2006). For the 17-item version of the HDRS (HDRS-17), the accepted definition of response is a reduction in total score of 50% from baseline at a given follow up time point (Cusin et al., 2009).

Despite its wide use, very few empiric studies have examined the validity of the consensus-derived definition of response (Furukawa et al., 2007). One conceptually appealing approach has been to equate continuous HDRS scores (or change in HDRS scores) with scores on the Clinical Global Impression (CGI) scale (Guy, 1976), a tool that was developed for use in clinical trials to assess clinician's view of patients' symptoms and functioning after initiating study medications. The CGI has been shown to correlate well with standard depression rating scales (including the HDRS), and to be a useful measure of change in symptoms and functioning under treatment according to clinical judgment (Busner & Targum, 2007;Spielmanns & McFall, 2006). Two empirical studies used pooled data sets from clinical trials of various antidepressants and of mirtazapine and employed varying statistical approaches to equate HDRS-17 scores with scores on the Clinical Global Impression–Improvement (CGI-I) scales, the results of which supported the current consensus definition of response (Furukawa et al., 2007;Leucht et al., 2013).

We sought to replicate and extend this prior work by linking change in HDRS-17 and CGI-I scores in a large, single-site, clinical antidepressant trial. We conducted an additional set of analyses that linked the CGI-I scale with two HDRS-17-derived sub-scales, the HDRS-7 and Bech-6 (Bech, 1981;McIntyre et al., 2002). Both sub-scales were designed to overcome problems associated with the multidimensional nature of the HDRS-17 that may limit its usefulness as a measure of depression severity (Bagby et al., 2004).

METHODS

Source of Data

We conducted a secondary analysis of individual patient data from the Pharmacogenomic Research Network Antidepressant Medication Pharmacogenomic Study (PGRN-AMPS) (Mrazek et al., 2014). PGRN-AMPS was designed to assess the clinical outcomes of adults with non-psychotic MDD after 8 weeks of open-label treatment with citalopram or escitalopram, and to examine genetic factors associated with these outcomes. All PGRN-AMPS participants provided written informed consent. The PGRN-AMPS study protocol was approved by the institutional review board of the Mayo Clinic, Rochester, MN.

Sample and Treatment

A total of 922 adults (aged 18–84 years) with diagnoses of non-psychotic MDD were enrolled from the inpatient and outpatient practices of the Department of Psychiatry and Psychology at Mayo Clinic (Rochester, MN) between May 4, 2005 and October 10, 2012. Eligible participants had a HDRS-17 score ≥ 14 and a confirmed MDD diagnosis using the Structured Clinical Interview for DSM-IV (SCID) at the screening visit.

Persons were ineligible for trial participation if they had a medical contraindication to citalopram or escitalopram treatment; a history of poor response to an adequate therapeutic trial of citalopram or escitalopram; diagnosed schizophrenia, schizoaffective disorder, or bipolar disorder (type I, type II, or not otherwise specified); an active substance use disorder; or were pregnant or nursing or deemed by the study clinicians as being actively suicidal or at high risk for completed suicide. The SCID was also used at screening to identify and exclude persons with evidence of mania, bipolar disorders, and psychotic symptoms.

Eligible subjects received open-label citalopram (starting at 20 mg/day) or escitalopram (starting at 10 mg/day) treatment. The choice of study drug was based on the preference of the patient or the patient's referring physician. Face-to-face study visits occurred at baseline (the first day of study drug treatment) and at 4- and 8 weeks following the baseline visit to assess clinical response to study medications. The daily doses of study medications were increased at the week 4 visit (to 40 mg/day of citalopram or 20 mg/day of escitalopram) if there was an inadequate treatment response, defined as a 16-item, clinician-rated Quick Inventory of Depression Symptoms (QIDS-C-16) total score ≥ 9 (Rush et al., 2003).

Depressive Symptom Measures

The HDRS-17 was administered by experienced clinical raters at baseline, and at the week 4 and week 8 study visits. All clinical raters were certified as having a high rate of inter-rater reliability on both measures, which was reassessed on a quarterly basis in order to minimize rater drift. The present analysis focused on the HDRS-17, a clinician-rated measure of depressive symptoms that consists of 17 items rated using a semi-structured interview. Eight of the 17 HDRS-17 items are rated on a 5-point scale (0=absent; 1=doubtful or mild; 2=mild to moderate; 3=moderate to severe; 4=very severe), while the remaining 9 items are rated on a 3-point scale (0=absent; 1=doubtful or mild; 2=clearly present), yielding a minimum total

score of 0 (least severe) and a maximum score of 52 (most severe). Positive response was defined as a reduction (improvement) in HDRS-17 total scores by 50% from baseline.

Measure of Global Clinical State

In the present analysis, HDRS-17 scores were linked with CGI-I ratings, a proxy measure for clinically significant change. The CGI-I is commonly employed in psychopharmacology clinical trials as a measure of the clinical significance of subjects' change in symptoms and functioning, based on the clinician-rater's experience with other patients having the same diagnosis.

The CGI-I was rated on a 7-point scale using the following scores: 1=very much improved; 2=much improved; 3=minimally improved; 4=no change; 5=minimally worse; 6=much worse; and 7=very much worse. In the PGRN-AMPS study, all CGI-I ratings were completed by study clinicians at the week 4 and week 8 study visits. As with prior studies, a CGI-I score 2 was used as the threshold for defining response (Leucht et al., 2013). All study clinicians were experienced in the evaluation and treatment of depressed adults, and were considered to have sufficient experience to provide valid CGI ratings.

Statistical Analysis

Descriptive statistics were used to summarize the demographic and clinical characteristics of the PGRN-AMPS study participants, and were presented as means $\pm$ SD and proportions. The analyses included data on study participants with complete HDRS-17 and CGI-I ratings at baseline (HDRS-17 only) and at a given follow-up time point (week 4, week 8).

The main interest was to equate the scales of HDRS-17-based measures (change in depressive symptoms) and CGI-I measures (clinical significance of symptom change) using equipercentile linking, a statistical process that is used to find equivalent points on different but correlated scales. First, correlation between the CGI- and HDRS-17-based measures was assessed using Spearman rank correlation coefficients, testing the coefficients versus no correlation using an F-test with a threshold for statistical significance set at $p < 0.05$. Equipercentile linking was then performed using the *equate* v2.0.3 package in R v3.1.1 by, first, calculating the empirical distribution functions for both of the measures to be linked (i.e., percentiles of each measure). The percentiles were then matched between the two measures creating a link between the two scales. Thus, for a given score on a CGI-based measure, a corresponding score (or range of scores) on the HDRS-17-based measure with the same percentile rank was identified, linking the two types of measures. All of the resulting pairs of scores were plotted on a graph, such that each point in the graph represented equivalent scores on CGI- and HDRS-17-based measures. Each of these points was connected by a smooth curve, thus displaying the equipercentile relationship between CGI- and HDRS-17-based measures across the observed range of values on both measures. Although other analytic approaches were considered, the equipercentile method was preferred given that it is non-parametric (i.e., it does not require a specific type of distribution of measured values) and accounts for possible measurement error for both scales (Kolen & Brennan, 2014).

To address potential concerns about the multidimensional factor structure of the HDRS-17 (Bagby, Ryder, Schuller, & Marshall 2004), we repeated the aforementioned analyses after replacing the HDRS-17 with the HDRS-7 and the Bech 6. The HDRS-7 and Bech-6 are two validated subscales of the HDRS-17 that were designed to measure core depressive symptoms (Bech, 2006; Faries et al., 2000; Licht et al., 2005; McIntyre et al., 2002; Tomba & Bech, 2012). The HDRS-7 contains the full-scale HDRS items that measure mood, guilt, work and interest, psychic anxiety, energy, somatic anxiety, and suicide; whereas the Bech 6 contains corresponding full-scale HDRS items assessing mood, guilt, work and interest, psychic anxiety, energy, and psychomotor retardation (Kennedy, 2008). Additionally, we examined the potential effect of baseline depression severity on the relationship between HDRS- and CGI-based measures by again repeating the analyses within strata based on a median split of baseline HDRS-17 scores (Leucht et al., 2013). The higher severity stratum was composed of study subjects with the median HDRS-17 score at baseline; all others were classified into the lower severity stratum.

RESULTS

Subject demography and clinical characteristics

The demographic and clinical characteristics of the study sample at baseline, week 4, and week 8 are summarized in Table 1. Of the 922 subjects enrolled in the PGRN-AMPS study, 920, 677, and 603 subjects had complete data at baseline (HDRS-17 only), week 4, and week 8, respectively. The most common reasons for early withdrawal from the study by week 4 ($n=245$) were no longer meeting eligibility criteria ($n=96$), loss to follow-up ($n=65$), adverse effects ($n=32$), subject withdrawal of consent ($n=14$), request of referring physician ($n=13$), and subject refusal of further treatment with citalopram or escitalopram ($n=12$). The most common reasons for early withdrawal between weeks 4 and 8 ($n=74$) were loss to follow-up ($n=45$), side-effects ($n=13$), and no longer meeting eligibility criteria ($n=7$). At each time point, study subjects were predominantly middle-aged, Caucasian, and female (Table 1).

Correlation between HDRS- and CGI-based measures

Spearman correlations between HDRS- and CGI-based measures are presented in Table 2. All HDRS- and CGI-based measures were strongly and significantly correlated at all time points, thus permitting equipercetile linking to occur.

Linkage of absolute and percent change in HDRS-17 with CGI-I scores

The results of equipercetile linking of absolute and percent changes (from baseline) in HDRS-17 scores with CGI-I scores at weeks 4 and 8 are graphically displayed in Figures 1A and 1B. CGI-I scores of 3 (minimally improved), 2 (much improved), and 1 (very much improved) were linked with absolute reductions (improvement) of 5 to 6 points, 11 points, and 17 to 18 points in HDRS-17 total scores from baseline values, consistently, at both time points (Figure 1A).

As shown in Figure 1B, the relationship between percent change in HDRS-17 scores from baseline and CGI-I scores were consistent at both time points. CGI-I scores of 3 (minimally

improved), 2 (much improved), and 1 (very much improved) were linked with reductions (improvement) in HDRS-17 total scores of 25% to 30%, 50% to 57%, and 79% to 82%, respectively, from baseline values.

Linkage of HDRS-17 subscale (HDRS-7, Bech-6) and CGI-I scores

Results of equipercetile linking of absolute and percent changes (from baseline) in HDRS-7 scores with CGI-I scores at weeks 4 and 8 are graphically displayed in Figures 2A and 2B. CGI-I scores of 3 (minimally improved), 2 (much improved), and 1 (very much improved) were linked with absolute reductions of 3 to 4 points, 7 to 8 points, and 12 points in HDRS-7 scores from baseline values, at weeks 4 and 8 (Figure 2A); and with reductions in HDRS-7 total scores of 25% to 31%, 53% to 67%, and 80% to 87%, respectively, from baseline values (Figure 2B).

Similar results were obtained for equipercetile linking of Bech-6 measures with CGI-I scores (Figures 3A and 3B). CGI-I scores of 3 (minimally improved), 2 (much improved), and 1 (very much improved) were linked with absolute reductions of 3 to 4 points, 7 points, and 11 to 12 points in Bech-6 scores from baseline values, at weeks 4 and 8 (Figure 3A); and with reductions in Bech-6 total scores of 25% to 32%, 57% to 63%, and 83% to 90%, respectively, from baseline values (Figure 3B).

Stratified analyses according to baseline depressive symptom severity

Figures 4A and 4B display results of equipercetile linking of absolute and percent change (from baseline) in HDRS-17 scores with CGI-I scores at weeks 4 and 8, stratified into higher- and lower-severity groups based on a median split of baseline HDRS-17 values, using a median value of 21. The relationship between absolute change in HDRS-17 scores and CGI-I scores was affected by baseline depression severity at both 4 and 8 weeks—in general, a greater absolute reduction in HDRS-17 score from baseline was associated with a given CGI-I score for persons in the higher severity group, as compared with the lower severity group (Figure 4A). This effect was not observed, however, in analyses that linked percent (rather than absolute) change in HDRS-17 scores with CGI-I scores (Figure 4B). For example, a CGI-I score of 2 (much improved) was linked with absolute reductions in HDRS-17 scores (from baseline) of 13 to 14 points in the higher severity group and 9 points in the lower severity group at 4 and 8 weeks. Percent reductions (from baseline) in HDRS-17 scores corresponding to a CGI-I score of 2 were 50% to 57% in the higher severity group and 50% to 56% in the lower severity group.

DISCUSSION

Changes in continuous measures of depressive symptoms, such as the HDRS-17, have been used as primary outcome measures in clinical antidepressant trials for decades, but this approach has been criticized for lacking clear and empirically validated thresholds for clinically significant change in patients' symptoms or functioning (Bech, 2006; Kriston & von Wolff, 2011; Masson & Tejani, 2013). Using data from a large, 8-week, single-site clinical trial of citalopram or escitalopram for treating major depression in adults, we found that a CGI-I score of 2 (much improved) equated to an absolute reduction (improvement) in

HDRS-17 scores of 11 points, and a percent reduction of 50%–57%, from baseline values. The latter finding supports the widely accepted definition of clinical response, i.e., improvement in HDRS-17 total score by 50% from baseline.

Leucht and colleagues were the first to apply equipercentile linking methods to define clinically significant thresholds for continuous psychiatric symptom ratings (and changes in the total scores), using CGI-S and CGI-I ratings as a measure of clinical significance (Leucht et al., 2005a; Leucht et al., 2005b; Leucht et al., 2006). In an analysis of a pooled data set that consisted of all manufacturer-sponsored clinical trials of mirtazapine for MDD in adults (43 studies of various design, totaling 7,131 patients), a CGI-I score of 2 corresponded to an absolute reduction in HDRS-17 total score of 10 points and a percent reduction of 50%–60% (Leucht et al., 2013). In that data set, follow-up time extended up to 4 weeks. Our results from a separate sample of adults with MDD who received citalopram or escitalopram treatment for 8 weeks replicate those of Leucht et al., and are also consistent with results of an anchor-based analysis of a data set consisting of 7 short-term (6–13 weeks) clinical trials of various antidepressants (imipramine, amitriptyline, trazodone, fluoxetine, paroxetine, fluvoxamine) in patients with MDD (Furukawa et al., 2007). In that study, percent reductions in HDRS scores by 46%–72% were equated to a CGI-I score of 2.

In addition to replicating these results, we were also interested in identifying thresholds of clinically significant change in HDRS-7 and Bech-6 scores. The use of HDRS total scores as a measure of depressive symptoms has been criticized, in part, on the basis of multidimensionality that may limit its sensitivity to differences between antidepressive treatments (Bagby et al., 2004; Faries et al., 2000). While some degree of multidimensionality ensures adequate coverage of the clinical features of MDD, decreases in total scores during antidepressant treatment may not necessarily reflect improvement in core depressive symptoms (Kennedy, 2008). Both the HDRS-7 and Bech-6 are unidimensional sub-scales of the HDRS that measure core depressive symptoms (Bech, 2006; Faries et al., 2000; McIntyre et al., 2002; Tomba & Bech, 2012). Five HDRS items (depressed mood, anhedonia, guilt, fatigue, and psychological anxiety) are common to both sub-scales. The HDRS-7 also includes somatic anxiety and suicidality, whereas the Bech-6 includes psychomotor retardation. Based on our findings, it would appear that a similar threshold of relative change from baseline (50% or greater) used to define response when using HDRS-17 may also be used for the HDRS-7 and Bech-6.

To assess the effect of baseline depression severity on our main findings, we performed a median split of baseline HDRS-17 data to classify subjects into higher- and lower-severity groups in a manner similar to that of Leucht and colleagues (Leucht et al., 2013). In their report, Leucht et al. observed a clear impact of baseline depression severity at all follow-up time points, wherein larger absolute changes in HDRS-17 scores corresponded with a given CGI-I score in patients with higher baseline depression severity, as compared with those having lower baseline depression severity. This effect was not observed when considering percent change in HDRS-17 scores. As pointed out by the authors, the mean (SD) HDRS-17 at baseline in their data set was 23.8 (4.6), likely a reflection of a higher HDRS-17 threshold for study entry than was the case in the PGRN-AMPS study. The mean (SD) baseline HDRS-17 score in PGRN-AMPS was 21.2 (5.9). Differences between the two studies in

HDRS-17 total scores at baseline were therefore small, and we also observed a clear effect of baseline depression severity on the absolute—but not relative (percent)—change in HDRS-17 scores equating with a given CGI-I score used to define clinically significant improvement.

Strengths of our study include a large sample size, the use of experienced study clinicians for completing CGI evaluations, and the use of trained and certified clinical raters who underwent periodic inter-rater reliability assessments of HDRS assessments. Additionally, CGI-I and HDRS ratings were completed by different individuals at each study visit, and the correlations between these measures were very strong at both follow-up time points. There are also limitations to consider. Our study used an open-label design, and did not include CGI-severity of illness ratings; thus, we were unable to explore or validate HDRS-17 thresholds for remission. Periodic assessment of inter-rater reliability for CGI assessments was not performed. Additionally, study procedures did not prevent access to HDRS-17 scores by study clinicians at the face-to-face visits, which could have influenced the CGI-I ratings. Additional information on attrition from the study due to inefficacy was unavailable. Patients who were responding poorly to study medications may have been represented among those who were lost to follow up, refused ongoing study drug treatment, or withdrew consent for further participation. It is therefore difficult to ascertain whether subjects remaining in the study at weeks 4 and 8 represented a more homogeneously positive-responding cohort, and what effect this may have had on study findings. And finally, although our findings are remarkably consistent with prior research, they are ultimately derived from a single-site antidepressant trial and generalizability may be therefore limited.

CONCLUSION

Even with these limitations, our findings support the accepted definition of clinical response based on HDRS-17 scores, and suggest that a similar definition of response may be applied to two HDRS-17-derived sub-scales focused on core depressive symptoms.

Acknowledgments

The PGRN-AMPS study was supported by U19 GM61388 and R01 GM28157 (to Drs. Liewei Wang and Richard Weinshilboum). Dr. Bobo's research has been supported by the National Institute of Mental Health, the Mayo Foundation, and the Brain and Behavior Research Foundation (formerly NARSAD).

References

- ECDEU Assessment Manual for Psychopharmacology, revised (DHEW Publ No ADM 76-338). National Institute of Mental Health; Rockville, MD: 1976. Clinical Global Impressions.
- 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. Rating scales for affective disorders: their validity and consistency. *Acta Psychiatr Scand Suppl*. 1981; 295:1–101.
- Bech P. Rating scales in depression: limitations and pitfalls. *Dialogues Clin Neurosci*. 2006; 8(2):207–215.
- Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. *Psychiatry (Edgmont)*. 2007; 4(7):28–37. [PubMed: 20526405]

- Cusin, C.; Yang, H.; Yeung, A.; Fava, M. Rating scales for depression. In: Baer, L.; Blais, MA., editors. *Handbook of Clinical Rating Scales and Assessment in Psychiatry and Mental Health*. Humana Press; New York: 2009. p. 7-36.
- 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]
- Furukawa TA, Akechi T, Azuma H, Okuyama T, Higuchi T. Evidence-based guidelines for interpretation of the Hamilton Rating Scale for Depression. *J Clin Psychopharmacol*. 2007; 27(5): 531–534. [PubMed: 17873700]
- HAMILTON M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960; 23:56–62. [PubMed: 14399272]
- Kennedy SH. Core symptoms of major depressive disorder: relevance to diagnosis and treatment. *Dialogues Clin Neurosci*. 2008; 10(3):271–277. [PubMed: 18979940]
- Kolen, MJ.; Brennan, RL. *Test Equating, Scaling, and Linking - Methods and Practices*. Springer; New York: 2014.
- Kriston L, von Wolff WA. Not as golden as standards should be: interpretation of the Hamilton Rating Scale for Depression. *J Affect Disord*. 2011; 128(1–2):175–177. [PubMed: 20696481]
- Leucht S, Fennema H, Engel R, Kaspers-Janssen M, Lepping P, Szegedi A. What does the HAMD mean? *J Affect Disord*. 2013; 148(2–3):243–248. [PubMed: 23357658]
- Leucht S, Kane JM, Etschel E, Kissling W, Hamann J, Engel RR. Linking the PANSS, BPRS, and CGI: clinical implications. *Neuropsychopharmacology*. 2006; 31(10):2318–2325. [PubMed: 16823384]
- Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel R. Clinical implications of Brief Psychiatric Rating Scale scores. *Br J Psychiatry*. 2005a; 187:366–371. [PubMed: 16199797]
- Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel RR. What does the PANSS mean? *Schizophr Res*. 2005b; 79(2–3):231–238. [PubMed: 15982856]
- Licht RW, Qvitzau S, Allerup P, Bech P. Validation of the Bech-Rafaelsen Melancholia Scale and the Hamilton Depression Scale in patients with major depression; is the total score a valid measure of illness severity? *Acta Psychiatr Scand*. 2005; 111(2):144–149. [PubMed: 15667434]
- Masson SC, Tejani AM. Minimum clinically important differences identified for commonly used depression rating scales. *J Clin Epidemiol*. 2013; 66(7):805–807. [PubMed: 23618794]
- McIntyre R, Kennedy S, Bagby RM, Bakish D. Assessing full remission. *J Psychiatry Neurosci*. 2002; 27(4):235–239. [PubMed: 12174732]
- Mrazek DA, Biernacka JM, McAlpine DE, Benitez J, Karpyak VM, Williams MD, Hall-Flavin DK, Netzel PJ, Passov V, Rohland BM, Shinozaki G, Hoberg AA, Snyder KA, Drews MS, Skime MK, Sagen JA, Schaid DJ, Weinshilboum R, Katselnick DJ. Treatment outcomes of depression: the pharmacogenomic research network antidepressant medication pharmacogenomic study. *Journal of Clinical Psychopharmacology*. 2014; 34(3):313–317. [PubMed: 24743713]
- Rush AJ, Kraemer HC, Sackeim HA, Fava M, Trivedi MH, Frank E, Ninan PT, Thase ME, Gelenberg AJ, Kupfer DJ, Regier DA, Rosenbaum JF, Ray O, Schatzberg AF. Report by the ACNP Task Force on response and remission in major depressive disorder. *Neuropsychopharmacology*. 2006; 31(9):1841–1853. [PubMed: 16794566]
- Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, Markowitz JC, Ninan PT, Kornstein S, Manber R, Thase ME, Kocsis JH, Keller MB. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biological Psychiatry*. 2003; 54(5):573–583. [PubMed: 12946886]
- Spielmans GI, McFall JP. A comparative meta-analysis of Clinical Global Impressions change in antidepressant trials. *J Nerv Ment Dis*. 2006; 194(11):845–852. [PubMed: 17102709]
- Tomba E, Bech P. Clinimetrics and clinical psychometrics: macro- and micro- analysis. *Psychotherapy and Psychosomatics*. 2012; 81(6):333–343. [PubMed: 22964522]
- Williams JB. Standardizing the Hamilton Depression Rating Scale: past, present, and future. *Eur Arch Psychiatry Clin Neurosci*. 2001; 251(Suppl 2):II6–12. [PubMed: 11824839]

Figure 1.

Linkage of absolute (A) and percent (B) change in HDRS-17 score with CGI-I scores at week 4 (black line) and week 8 (red line)

Figure 2.
Linkage of absolute (A) and percent (B) change in HDRS-7 score with CGI-I scores at week 4 (black line) and week 8 (red line)

Figure 3.
Linkage of absolute (A) and percent (B) change in Bech-6 score with CGI-I scores at week 4 (black line) and week 8 (red line)

Figure 4. Linkage of absolute (A) and percent (B) change in HDRS-17 and CGI-I scores at week 4 (black lines) and week 8 (red lines), stratified by higher (dashed lines) or lower (solid lines) depression severity at baseline

Table 1

Subject demography and clinical ratings at baseline, 4 weeks, and 8 weeks

	Baseline	4 weeks	8 weeks
N	922	677	603
Age, mean (SD)	39.1 (14.3)	39.7 (13.7)	40.2 (13.5)
Female, n (%)	570 (61.8%)	426 (62.9%)	381 (63.2%)
Caucasian race, n (%)	848 (94.2%)	643 (96.5%)	575 (97.0%)
HDRS-17	21.2 (5.9)	12.2 (6.6)	8.9 (6.0)
HDRS-7	13.2 (3.4)	7.2 (4.1)	5.2 (3.9)
Bech-6	12.1 (3.1)	6.4 (3.9)	4.4 (3.6)

Key: Bech-6 = six-item subscale derived from the full Hamilton Depression Rating Scale; CGI-S = severity of illness subscale of the Clinical Global Impression Scale; HDRS-7 = seven-item subscale derived from the full Hamilton Depression Rating Scale; HDRS-17 = seventeen-item Hamilton Depression Rating Scale.

Table 2

Correlations between absolute and percent change in HDRS-17 scores and CGI-I scores at 4 weeks and at 8 weeks

-----Linking variables ^a -----		Time point	Spearman correlation coefficient	p-value ^b
HDRS-17, absolute	CGI-I	Week 4	0.657	$3.1 * 10^{-84}$
		Week 8	0.617	$2.7 * 10^{-64}$
HDRS-17, percent	CGI-I	Week 4	0.726	$2.5 * 10^{-111}$
		Week 8	0.729	$1.2 * 10^{-100}$
HDRS-7, absolute	CGI-I	Week 4	0.661	$1.1 * 10^{-85}$
		Week 8	0.627	$5.5 * 10^{-67}$
HDRS-7, percent	CGI-I	Week 4	0.718	$1.3 * 10^{-107}$
		Week 8	0.722	$1.4 * 10^{-97}$
Bech-6, absolute	CGI-I	Week 4	0.668	$4.4 * 10^{-88}$
		Week 8	0.622	$1.4 * 10^{-65}$
Bech-6, percent	CGI-I	Week 4	0.719	$5.9 * 10^{-108}$
		Week 8	0.707	$3.3 * 10^{-92}$

Key: Bech-6 = six-item subscale derived from the full Hamilton Depression Rating Scale; CGI-S = severity of illness subscale of the Clinical Global Impression Scale; HDRS-7 = seven-item subscale derived from the full Hamilton Depression Rating Scale; HDRS-17 = seventeen-item Hamilton Depression Rating Scale; = change from baseline for a given rating scale score.

^aDenotes the two variables being correlated using Spearman correlation, prior to being equated using equipercntile linking.

^bP-values are expressed using scientific notation, where a p-value of $5.5 * 10^{-8}$ equates to 0.000000055.