Translating the BDI and BDI-II into the HAMD and vice versa with equipercentile linking

David J. A. Dozois

University of Western Ontario, ddozois@uwo.ca

Follow this and additional works at: https://ir.lib.uwo.ca/huronpsychologypub

Part of the Clinical Psychology Commons

Citation of this paper:
Dozois, David J. A., "Translating the BDI and BDI-II into the HAMD and vice versa with equipercentile linking" (2019). Faculty Publications. 5.
https://ir.lib.uwo.ca/huronpsychologypub/5
Translating the BDI and BDI-II into the HAMD and vice versa with equipercentile linking

Toshi A. Furukawa1, Mirjam Reijnders2, Sanae Kishimoto3, Masatsugu Sakata3, Robert J. DeRubeis4, Sona Dimidjian5, David J.A. Dozois6, Ulrich Hegerl7, Steven D. Hollon8, Robin B. Jarrett9, François Lеспérance10, Zindel V. Segal11, David C. Mohr12, Anne D. Simons13, Lena C. Quilty14, Charles F. Reynolds III15, Claudio Gentili16, Stefan Leucht17, Rolf R. Engel18 and Pim Cuijpers2

Abstract

Aims. The Hamilton Depression Rating Scale (HAMD) and the Beck Depression Inventory (BDI) are the most frequently used observer-rated and self-report scales of depression, respectively. It is important to know what a given total score or a change score from baseline on one scale means in relation to the other scale.

Methods. We obtained individual participant data from the randomised controlled trials of psychological and pharmacological treatments for major depressive disorders. We then identified corresponding scores of the HAMD and the BDI (369 patients from seven trials) or the BDI-II (683 patients from another seven trials) using the equipercentile linking method.

Results. The HAMD total scores of 10, 20 and 30 corresponded approximately with the BDI scores of 10, 27 and 42 or with the BDI-II scores of 13, 32 and 50. The HAMD change scores with the BDI of −20 and −10 and with the BDI-II of −35 and −16. The results can help clinicians interpret the HAMD or BDI scores of their patients in a more versatile manner and also help clinicians and researchers evaluate such scores in the literature or the database, when scores on only one of these scales are provided. We present a conversion table for future research.

Introduction

It is important to evaluate the course of major depressive disorder (MDD) using quantitative rating scales of symptoms. Various rating scales have been developed to evaluate the severity of MDD in research and clinical settings. These measures can be categorised as clinician-rated scales such as the Hamilton Rating Scale for Depression (HAMD) (Hamilton, 1960; Williams et al., 2008), Montgomery Åsberg Depression Rating Scale (MADRS) (Montgomery and Asberg, 1979) or Quick Inventory of Depression Symptomatology Clinician Rating (Rush et al., 2003), and self-report scales such as the Beck Depression Inventory (BDI) (Beck et al., 1961) and its revised version (BDI-II) (Beck et al., 1996), Patient Health Questionnaire-9 (Kroenke et al., 2001) or Quick Inventory of Depression Symptomatology self-report version (Rush et al., 2003). Although numerous scales for rating depression severity have been developed to date, the HAMD is the most commonly used clinician-rated scale in research and clinical settings. The HAMD has been used as a main outcome measure in randomised controlled trials of pharmacotherapy and psychotherapy for depression. In the latest network meta-analysis of antidepressant medications for MDD, 464 of 522 eligible studies
reported baseline severity scores using the HAMD (Cipriani et al., 2018). Similarly, the network meta-analysis of psychotherapy for MDD showed that 75 of 198 studies reported outcomes using the HAMD (Barth et al., 2013). On the other hand, the BDI is one of the most widely used self-rating scales. The BDI/BDI-II have been used particularly often as the outcome measure in psychotherapy trials. According to the above-mentioned network meta-analysis studies of psychotherapies for depression, 116 of 198 studies used the BDI and 25 of 198 studies used the BDI-II as an outcome measure of the trial (Barth et al., 2013).

Although both the HAMD and the BDI/BDI-II are standard measures to assess depression severity, no study has yet examined how scores on the HAMD can be converted to the BDI/BDI-II scores or vice versa. It is important to link these two most commonly used scales for comparison of the baseline severity or treatment outcome. Several studies identified the corresponding scores of simultaneous HAMD and other scales such as MADRS (Leucht et al., 2018) and the Clinical Global Impression (Leucht et al., 2013a) using the equipercentile linking method (Linn, 1993). The equipercentile linking method has been used extensively for various other scales in previous publications (Leucht et al., 2005, 2006, 2013b, 2016; Furukawa et al., 2009; Levine and Leucht, 2013; Samara et al., 2014). In the current study, we attempted to link the HAMD and the BDI/BDI-II applying the same procedure.

Method

Database

We used an existing database of psychological treatments for depression which is updated annually through comprehensive literature searches in the bibliographic databases of PubMed, PsycINFO, EMBASE and the Cochrane Library (Cuipers et al., 2008). Appendix A provides the full search strings used. This database has been used in a series of previously published meta-analyses (Bower et al., 2013; Furukawa et al., 2017; Karyotaki et al., 2017). For this linking study, we focused on the individual participant data (IPD) that we had assembled to conduct IPD meta-analytic studies comparing cognitive-behavioural therapy (CBT), antidepressant pharmacotherapy and their combination (Weitz et al., 2017).

Rating scales

The HAMD is based on clinical interviews. We used the HAMD 17-item version in this analysis. The 17 items consist of nine symptoms (depressed mood, self-deprecation and guilt feelings, suicidal impulses, work and interests, psychomotor retardation, agitation, anxiety psychic, anxiety somatic, hypochondriasis) rated between 0 (absent) to 4 (very severe), and eight symptoms (initial insomnia, middle insomnia, delayed insomnia, gastrointestinal, general somatic, sexual interests, loss of insight, weight loss) rated between 0 (absent) to 2 (clearly present) (Hamilton, 1960). The maximum score of the HAMD is therefore 52. A meta-analysis showed that the HAMD has sufficient internal consistency (Cronbach’s $\alpha = 0.79$), inter-rater reliability (intra-class correlation coefficient (ICC) = 0.94) and test–retest reliability (ICC = 0.93) (Trajkovic et al., 2011).

The BDI is a 21-item patient’s self-report questionnaire that measures the depression severity (Beck et al., 1961). All items of the BDI are rated on a four-point Likert scale ranging from 0 to 3, and the total score therefore ranges from 0 to 63. Beck et al. developed the revised version of the BDI to harmonise its item contents with the modern diagnostic criteria for MDD in Diagnostic and Statistical Manual of Mental Disorders (DSM)-IV, while maintaining the same number of items and range of scale as the BDI (Beck et al., 1996). The BDI has sufficient internal consistency in psychiatric patients (Cronbach’s $\alpha$ ranging from 0.76 to 0.95) and non-psychiatric populations (Cronbach’s $\alpha$ ranging from 0.73 to 0.92) (Beck et al., 1988). The BDI-II also has sufficient internal consistency ($\alpha = 0.93$ among college students, $\alpha = 0.92$ among outpatients) (Beck et al., 1996). According to a survey of 1022 undergraduate students, the mean score of the BDI-II was 1.54 points higher than that of the BDI (Dozois et al., 1998). However, the two scales showed high correlation ($r = 0.93$), suggesting convergence of the two scales.

Statistical analysis

We first drew scatterplots and calculated Spearman correlation coefficients between HAMD and BDI or BDI-II, at baseline and at end of treatment. We then applied the equipercentile linking procedure (Linn, 1993), which is a technique that identifies those scores on the HAMD and the BDI or the BDI-II that have the same percentile ranks, thus allowing for a nominal translation from HAMD scores to BDI or BDI-II scores or vice versa by using their percentile values. We used Microsoft Excel to realise the analytical procedures described in Chapter 2 of Kolen and Brennan (1995) and to draw the diagrams. We merged the baseline and endpoint measurements to produce the final linking curves and the table of conversion.

Because many trials take the change scores from baseline to end of treatment, instead of raw scores at end of treatment, as the primary outcome, we also examined the linking relationships between change scores of the HAMD and the BDI/BDI-II.

Subgroup and sensitivity analyses

In order to examine possible subgroup differences, we conducted the same analyses for men and women separately, and also for dropouts.

Results

Included studies

Figure 1 presents the flow of the literature search. For this study we used the search that was conducted in January 2016. After removing the duplicates from different data sources, two independent reviewers examined 13 384 titles and abstracts, retrieved 1885 full-text articles and finally identified 75 studies that compared CBT, antidepressant pharmacotherapy or their combination in the acute phase treatment of depression.

Of these, authors of 14 studies provided IPD including both HAMD and BDI (Rush et al., 1977; Murphy et al., 1984; Elkin et al., 1989; Holton et al., 1992; Jarrett et al., 1999; Reynolds et al., 1999; Mohr et al., 2001) or BDI-II (DeRubeis et al., 2005; Dimidjian et al., 2006; Lesperance et al., 2007; McBride et al., 2007; Dozois et al., 2009; Hegerl et al., 2010; Quilty et al., 2014) (Table 1). Studies using the BDI were published mainly before 2000, while those using the BDI-II were all published after 2000. The 14 studies included 1536 participants: their mean age
Fig. 1. Flowchart of study identification.

### Table 1. Included studies and their characteristics

<table>
<thead>
<tr>
<th>Studies</th>
<th>N</th>
<th>Age</th>
<th>Sex (male/female)</th>
<th>Treatments</th>
<th>Treatment duration (weeks)</th>
<th>Scale used for inclusion</th>
<th>Baseline HAMD</th>
<th>Baseline BDI</th>
<th>Endpoint HAMD</th>
<th>Endpoint BDI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elkin et al. (1989)</td>
<td>116</td>
<td>34.5</td>
<td>35/81</td>
<td>CBT v. ADM</td>
<td>16</td>
<td>HAMD ≥ 14</td>
<td>26.7 (8.5)</td>
<td>7.8 (6.8)</td>
<td>9.0 (9.7)</td>
<td></td>
</tr>
<tr>
<td>Hollon et al. (1992)</td>
<td>107</td>
<td>32.6</td>
<td>21/86</td>
<td>CBT v. ADM v. CBT + ADM</td>
<td>12</td>
<td>HAMD ≥ 14 and BDI ≥ 20</td>
<td>30.7 (7.0)</td>
<td>7.7 (8.2)</td>
<td>9.4 (9.9)</td>
<td></td>
</tr>
<tr>
<td>Jarrett et al. (1999)</td>
<td>72</td>
<td>39.3</td>
<td>21/51</td>
<td>CBT v. ADM</td>
<td>10</td>
<td>HAMD ≥ 14</td>
<td>25.3 (7.9)</td>
<td>8.1 (6.6)</td>
<td>8.6 (7.8)</td>
<td></td>
</tr>
<tr>
<td>Mohr et al. (2001)</td>
<td>45</td>
<td>43.9</td>
<td>9/35</td>
<td>CBT v. ADM</td>
<td>16</td>
<td>HAMD ≥ 16 and BDI ≥ 16</td>
<td>20.7 (5.3)</td>
<td>13.4 (6.5)</td>
<td>13.5 (9.5)</td>
<td></td>
</tr>
<tr>
<td>Murphy et al. (1984)</td>
<td>33</td>
<td>–</td>
<td>6/27</td>
<td>CBT v. ADM</td>
<td>12</td>
<td>HAMD ≥ 16 and BDI ≥ 16</td>
<td>28.9 (6.2)</td>
<td>6.0 (5.1)</td>
<td>8.1 (8.5)</td>
<td></td>
</tr>
<tr>
<td>Reynolds et al. (1999)</td>
<td>58</td>
<td>67.3</td>
<td>12/46</td>
<td>IPT v. ADM v. IPT + ADM</td>
<td>16</td>
<td>SADS-L and RDC and SCID</td>
<td>12.4 (4.1)</td>
<td>11.6 (8.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rush et al. (1977)</td>
<td>41</td>
<td>–</td>
<td>15/26</td>
<td>CBT v. ADM</td>
<td>12</td>
<td>HAMD ≥ 14 and BDI ≥ 20</td>
<td>30.2 (6.1)</td>
<td>7.3 (5.4)</td>
<td>9.2 (9.8)</td>
<td></td>
</tr>
<tr>
<td>All patients with BDI</td>
<td>472</td>
<td>40.5</td>
<td>117/350</td>
<td></td>
<td></td>
<td></td>
<td>26.2 (8.5)</td>
<td>8.9 (6.9)</td>
<td>9.8 (9.2)</td>
<td></td>
</tr>
<tr>
<td>Using BDI-II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DeRubeis et al. (2005)</td>
<td>180</td>
<td>39.9</td>
<td>75/105</td>
<td>CBT v. ADM</td>
<td>16</td>
<td>HAMD ≥ 20</td>
<td>32.6 (9.4)</td>
<td>8.3 (6.3)</td>
<td>10.3 (10.4)</td>
<td></td>
</tr>
<tr>
<td>Dimidjian et al. (2006)</td>
<td>145</td>
<td>38.9</td>
<td>44/101</td>
<td>CBT v. ADM</td>
<td>16</td>
<td>HAMD ≥ 14 and BDI-II ≥ 20</td>
<td>31.9 (7.4)</td>
<td>7.1 (5.6)</td>
<td>9.9 (10.5)</td>
<td></td>
</tr>
<tr>
<td>Dozois et al. (2009)</td>
<td>48</td>
<td>45.3</td>
<td>12/36</td>
<td>ADM v. CBT + ADM</td>
<td>15</td>
<td>SCID</td>
<td>28.6 (9.9)</td>
<td>7.0 (6.6)</td>
<td>12.6 (11.3)</td>
<td></td>
</tr>
<tr>
<td>Hegel et al. (2010)</td>
<td>144</td>
<td>46.7</td>
<td>47/97</td>
<td>CBT v. ADM</td>
<td>10</td>
<td>HAMD ≥ 8 ≤ 22</td>
<td>21.1 (8.3)</td>
<td>9.0 (6.7)</td>
<td>11.2 (8.3)</td>
<td></td>
</tr>
<tr>
<td>Lesperance et al. (2007)</td>
<td>142</td>
<td>57.9</td>
<td>109/33</td>
<td>ADM v. IPT + ADM</td>
<td>12</td>
<td>HAMD ≥ 20</td>
<td>30.3 (9.1)</td>
<td>11.7 (7.9)</td>
<td>15.4 (11.1)</td>
<td></td>
</tr>
<tr>
<td>McBride et al. (2007)</td>
<td>301</td>
<td>37.4</td>
<td>133/168</td>
<td>CBT v. ADM</td>
<td>24</td>
<td>SCID</td>
<td>31.9 (9.2)</td>
<td>6.6 (4.9)</td>
<td>11.4 (10.7)</td>
<td></td>
</tr>
<tr>
<td>Quilty et al. (2014)</td>
<td>104</td>
<td>33.5</td>
<td>50/54</td>
<td>CBT v. ADM</td>
<td>16</td>
<td>SCID</td>
<td>29.8 (8.6)</td>
<td>7.9 (6.2)</td>
<td>12.0 (10.8)</td>
<td></td>
</tr>
<tr>
<td>All patients with BDI-II</td>
<td>1064</td>
<td>42.0</td>
<td>447/527</td>
<td></td>
<td></td>
<td></td>
<td>30.4 (9.4)</td>
<td>8.5 (6.5)</td>
<td>12.0 (10.6)</td>
<td></td>
</tr>
</tbody>
</table>

ADM, antidepressant medication; BDI, Beck Depression Inventory; CBT, cognitive-behavioural therapy; HAMD, Hamilton Rating Scale for Depression; IPT, interpersonal psychotherapy; RDC, Research Diagnostic Criteria; SADS-L, Schedule for Affective Disorders and Schizophrenia-Lifetime Version; SCID, Structured Interview for DSM.

Standard deviations in parentheses.
was around 40 years, and 61% were women. The treatment lasted between 10 and 24 weeks, typically for 16 weeks. At baseline, participants presented with average HAMD scores around 20, which dropped to scores around 9 at end of treatment. Seven studies used the BDI, which dropped from around 26 to around 10; another seven studies used the BDI-II, which dropped from around 30 to 12, on average.

Correlations between HAMD and BDI, BDI-II

Figure 2 presents the scatterplots between HAMD and BDI or BDI-II at baseline and at end of treatment. The correlations between the HAMD and BDI or BDI-II were relatively weak, with Spearman correlation coefficients of 0.37 and 0.36, respectively; the raw data were scattered relatively widely, and there were few data points with a HAMD score of 10 or below, or 30 or higher. At the end of treatment, the correlations between the HAMD and BDI or BDI-II were stronger ($r = 0.73$ and $0.74$, respectively), with raw data distributed in a more elliptic manner predominantly below a HAMD score of 20. When the baseline observations and end-of-treatment observations were combined, the correlations between the scales rose to 0.77 and 0.76, respectively.

There were moderately strong correlations between change scores: the Spearman correlation coefficients were 0.69 and 0.61 for the HAMD and the BDI or BDI-II change scores, respectively.

Subgroup and sensitivity analyses

Appendix B shows the scatterplots for men and women separately at baseline and at endpoint. Appendix C provides the scatterplots for those who would later drop out and those who would complete the studies separately. The equipercentile linkings were essentially similar across these subgroups, and hence with the overall findings.

Linking HAMD and BDI, BDI-II

Figure 3 depicts the linking curves between HAMD and BDI or BDI-II: 3a in terms of raw scores and 3b in terms of change scores. Table 2 summarises the correspondences on each of these scales. Outside of the ranges displayed and tabled; there were too few data for linking.

Discussion

We have obtained IPD from 14 randomised controlled trials of psychotherapies for the acute phase treatment of depression.
(total \( n = 1536 \) participants), in which the HAMD and the BDI/BDI-II were administered concurrently both at baseline and at end of treatment. The equipercentile linking between the HAMD and the BDI/BDI-II raw scores or change scores established that the HAMD scores of 10, 20 and 30 corresponded approximately with the BDI of 10, 27 and 42 or with the BDI-II of 13, 32 and 50; the HAMD change scores of \(-20\) and \(-10\) with the BDI of \(-29\) and \(-15\) and with the BDI-II of \(-35\) and \(-16\).

It is worthwhile to note that the BDI-II tended to produce higher scores than the original BDI. This was noted originally when the BDI-II was first developed (Beck et al., 1996) and replicated subsequently (Dozois et al., 1998), as the BDI-II dropped or reworded items that poorly reflected depression severity in the original BDI. Our linking analyses correctly reflected this difference between the BDI and the BDI-II.

Possible weaknesses of this study include the following. First, our IPD dataset included only trials that compared psychotherapies against pharmacotherapies or their combinations. Some might suspect that the relationship between the HAMD and the BDI/BDI-II could be different if the data were derived from pharmacotherapy trials. Likewise, the datasets were limited to individuals with major depression who sought treatment. It is possible that linking results could be different among people in the community suffering from major depression but not seeking treatment. However, as this linking study is not about treatments

### Table 2. Conversion from HAMD to BDI or BDI-II scores

<table>
<thead>
<tr>
<th>Total scores</th>
<th>Change scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAMD</td>
<td>BDI</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>13</td>
<td>15</td>
</tr>
<tr>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td>15</td>
<td>17</td>
</tr>
<tr>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>17</td>
<td>21</td>
</tr>
<tr>
<td>18</td>
<td>23</td>
</tr>
<tr>
<td>19</td>
<td>25</td>
</tr>
<tr>
<td>20</td>
<td>27</td>
</tr>
<tr>
<td>21</td>
<td>28</td>
</tr>
<tr>
<td>22</td>
<td>30</td>
</tr>
<tr>
<td>23</td>
<td>31</td>
</tr>
<tr>
<td>24</td>
<td>33</td>
</tr>
<tr>
<td>25</td>
<td>34</td>
</tr>
<tr>
<td>26</td>
<td>36</td>
</tr>
<tr>
<td>27</td>
<td>38</td>
</tr>
<tr>
<td>28</td>
<td>39</td>
</tr>
<tr>
<td>29</td>
<td>41</td>
</tr>
<tr>
<td>30</td>
<td>42</td>
</tr>
</tbody>
</table>

BDI, Beck Depression Inventory; BDI-II, Beck Depression Inventory, 2nd Edition; HAMD, Hamilton Rating Scale for Depression.
but about measurements, we do not foresee any strong reason that there would be major differences. Second, the correlations between the HAMD and the BDI/BDI-II were only moderate at baseline. This is reflected by rounder, rather than elliptic, scatterplots between the HAMD and the BDI/BDI-II at baseline (Fig. 2). We originally suspected that there may have arisen some ‘baseline inflation’ through which people tended to overestimate the depression severity at baseline when a certain threshold on that scale was used as a cutoff for eligibility. Focusing on the four studies that did not use a cutoff (cf Table 1), however, did not improve the correlation coefficients at baseline. A possibility remains that the observed low correlation at baseline is due to range restriction of the available scores on HAMD and BDI/BDI-II as indicated by smaller standard deviations of these scores at baseline than at endpoint. Another possibility is that participants may have been engaging in impression management, either by overreporting or underreporting their symptoms in the self-reports, especially at the start of the trial: as the trial progresses, they may feel less need for such impression management. Third, it must be pointed out that observer- and self-ratings of depression severity do not in general show perfect correlations and that their contrasts can sometimes provide clinically useful information (Petkova et al., 2000; Targum et al., 2013). The conversion algorithm as presented in this study must therefore serve as a rough guide when only one of HAMD/BDI/BDI-II is available and one wishes to know the approximately equivalent scores. Last, the linking above the HAMD scores of 30, where there were few endpoint measurements, may require appropriate caution. Alternatively it may be safer to convert the change scores rather than raw scores when researchers would like to use one common scale across different studies.

On the other hand, the current study also has several major strengths. This is the first study to empirically link the most representative observer-rated instrument and the most frequently used self-rating instrument for depression, based on data from over 1500 participants. The conversion table will help clinicians interpret the HAMD or BDI/BDI-II scores of their patients in a more versatile manner as they can now convert each scale into another. Clinicians will also find it easier to compare their patients’ scores with those reported in the literature when the latter only reports one of these scales while they have only scores from the other scales for their patients. The conversion table will also be informative for researchers when they compare trials using one but not the other of these scales; in particular, the table will allow researchers to convert these scores onto the common scale so that they would need less assumption when they conduct IPD meta-analysis (Furukawa et al., 2018); without the conversion the only way to pool individual data was via standardisation assuming a consistent and common standard deviation (Bower et al., 2013). For the latter purpose one might prefer to use the conversion of the change scores as they showed higher correlations among the scales.

In conclusion, this study provided the first empirically-derived conversion table between the HAMD and the BDI/BDI-II. The table is expected to be of help to both clinicians and researchers.

**Data and materials.** Data used in this study are not available for sharing due to ethical and data management requirements. The researchers are open to collaboration.

**Author ORCIDs.** Mirjam Reijnders, 0000-0002-4272-2576

**Acknowledgement.** None.

**Financial support.** This study was supported in part by a grant-in-aid from Japan Agency for Medical Research and Development (AMED) to TAF (grant number 18dk0307072). The funder had no role in study design, data collection or analysis, decision to publish or preparation of the manuscript.

**Conflict of interest.** TAF has received lecture fees from Meiji, Mitsubishi-Tanabe, MSD and Pfizer. He has received research support from Mitsubishi-Tanabe, Bristol Meyers Squib and Pfizer have provided pharmaceutical supplies for CERF NHI sponsored research. UH was an advisory board member for Lundbeck, Janssen and Servier; a consultant for Bayer Pharma; and a speaker for Servier. RBIs medical centre collects the payments from the cognitive therapy she provides to patients. RB is a paid consultant to the National Institutes of Health and is a paid reviewer for UpToDate. DCM received consulting fees from Apple Inc., Optum Behavioral Health and the One Mind Foundation. He also has an ownership interest in Actualize Therapy. SL has received honoraria for consulting from LB Pharma, Lundbeck, Otsuka, Roche, and TEVA, for lectures from AOP Orphan, ICON, Janssen, Lilly, Lundbeck, Otsuka, Sanofi, Roche, and Servier, and for a publication from Roche. All the other authors declare that they have no conflict of interest.

**Ethical standards.** All procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees about studies on human participants and with the Declaration of Helsinki and its amendments. The investigators of the primary trials have obtained ethical approval for the data used in the current study and for sharing the data, if this was necessary, according to local requirements and was not covered from the initial ethical assessment.

**References**


Appendix
Translating the BDI and BDI-II into the HAMD and vice versa with equipercentile linking

Appendix A. Search string for PubMed

Appendix B. Subgroup analyses by sex

B1. Men

a. HAMD and BDI at baseline

b. HAMD and BDI at endpoint

r = 0.35

r = 0.70

c. and BDI-II at baseline

d. HAMD and BDI-II at endpoint

r = -0.33

r = 0.73

White circles: Raw data, Purple circles: Equipercntile linking
B2. Women

**a. HAMD and BDI at baseline**

![Graph](image1)

$r = 0.38$

**b. HAMD and BDI at endpoint**

![Graph](image2)

$r = 0.75$

**c. HAMD and BDI-II at baseline**

![Graph](image3)

$r = 0.43$

**d. HAMD and BDI-II at endpoint**

![Graph](image4)

$r = 0.75$

White circles: Raw data, Purple circles: Equipercentile linking
Appendix C. Subgroup analyses by completers and dropouts

C1. Dropouts

a. HAMD and BDI at baseline

![Graph showing correlation between HAMD and BDI at baseline.]

\[ r = 0.39 \]

b. HAMD and BDI at endpoint

(For dropouts, there are no measurement of HAMD/BDI at endpoint.)

c. HAMD and BDI-II at baseline

![Graph showing correlation between HAMD and BDI-II at baseline.]

\[ r = 0.38 \]

d. HAMD and BDI-II at endpoint

(For dropouts, there are no measurement of HAMD/BDI-II at endpoint.)

White circles: Raw data, Purple circles: Equipercentile linking
C2. Completers

a. HAMD and BDI at baseline

\[ r = 0.35 \]

b. HAMD and BDI at endpoint

\[ r = 0.75 \]

c. HAMD and BDI-II at baseline

\[ r = 0.39 \]

d. HAMD and BDI-II at endpoint

\[ r = 0.75 \]

White circles: Raw data, Purple circles: Equipercentile linking