Epidemiology and Psychiatric Sciences

cambridge.org/eps

Original Article

Cite this article: Furukawa TA *et al* (2020). Translating the BDI and BDI-II into the HAMD and vice versa with equipercentile linking. *Epidemiology and Psychiatric Sciences* **29**, e24, 1–13. https://doi.org/10.1017/S2045796019000088

Received: 14 October 2018
Revised: 16 January 2019
Accepted: 7 February 2019

Key words:
Assessment; Beck Depression Inventory; depressive disorder; Hamilton Rating Scale for Depression; rating instrument

Author for correspondence:
Toshi A. Furukawa, E-mail: furukawa@kuhp.kyoto-u.ac.jp

CAMBRIDGE UNIVERSITY PRESS

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

Toshi A. Furukawa[1], Mirjam Reijnders[2], Sanae Kishimoto[3], Masatsugu Sakata[3], Robert J. DeRubeis[4], Sona Dimidjian[5], David J.A. Dozois[6], Ulrich Hegerl[7], Steven D. Hollon[8], Robin B. Jarrett[9], François Lespérance[10], Zindel V. Segal[11], David C. Mohr[12], Anne D. Simons[13], Lena C. Quilty[14], Charles F. Reynolds III[15], Claudio Gentili[16], Stefan Leucht[17], Rolf R. Engel[18] and Pim Cuijpers[2]

[1]Departments of Health Promotion and Human Behavior and of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan; [2]Department of Clinical, Neuro-, and Developmental Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, The Netherlands; [3]Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan; [4]Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA; [5]Department of Psychology and Neuroscience, University of Colorado, Boulder, USA; [6]Department of Psychology, University of Western Ontario, Westminster Hall, London, Ontario N6A 3K7, Canada; [7]Department of Psychiatry and Psychotherapy, University of Leipzig, Leipzig, Germany; [8]Department of Psychology, Vanderbilt University, Nashville, TN, USA; [9]Department of Psychiatry, The University of Texas Southwestern Medical Center, Dallas, TX, USA; [10]Department of Psychiatry and Addiction, Université de Montréal, Montréal, Québec, Canada; [11]Department of Psychology, University of Toronto – Scarborough, Toronto, Canada; [12]Center for Behavioral Intervention Technologies, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA; [13]Department of Psychology, University of Oregon, Eugene, OR, USA; [14]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health; Department of Psychiatry, University of Toronto, Toronto, Canada; [15]Department of Psychiatry, University of Pittsburgh Medical Center, Western Psychiatric Institute and Clinic, Pittsburgh, USA; [16]Department of General Psychology, University of Padova, Padova, Italy; [17]Department of Psychiatry and Psychotherapy, Technische Universität München, Kkinikum rechts der Isar, Germany and [18]Department of Psychiatry and Psychotherapy, Ludwig-Maximillians Universität München, Germany

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 of −20 and −10 with the BDI of −29 and −15 and with the BDI-II of −35 and −16.

Conclusions. 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 reported 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.*, 2013*a*) 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, 2013*b*, 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 (Cuijpers *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 consists of nine symptoms (depressed mood, self-depreciation 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 α ranging from 0.76 to 0.95) and non-psychiatric populations (Cronbach's α 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; Hollon *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

https://doi.org/10.1017/S2045796019000088 Published online by Cambridge University Press

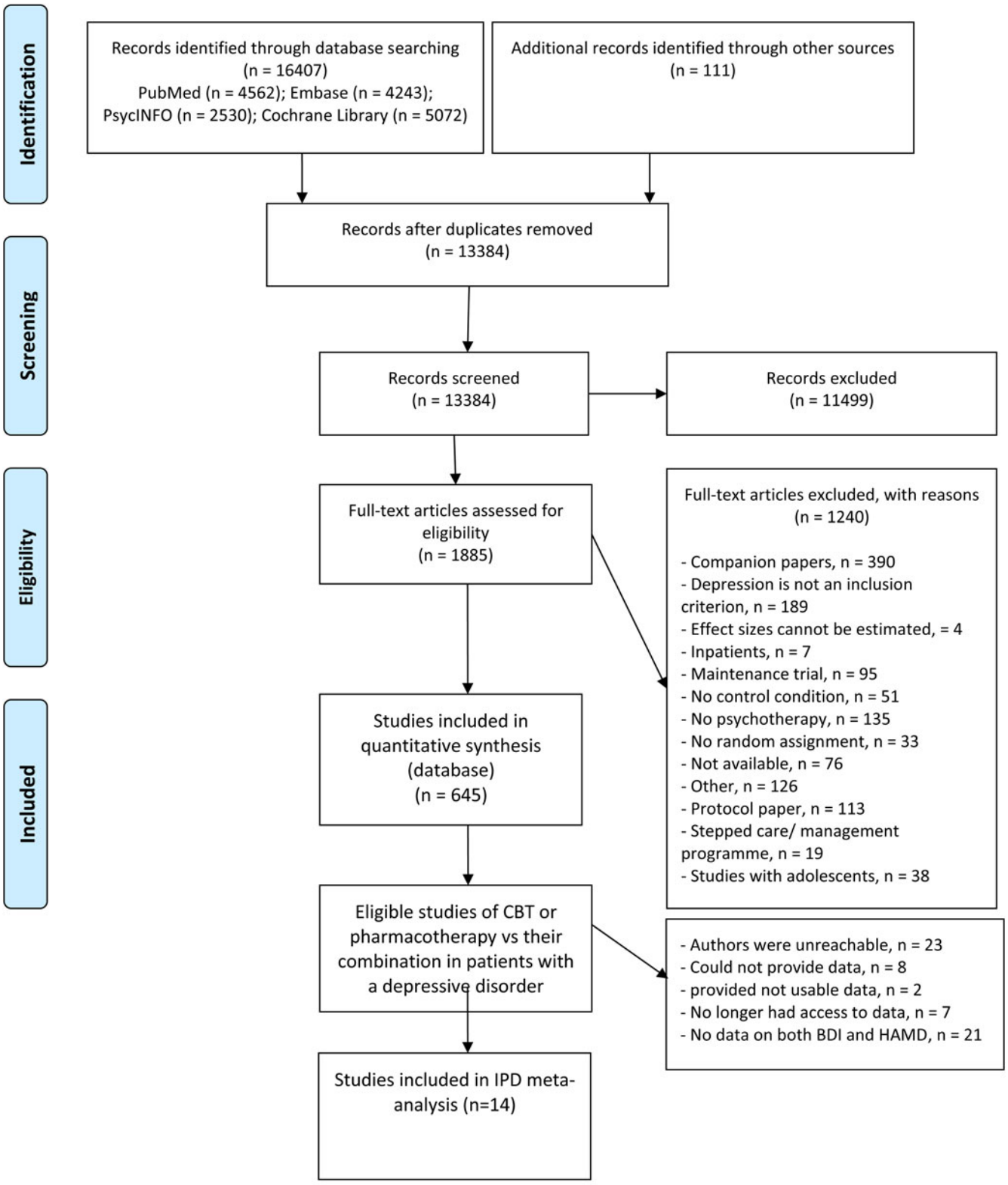


Fig. 1. Flowchart of study identification.

https://doi.org/10.1017/S2045796019000088 Published online by Cambridge University Press

Table 1. Included studies and their characteristics

Studies							Baseline		Endpoint	
Using BDI	*N*	Age	Sex (male/female)	Treatments	Treatment duration (weeks)	Scale used for inclusion	HAMD	BDI	HAMD	BDI
Elkin *et al.* (1989)	116	34.5	35/81	CBT *v.* ADM	16	HAMD ⩾ 14	19.5 (4.2)	26.7 (8.5)	7.8 (6.8)	9.0 (9.7)
Hollon *et al.* (1992)	107	32.6	21/86	CBT *v.* ADM *v.* CBT + ADM	12	HAMD ⩾ 14 and BDI ⩾ 20	23.9 (4.9)	30.7 (7.0)	7.7 (8.2)	9.4 (9.9)
Jarrett *et al.* (1999)	72	39.3	21/51	CBT *v.* ADM	10	HAMD ⩾ 14	16.6 (3.1)	25.3 (7.9)	8.1 (6.6)	8.6 (7.8)
Mohr *et al.* (2001)	45	43.9	9/35	CBT *v.* ADM	16	HAMD ⩾ 16 and BDI ⩾ 16	19.5 (3.8)	20.7 (5.3)	13.4 (6.5)	13.5 (9.5)
Murphy *et al.* (1984)	33	–	6/27	CBT *v.* ADM	12	HAMD ⩾ 14 and BDI ⩾ 20	18.9 (2.9)	28.9 (6.2)	6.0 (5.1)	8.1 (8.5)
Reynolds *et al.* (1999)	58	67.3	12/46	IPT *v.* ADM *v.* IPT + ADM	16	SADS-L and RDC and SCID	19.7 (4.1)	17.7 (8.0)	12.4 (4.1)	11.6 (8.1)
Rush *et al.* (1977)	41	–	15/26	CBT *v.* ADM	12	HAMD ⩾ 14 and BDI ⩾ 20	21.3 (4.0)	30.2 (6.1)	7.3 (5.4)	9.2 (9.8)
All patients with BDI	*472*	*40.5*	*117/350*				*20.2 (4.7)*	*26.2 (8.5)*	*8.9 (6.9)*	*9.8 (9.2)*
Using BDI-II							HAMD	BDI-II	HAMD	BDI-II
DeRubeis *et al.* (2005)	180	39.9	75/105	CBT *v.* ADM	16	HAMD ⩾ 20	21.5 (4.0)	32.6 (9.4)	8.3 (6.3)	10.3 (10.4)
Dimidjian *et al.* (2006)	145	38.9	44/101	CBT *v.* ADM	16	HAMD ⩾ 14 and BDI-II ⩾ 20	18.5 (4.1)	31.9 (7.4)	7.1 (5.6)	9.9 (10.5)
Dozois *et al.* (2009)	48	45.3	12/36	ADM *v.* CBT + ADM	15	SCID	18.0 (4.0)	28.6 (9.9)	7.0 (6.6)	12.6 (11.3)
Hegerl *et al.* (2010)	144	46.7	47/97	CBT *v.* ADM	10	HAMD ⩾ 8 ⩽ 22	16.5 (4.3)	21.1 (8.3)	9.0 (6.7)	11.2 (8.3)
Lesperance *et al.* (2007)	142	57.9	109/33	ADM *v.* IPT + ADM	12	HAMD ⩾ 20	22.3 (4.9)	30.3 (9.1)	11.7 (7.9)	15.4 (11.1)
McBride *et al.* (2007)	301	37.4	133/168	CBT *v.* ADM	24	SCID	19.3 (3.7)	31.9 (9.2)	6.6 (4.9)	11.4 (10.7)
Quilty *et al.* (2014)	104	33.5	50/54	CBT *v.* ADM	16	SCID	16.6 (5.1)	29.8 (8.6)	7.9 (6.2)	12.0 (10.8)
All patients with BDI-II	*1064*	*42.0*	*447/527*				*19.3 (4.7)*	*30.4 (9.4)*	*8.5 (6.5)*	*12.0 (10.6)*

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.

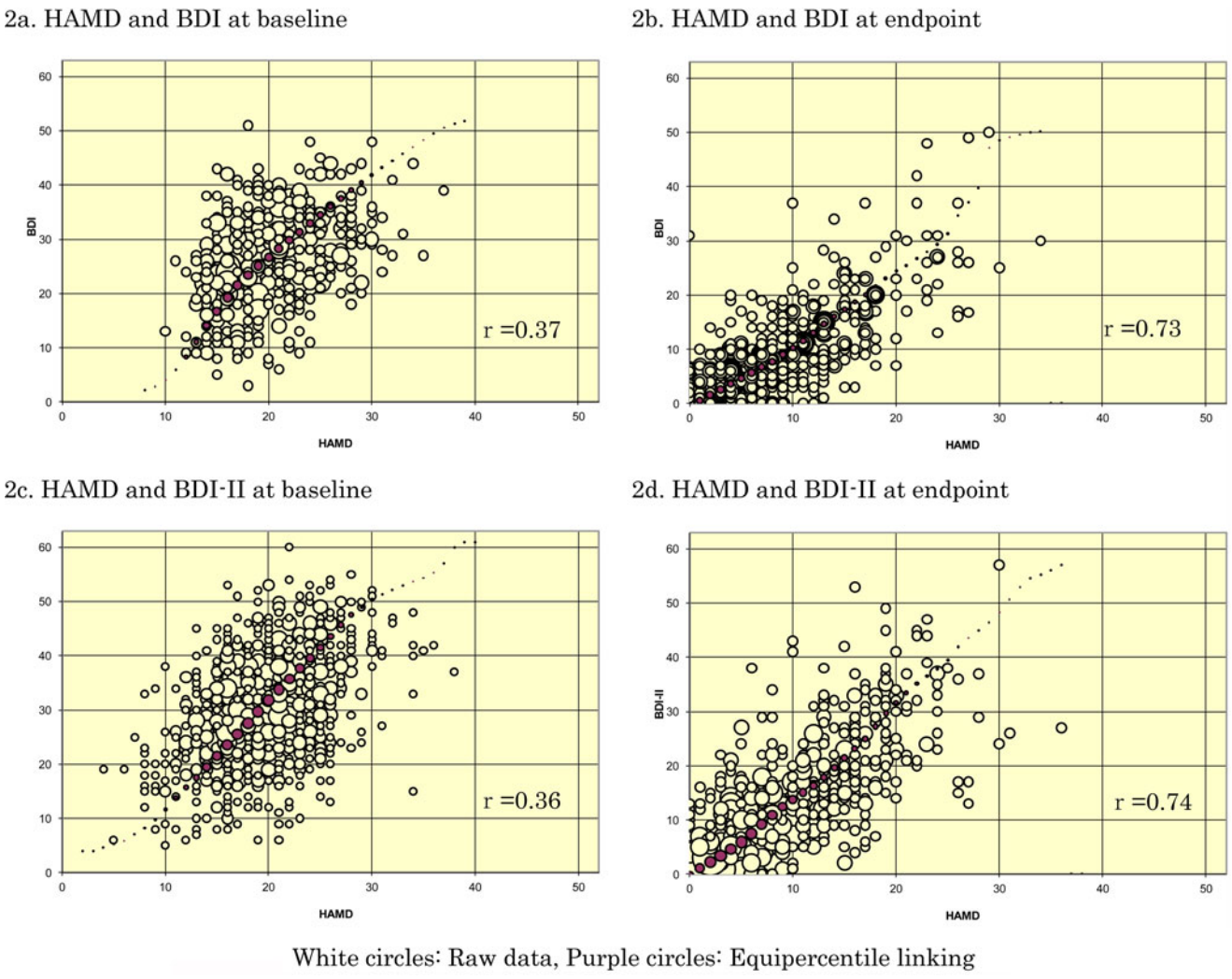


Fig. 2. Scatterplots of HAMD and BDI, BDI-II, superimposed with equipercentile linking.

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

https://doi.org/10.1017/S2045796019000088 Published online by Cambridge University Press

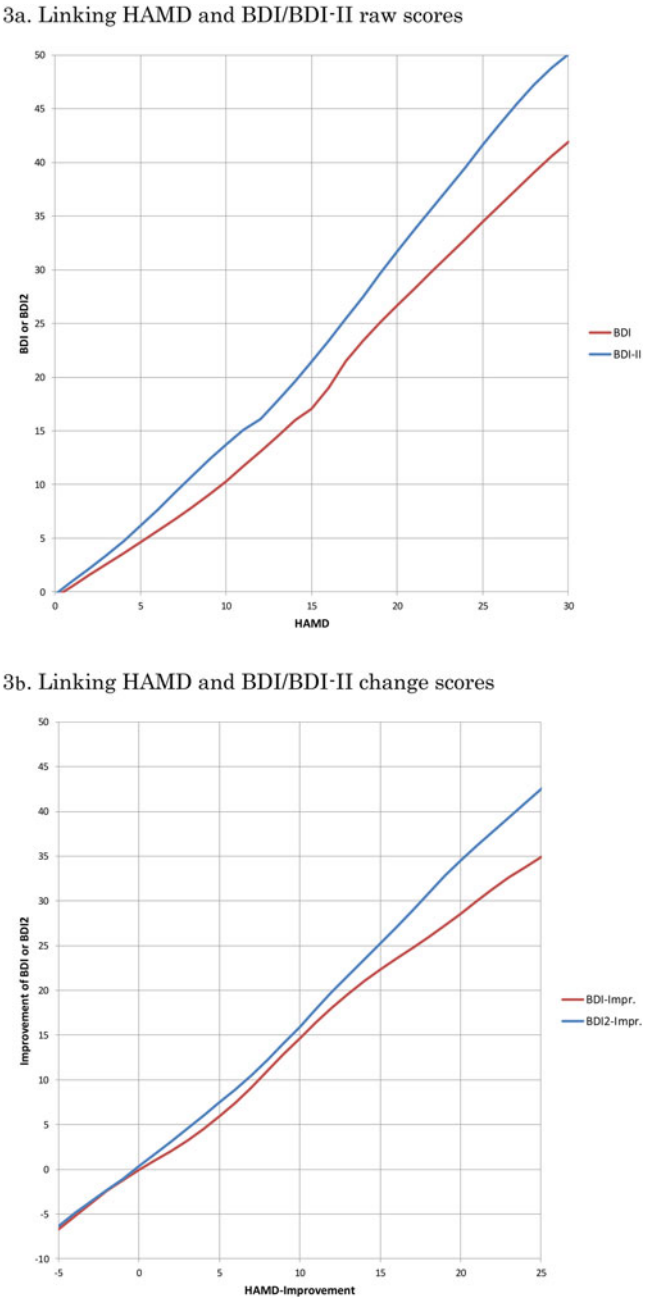


Fig. 3. Linking curves between HAMD and BDI, BDI-II.

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

Total scores			Change scores		
HAMD	BDI	BDI-II	HAMD	BDI	BDI-II
0	0	0	−5	−7	−6
1	1	1	−4	−5	−5
2	2	2	−3	−4	−4
3	3	3	−2	−2	−2
4	4	5	−1	−1	−1
5	5	6	0	0	0
6	6	8	1	1	2
7	7	9	2	2	3
8	8	11	3	3	5
9	9	12	4	4	6
10	10	13	5	6	7
11	12	15	6	7	9
12	13	16	7	9	11
13	15	18	8	11	12
14	16	20	9	13	14
15	17	21	10	15	16
16	19	23	11	16	18
17	21	25	12	18	20
18	23	27	13	20	22
19	25	30	14	21	23
20	27	32	15	22	25
21	28	34	16	24	27
22	30	36	17	25	29
23	31	38	18	26	31
24	33	40	19	27	33
25	34	42	20	29	35
26	36	44	21	30	36
27	38	45	22	31	38
28	39	47	23	33	39
29	41	49	24	34	41
30	42	50	25	35	43

BDI, Beck Depression Inventory; BDI-II, Beck Depression Inventory, 2nd Edition; HAMD, Hamilton Rating Scale for 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

https://doi.org/10.1017/S2045796019000088 Published online by Cambridge University Press

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 criterion 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 scales 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 CFRs NIH sponsored research. UH was an advisory board member for Lundbeck, Janssen and Servier; a consultant for Bayer Pharma; and a speaker for Servier. RBJs medical centre collects the payments from the cognitive therapy she provides to patients. RBJ 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 ethic assessment.

References

Barth J, Munder T, Gerger H, Nuesch E, Trelle S, Znoj H, Juni P and Cuijpers P (2013) Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. *PLoS Medicine* **10**, e1001454.

Beck AT, Ward CH, Mendelson M, Mock J and Erbaugh J (1961) An inventory for measuring depression. *Archives of General Psychiatry* **4**, 561–571.

Beck AT, Steer RA and Garbin MG (1988) Psychometric properties of the Beck Depression Inventory: twenty-five years of evaluation. *Clinical Psychology Review* **8**, 77–100.

Beck AT, Steer RA and Brown GK (1996) *BDI-II: Beck Depression Inventory, Second Edition, Manual*. San Antonia: The Psychological Corporation

Bower P, Kontopantelis E, Sutton A, Kendrick T, Richards DA, Gilbody S, Knowles S, Cuijpers P, Andersson G, Christensen H, Meyer B, Huibers M, Smit F, van Straten A, Warmerdam L, Barkham M, Bilich L, Lovell K and Liu ET (2013) Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. *BMJ (Clinical Research Ed.)* **346**, f540.

Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, Leucht S, Ruhe HG, Turner EH, Higgins JPT, Egger M, Takeshima N, Hayasaka Y, Imai H, Shinohara K, Tajika A, Ioannidis JPA and Geddes JR (2018) Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet* **391**, 1357–1366.

Cuijpers P, van Straten A, Warmerdam L and Andersson G (2008) Psychological treatment of depression: a meta-analytic database of randomized studies. *BMC Psychiatry* **8**, 36.

DeRubeis RJ, Hollon SD, Amsterdam JD, Shelton RC, Young PR, Salomon RM, O'Reardon JP, Lovett ML, Gladis MM, Brown LL and Gallop R (2005) Cognitive therapy vs medications in the treatment of moderate to severe depression. *Archives of General Psychiatry* **62**, 409–416.

Dimidjian S, Hollon SD, Dobson KS, Schmaling KB, Kohlenberg RJ, Addis ME, Gallop R, McGlinchey JB, Markley DK, Gollan JK, Atkins DC, Dunner DL and Jacobson NS (2006) Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology* **74**, 658–670.

Dozois DJ, Dobson KS and Ahnberg JL (1998) A psychometric evaluation of the Beck Depression Inventory-II. *Psychological Assessment* **10**, 83–89.

https://doi.org/10.1017/S2045796019000088 Published online by Cambridge University Press

Dozois DJ, Bieling PJ, Patelis-Siotis I, Hoar L, Chudzik S, McCabe K and Westra HA (2009) Changes in self-schema structure in cognitive therapy for major depressive disorder: a randomized clinical trial. *Journal of Consulting and Clinical Psychology* **77**, 1078–1088.

Elkin I, Shea MT, Watkins JT, Imber SD, Sotsky SM, Collins JF, Glass DR, Pilkonis PA, Leber WR, Docherty JP, Fiester SJ and Parloff MB. (1989) National Institute of Mental Health treatment of depression collaborative research program. General effectiveness of treatments. *Archives of General Psychiatry* **46**, 971–982, discussion 983.

Furukawa TA, Shear MK, Barlow DH, Gorman JM, Woods SW, Money R, Etschel E, Engel RR and Leucht S (2009) Evidence-based guidelines for interpretation of the Panic Disorder Severity Scale. *Depression and Anxiety* **26**, 922–929.

Furukawa TA, Weitz ES, Tanaka S, Hollon SD, Hofmann SG, Andersson G, Twisk J, DeRubeis RJ, Dimidjian S, Hegerl U, Mergl R, Jarrett RB, Vittengl JR, Watanabe N and Cuijpers P (2017) Initial severity of depression and efficacy of cognitive-behavioural therapy: individual-participant data meta-analysis of pill-placebo-controlled trials. *British Journal of Psychiatry* **210**, 190–196.

Furukawa TA, Efthimiou O, Weitz ES, Cipriani A, Keller MB, Kocsis JH, Klein DN, Michalak J, Salanti G, Cuijpers P and Schramm E (2018) Cognitive-Behavioral Analysis System of Psychotherapy, drug, or their combination for persistent depressive disorder: personalizing the treatment choice using individual participant data network metaregression. *Psychotherapy and Psychosomatics* **87**, 140–153.

Hamilton M (1960) A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry* **23**, 56–62.

Hegerl U, Hautzinger M, Mergl R, Kohnen R, Schutze M, Scheunemann W, Allgaier AK, Coyne J and Henkel V (2010) Effects of pharmacotherapy and psychotherapy in depressed primary-care patients: a randomized, controlled trial including a patients' choice arm. *The International Journal of Neuropsychopharmacology* **13**, 31–44.

Hollon SD, DeRubeis RJ, Evans MD, Wiemer MJ, Garvey MJ, Grove WM and Tuason VB (1992) Cognitive therapy and pharmacotherapy for depression. Singly and in combination. *Archives of General Psychiatry* **49**, 774–781.

Jarrett RB, Schaffer M, McIntire D, Witt-Browder A, Kraft D and Risser RC (1999) Treatment of atypical depression with cognitive therapy or phenelzine: a double-blind, placebo-controlled trial. *Archives of General Psychiatry* **56**, 431–437.

Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, Mackinnon A, Meyer B, Botella C, Littlewood E, Andersson G, Christensen H, Klein JP, Schroder J, Breton-Lopez J, Scheider J, Griffiths K, Farrer L, Huibers MJ, Phillips R, Gilbody S, Moritz S, Berger T, Pop V, Spek V and Cuijpers P (2017) Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. *JAMA Psychiatry* **74**, 351–359.

Kolen MJ and Brennan RL (1995) *Test Equating: Methods and Practices*. New York: Springer.

Kroenke K, Spitzer RL and Williams JB (2001) The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine* **16**, 606–613.

Lesperance F, Frasure-Smith N, Koszycki D, Laliberte MA, van Zyl LT, Baker B, Swenson JR, Ghatavi K, Abramson BL, Dorian P, Guertin MC and Investigators C (2007) Effects of citalopram and interpersonal psychotherapy on depression in patients with coronary artery disease: the Canadian cardiac randomized evaluation of antidepressant and psychotherapy efficacy (CREATE) trial. *JAMA* **297**, 367–379.

Leucht S, Kane JM, Kissling W, Hamann J, Etschel E and Engel RR (2005) What does the PANSS mean? *Schizophrenia Research* **79**, 231–238.

Leucht S, Kane JM, Etschel E, Kissling W, Hamann J and Engel RR (2006) Linking the PANSS, BPRS, and CGI: clinical implications. *Neuropsychopharmacology* **31**, 2318–2325.

Leucht S, Fennema H, Engel R, Kaspers-Janssen M, Lepping P and Szegedi A (2013a) What does the HAMD mean? *Journal of Affective Disorders* **148**, 243–248.

Leucht S, Rothe P, Davis JM and Engel RR (2013b) Equipercentile linking of the BPRS and the PANSS. *European Neuropsychopharmacology* **23**, 956–959.

Leucht S, Fennema H, Engel RR, Kaspers-Janssen M, Lepping P and Szegedi A (2016) What does MADRS mean? Equipercentile linking with the CGI using a company database of mirtazapine studies. *Journal of Affective Disorders* **210**, 287–293.

Leucht S, Fennema H, Engel RR, Kaspers-Janssen M and Szegedi A (2018) Translating the HAM-D into the MADRS and vice versa with equipercentile linking. *Journal of Affective Disorders* **226**, 326–331.

Levine SZ and Leucht S (2013) Identifying clinically meaningful symptom response cut-off values on the SANS in predominant negative symptoms. *Schizophrenia Research* **145**, 125–127.

Linn RL (1993) Linking results of distinct assessments. *Applied Measurements in Education* **6**, 83–102.

McBride C, Segal Z, Kennedy S and Gemar M (2007) Changes in autobiographical memory specificity following cognitive behavior therapy and pharmacotherapy for major depression. *Psychopathology* **40**, 147–152.

Mohr DC, Boudewyn AC, Goodkin DE, Bostrom A and Epstein L (2001) Comparative outcomes for individual cognitive-behavior therapy, supportive-expressive group psychotherapy, and sertraline for the treatment of depression in multiple sclerosis. *Journal of Consulting and Clinical Psychology* **69**, 942–949.

Montgomery SA and Asberg M (1979) A new depression scale designed to be sensitive to change. *British Journal of Psychiatry* **134**, 382–389.

Murphy GE, Simons AD, Wetzel RD and Lustman PJ (1984) Cognitive therapy and pharmacotherapy. Singly and together in the treatment of depression. *Archives of General Psychiatry* **41**, 33–41.

Petkova E, Quitkin FM, McGrath PJ, Stewart JW and Klein DF (2000) A method to quantify rater bias in antidepressant trials. *Neuropsychopharmacology* **22**, 559–565.

Quilty LC, Dozois DJ, Lobo DS, Ravindran LN and Bagby RM (2014) Cognitive structure and processing during cognitive behavioral therapy vs. pharmacotherapy for depression. *International Journal of Cognitive Therapy* **7**, 235–250.

Reynolds 3rd CF, Miller MD, Pasternak RE, Frank E, Perel JM, Cornes C, Houck PR, Mazumdar S, Dew MA and Kupfer DJ (1999) Treatment of bereavement-related major depressive episodes in later life: a controlled study of acute and continuation treatment with nortriptyline and interpersonal psychotherapy. *American Journal of Psychiatry* **156**, 202–208.

Rush AJ, Beck AT, Kovacs M and Hollon S (1977) Comparative efficacy of cognitive therapy and pharmacotherapy in the treatment of depressed outpatients. *Cognitive Therapy and Research* **1**, 17–37.

Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, Markowitz JC, Ninan PT, Kornstein S, Manber R, Thase ME, Kocsis JH and Keller MB (2003) 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* **54**, 573–583.

Samara MT, Engel RR, Millier A, Kandenwein J, Toumi M and Leucht S (2014) Equipercentile linking of scales measuring functioning and symptoms: examining the GAF, SOFAS, CGI-S, and PANSS. *European Neuropsychopharmacology* **24**, 1767–1772.

Targum SD, Wedel PC, Robinson J, Daniel DG, Busner J, Bleicher LS, Rauh P and Barlow C (2013) A comparative analysis between site-based and centralized ratings and patient self-ratings in a clinical trial of major depressive disorder. *Journal of Psychiatric Research* **47**, 944–954.

Trajkovic G, Starcevic V, Latas M, Lestarevic M, Ille T, Bukumiric Z and Marinkovic J (2011) Reliability of the Hamilton Rating Scale for Depression: a meta-analysis over a period of 49 years. *Psychiatry Research* **189**, 1–9.

Weitz E, Kleiboer A, van Straten A, Hollon SD and Cuijpers P (2017) Individual patient data meta-analysis of combined treatments versus psychotherapy (with or without pill placebo), pharmacotherapy or pill placebo for adult depression: a protocol. *BMJ Open* **7**, e013478.

Williams JB, Kobak KA, Bech P, Engelhardt N, Evans K, Lipsitz J, Olin J, Pearson J and Kalali A (2008) The GRID-HAMD: standardization of the Hamilton Depression Rating Scale. *International Clinical Psychopharmacology* **23**, 120–129.

https://doi.org/10.1017/S2045796019000088 Published online by Cambridge University Press

Appendix

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

Appendix A. Search string for PubMed

(Psychotherapy [MH] OR psychotherap*[All Fields] OR cbt[All Fields] OR "behavior therapies"[All Fields] OR "behavior therapy"[All Fields] OR "behavior therapeutic"[All Fields] OR "behavior therapeutical"[All Fields] OR "behavior therapeutics"[All Fields] OR "behavior therapeutist"[all Fields] OR "behavior therapeutists"[All Fields] OR "behavior treatment"[All Fields] OR "behavior treatments"[All Fields] OR "behaviors therapies"[All Fields] OR "behaviors therapy"[All Fields] OR "behaviors therapeutics"[All Fields] OR "behaviors therapeutic"[All Fields] OR "behaviors therapeutical"[All Fields] OR "behaviors therapeutist"[All Fields] OR "behaviors therapeutists"[All Fields] OR "behaviors treatment"[All Fields] OR "behaviors treatments"[All Fields] OR "behavioral therapies"[All Fields] OR "behavioral therapy"[All Fields] OR "behavioral therapeutics"[All Fields] OR "behavioral therapeutic"[All Fields] OR "behavioral therapeutical"[All Fields] OR "behavioral therapeutist"[All Fields] OR "behavioral therapeutists"[All Fields] OR "behavioral treatment"[All Fields] OR "behavioral treatments"[All Fields] OR "behaviour therapies"[All Fields] OR "behaviour therapy"[All Fields] OR "behaviour therapeutic"[All Fields] OR "behaviour therapeutical"[All Fields] OR "behaviour therapeutics"[All Fields] OR "behaviour therapeutist"[all Fields] OR "behaviour therapeutists"[All Fields] OR "behaviour treatment"[All Fields] OR "behaviour treatments"[All Fields] OR "behaviours therapies"[All Fields] OR "behaviours therapy"[All Fields] OR "behaviours therapeutics"[All Fields] OR "behaviours therapeutic"[All Fields] OR "behaviours therapeutical"[All Fields] OR "behaviours therapeutist"[All Fields] OR "behaviours therapeutists"[All Fields] OR "behaviours treatment"[All Fields] OR "behaviours treatments"[All Fields] OR "behavioural therapies" [All Fields] OR "behavioural therapy"[All Fields] OR "behavioural therapeutics"[All Fields] OR "behavioural therapeutic"[All Fields] OR "behavioural therapeutical"[All Fields] OR "behavioural therapeutist"[All Fields] OR "behavioural therapeutists"[All Fields] OR "behavioural treatment"[All Fields] OR "behavioural treatments"[All Fields] OR "cognition therapies"[All Fields] OR "cognition therapie"[All Fields] OR "cognition therapy"[All Fields] OR "cognition therapeutical"[All Fields] OR "cognition therapeutic"[All Fields] OR "cognition therapeutics"[All Fields] OR "cognition therapeutist"[All Fields] OR "cognition therapeutists"[All Fields] OR "cognition treatment"[All Fields] OR "cognition treatments"[All Fields] OR psychodynamic[All Fields] OR Psychoanalysis[MH] OR psychoanalysis[All Fields] OR psychoanalytic*[All Fields] OR counselling[All Fields] OR counseling [All Fields] OR Counseling[MH] OR "problem-solving"[All Fields] OR mindfulness[All Fields] OR (acceptance[All Fields] AND commitment[All Fields]) OR "assertiveness training"[All Fields] OR "behavior activation"[All Fields] OR "behaviors activation"[All Fields] OR "behavioral activation"[All Fields] OR "cognitive therapies"[All Fields] OR "cognitive therapy"[All Fields] OR "cognitive therapeutic"[All Fields] OR "cognitive therapeutics"[All Fields] OR "cognitive therapeutical"[All Fields] OR "cognitive therapeutist"[All Fields] OR "cognitive therapeutists"[All Fields] OR "cognitive treatment"[All Fields] OR "cognitive treatments"[All Fields] OR "cognitive restructuring"[All Fields] OR (("compassion-focused"[All Fields] OR "compassion-focussed"[All Fields]) AND (therapy[SH] OR therapies[All Fields] OR therapy[All Fields] OR therape*[All Fields] OR therapis*[All Fields]OR Therapeutics [OR treatment*[All Fields])) OR ((therapy[SH] OR therapies[All Fields]

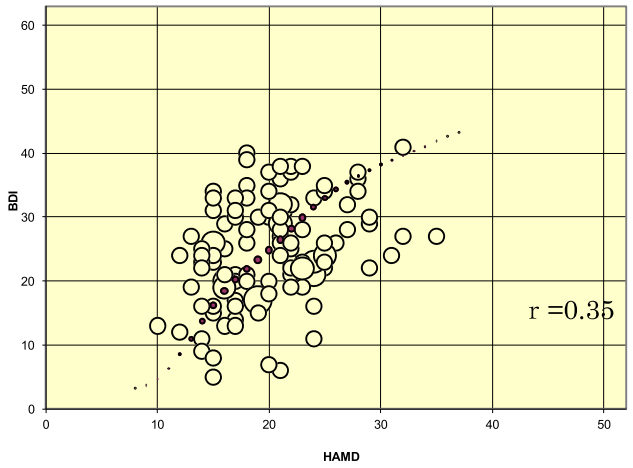
OR therapy [All Fields] OR therape*[All Fields] OR therapis*[All Fields] OR Therapeutics[MH] OR treatment*[All Fields]) AND constructivist*[All Fields]) OR "metacognitive therapies"[All Fields] OR "metacognitive therapy"[All Fields] OR "metacognitive therapeutic"[All Fields] OR "metacognitive therapeutics"[All Fields] OR "metacognitive therapeutical"[All Fields] OR "metacognitive therapeutist"[All Fields] OR "metacognitive therapeutists"[All Fields] OR "metacognitive treatment"[All Fields] OR "metacognitive treatments"[All Fields] OR "meta-cognitive therapies"[All Fields] OR "meta-cognitive therapy"[All Fields] OR "meta-cognitive therapeutic"[All Fields] OR "meta-cognitive therapeutics"[All Fields] OR "meta-cognitive therapeutical"[All Fields] OR "meta-cognitive therapeutist"[All Fields] OR "meta-cognitive therapeutists"[All Fields] OR "meta-cognitive treatment"[All Fields] OR "meta-cognitive treatments"[All Fields] OR "solution-focused therapies"[All Fields] OR "solution-focused therapy"[All Fields] OR "solution-focused therapeutic"[All Fields] OR "solution-focused therapeutics" [All Fields] OR "solution-focused therapeutical"[All Fields] OR "solution focused therapies"[All Fields] OR "solution focused therapy"[All Fields] OR "solution focused therapeutic"[All Fields] OR "solution focused therapeutics"[All Fields] OR "solution focused therapeutical"[All Fields]OR "solution-focussed therapies"[All Fields] OR "solution-focussed therapy"[All Fields] OR "solution-focussed therapeutic"[All Fields] OR "solution-focussed therapeutics"[All Fields] OR "solution-focussed therapeutical"[All Fields]OR "solution focussed therapies"[All Fields] OR "solution focussed therapy"[All Fields] OR "solution focussed therapeutic"[All Fields] OR "solution focussed therapeutics"[All Fields] OR "solution focussed therapeutical"[All Fields] OR "self-control therapies"[All Fields] OR "self-control therapy"[All Fields] OR "self-control therapeutics"[All Fields] OR "self-control therapeutical"[All Fields] OR "self-control therapeutic"[All Fields] OR "self-control training"[All Fields] OR "self-control trainings"[All Fields] OR "self control therapies"[All Fields] OR "self control therapy"[All Fields] OR "self control therapeutics"[All Fields] OR "self control therapeutical"[All Fields] OR "self control therapeutic"[All Fields] OR "self control training"[All Fields] OR "self control trainings"[All Fields] AND (Depressive Disorder[MH] OR Depression[MH]OR dysthymi*[All Fields] OR "affective disorder"[All Fields]OR "affective disorders"[All Fields] OR "mood disorder"[All Fields] OR "mood disorders"[All Fields] OR depression*[All Fields] OR depressive* [All Fields] OR "dysthymic disorder"[MeSH Terms]) AND ((randomized controlled trial [pt] OR controlled clinical trial [pt] OR randomized [tiab] OR randomly [tiab] NOT (animals[mh] NOT (animals[mh] AND humans [mh]))

https://doi.org/10.1017/S2045796019000088 Published online by Cambridge University Press

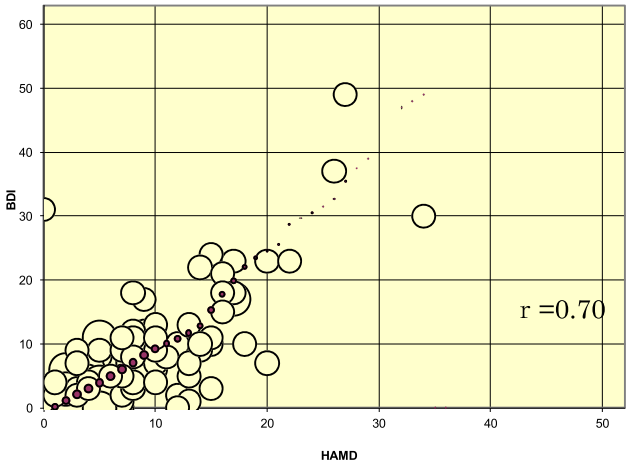
Appendix B. Subgroup analyses by sex

B1. Men

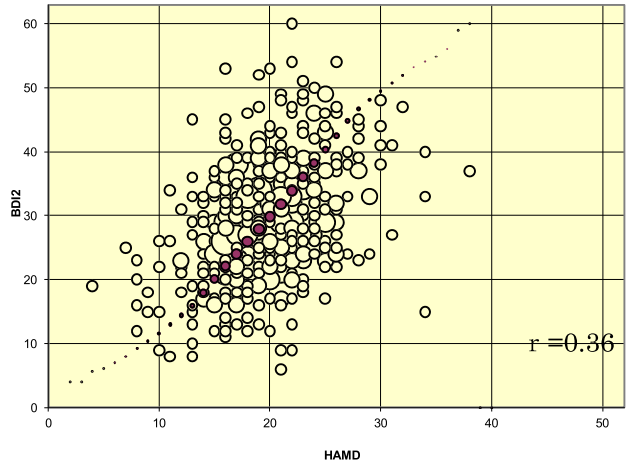
a. HAMD and BDI at baseline



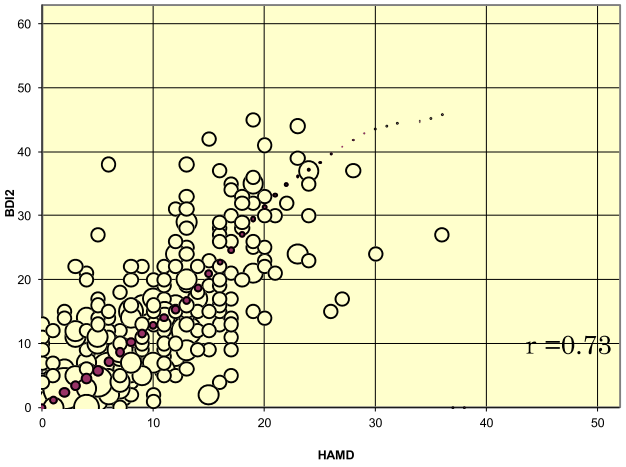
b. HAMD and BDI at endpoint



c. and BDI-II at baseline



d. HAMD and BDI-II at endpoint

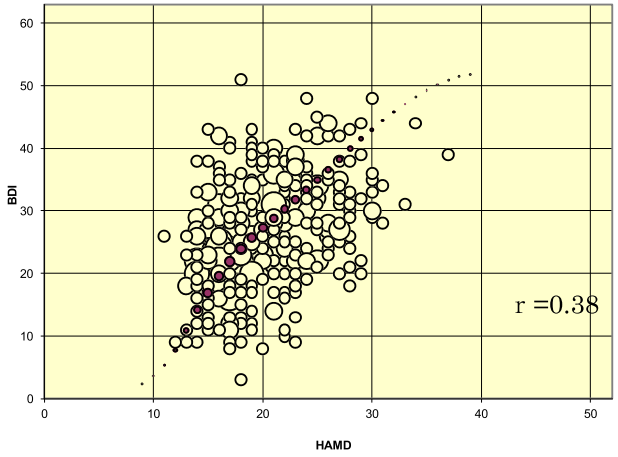


White circles: Raw data, Purple circles: Equipercentile linking

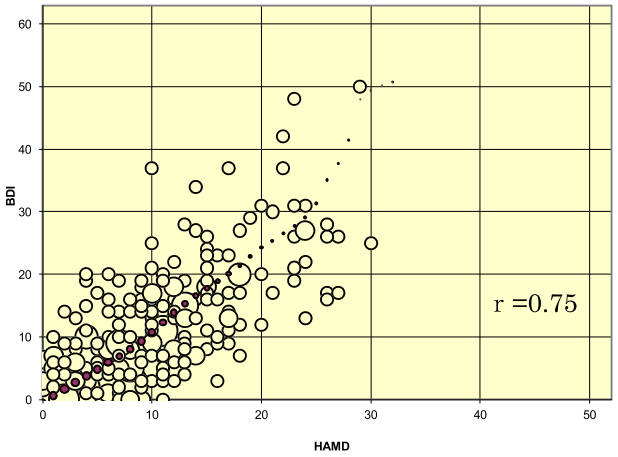
https://doi.org/10.1017/S2045796019000088 Published online by Cambridge University Press

B2. Women

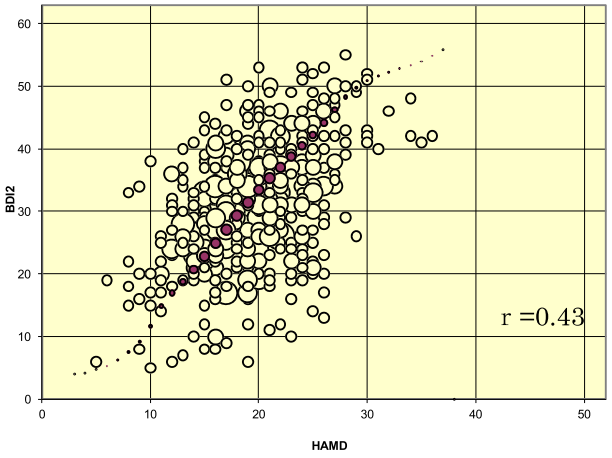
a. HAMD and BDI at baseline



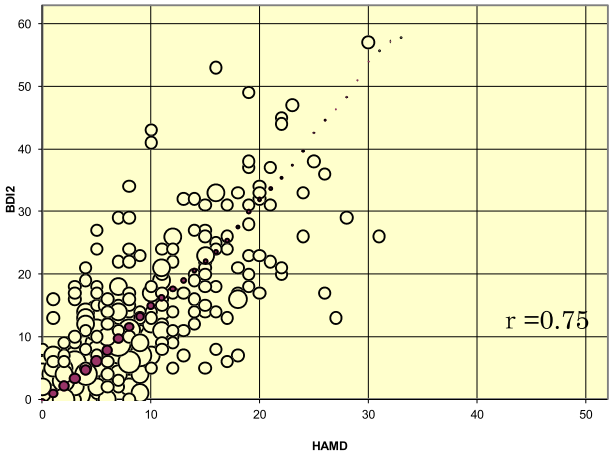
b. HAMD and BDI at endpoint



c. HAMD and BDI-II at baseline



d. HAMD and BDI-II at endpoint



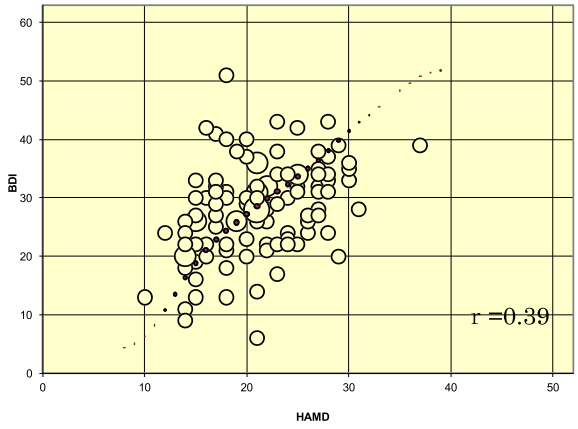
White circles: Raw data, Purple circles: Equipercentile linking

https://doi.org/10.1017/S2045796019000088 Published online by Cambridge University Press

Appendix C. Subgroup analyses by completers and dropouts

C1. Dropouts

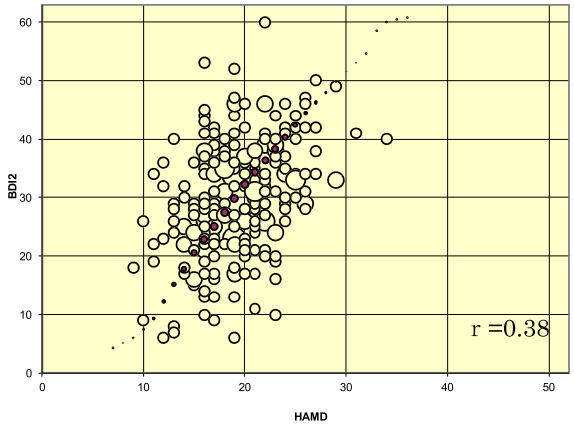
a. HAMD and BDI at baseline



b. HAMD and BDI at endpoint

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

c. HAMD and BDI-II at baseline



d. HAMD and BDI-II at endpoint

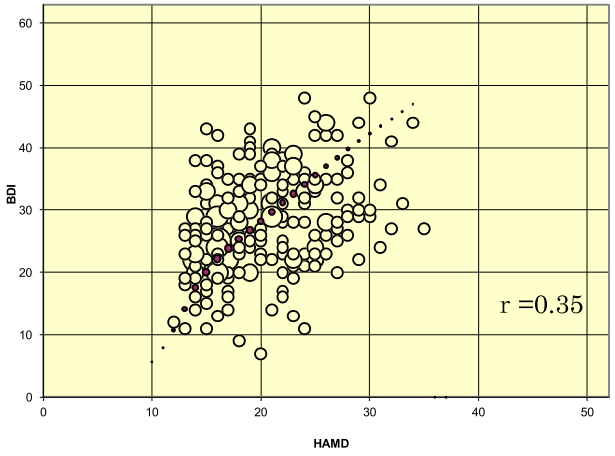
(For dropouts, there are no measurement of HAMD/BDI-II at endpoint.)

White circles: Raw data, Purple circles: Equipercentile linking

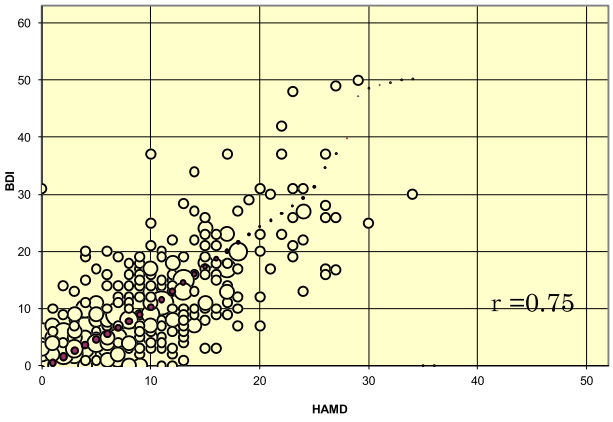
https://doi.org/10.1017/S2045796019000088 Published online by Cambridge University Press

C2. Completers

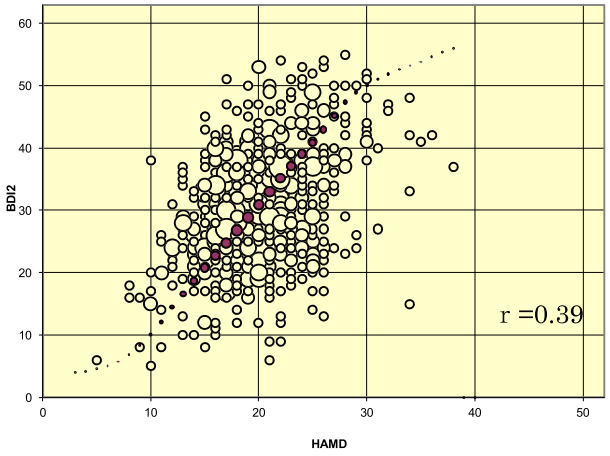
a. HAMD and BDI at baseline



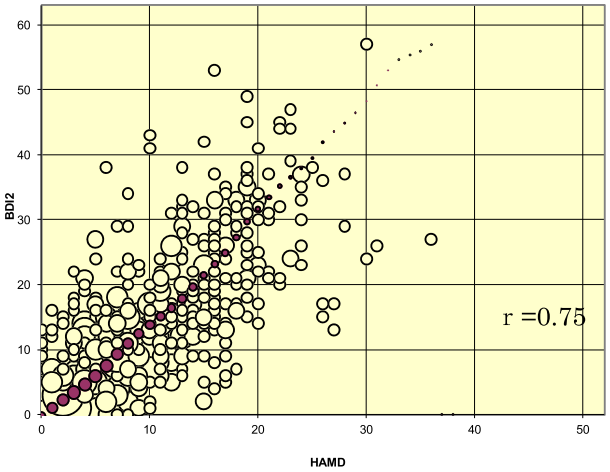
b. HAMD and BDI at endpoint



c. HAMD and BDI-II at baseline



d. HAMD and BDI-II at endpoint



White circles: Raw data, Purple circles: Equipercentile linking

https://doi.org/10.1017/S2045796019000088 Published online by Cambridge University Press