

## Review article

# Initial severity of depression and efficacy of cognitive–behavioural therapy: individual-participant data meta-analysis of pill-placebo-controlled trials

Toshi A. Furukawa,\* Erica S. Weitz,\* Shiro Tanaka,\* Steven D. Hollon, Stefan G. Hofmann, Gerhard Andersson, Jos Twisk, Robert J. DeRubeis, Sona Dimidjian, Ulrich Hegerl, Roland Mergl, Robin B. Jarrett, Jeffrey R. Vittengl, Norio Watanabe and Pim Cuijpers

## Background

The influence of baseline severity has been examined for antidepressant medications but has not been studied properly for cognitive–behavioural therapy (CBT) in comparison with pill placebo.

## Aims

To synthesise evidence regarding the influence of initial severity on efficacy of CBT from all randomised controlled trials (RCTs) in which CBT, in face-to-face individual or group format, was compared with pill-placebo control in adults with major depression.

## Method

A systematic review and an individual-participant data meta-analysis using mixed models that included trial effects as random effects. We used multiple imputation to handle missing data.

## Results

We identified five RCTs, and we were given access to individual-level data ( $n=509$ ) for all five. The analyses revealed that the difference in changes in Hamilton Rating Scale for Depression between CBT and pill placebo was not influenced by baseline severity (interaction  $P=0.43$ ). Removing the non-significant interaction term from the model, the difference between CBT and pill placebo was a standardised mean difference of  $-0.22$  (95% CI  $-0.42$  to  $-0.02$ ,  $P=0.03$ ,  $I^2=0\%$ ).

## Conclusions

Patients suffering from major depression can expect as

much benefit from CBT across the wide range of baseline severity. This finding can help inform individualised treatment decisions by patients and their clinicians.

## Declaration of interest

T.A.F. has received lecture fees from Eli Lilly, Meiji, Mochida, MSD, Otsuka, Pfizer and Tanabe-Mitsubishi and consultancy fees from Sekisui Chemicals and Takeda Science Foundation. He has received royalties from Igaku-Shoin, Seiwa-Shoten and Nihon Bunka Kagaku-sha publishers. He has received grant or research support from the Japanese Ministry of Education, Science and Technology, the Japanese Ministry of Health, Labour and Welfare, the Japan Society for the Promotion of Science, the Japan Foundation for Neuroscience and Mental Health, Mochida and Tanabe-Mitsubishi. He is diplomate of the Academy of Cognitive Therapy. U.H. has received lecture fees from Bristol-Myers Squibb, Medice Arzneimittel, Novartis and Roche Pharma and was an advisory board member for Lilly, Lundbeck, Takeda Pharmaceuticals, Servier and Otsuka Pharma. R.M. has served as a consultant for Nycomed, a Takeda company. R.B.J. received support from the NIMH for the trial cited here (ROMH45043) and is a paid consultant to NIH including the NIMH and to UpToDate. J.R.V. is a paid consultant to UpToDate. N.W. has received royalties from Sogensha and Paquet, and speaker fees and research funds from Asahi Kasei, Dai-Nippon Sumitomo, Eli Lilly, GlaxoSmithKline, Janssen, Meiji, MSD, Otsuka and Pfizer.

## Copyright and usage

© The Royal College of Psychiatrists 2017.

Antidepressant medications and psychotherapies, especially cognitive–behavioural therapy (CBT), constitute two major options as empirically supported treatments for adults with major depression.<sup>1,2</sup> Psychotherapies have traditionally been believed to be better suited for treating people with mild depression. For example, the third edition of the practice guidelines for major depression published by the American Psychiatric Association states: ‘Use of a depression-focused psychotherapy alone is recommended as an initial treatment choice for patients with mild to moderate major depressive disorder’. However, controversies have arisen recently with regard to the influence of baseline severity on the efficacy of psychotherapies for depression. One study-level analysis of trials comparing various psychotherapies, including CBT, interpersonal psychotherapy and supportive therapy, against pill placebo did not find an association between the study’s effect size and mean baseline depression severity.<sup>3</sup> Yet

another analysis at the study level found no effect of pre-treatment depression scores on the effect size of different psychotherapies over various control conditions (including waiting list, no treatment or pill placebo), but did find a clear indication that the effect size was bigger and statistically significant in high-severity patients and smaller and non-significant in low-severity patients in the small subset of studies that had examined such differences within each study.<sup>4</sup> However, these study-level analyses are limited in statistical power and are subject to the ecological fallacy that the relationships observed at the group level may not reflect the true relationships at the individual level.<sup>5</sup> Meta-analyses based on individual-participant data (IPD-MA) are necessary to examine patient-level effect moderators, including baseline severity. An IPD-MA of low-intensity psychotherapies (such as guided self-help bibliotherapy and internet-delivered therapy: 16 randomised controlled trials (RCTs), 2470 patients) found a significant albeit small interaction between baseline severity and treatment effect, suggesting that patients who are

\*These authors contributed equally to the work.

initially more depressed demonstrate slightly larger treatment effects with such low-intensity interventions.<sup>6</sup>

Given the direct relevance of baseline severity with respect to tailoring treatment and informed consent for patients with mild, moderate and severe symptoms, the possibility of the influence of baseline severity on efficacy of treatments for adults with major depression merits further investigation. Such investigation needs to be methodologically rigorous in applying individual-participant-level meta-analysis and procedurally comprehensive in identifying and assembling individual data to be so analysed. The current study aims to systematically identify all RCTs examining CBT in a face-to-face format in comparison with pill placebo in the treatment of adults with major depression, retrieve their individual-level data and apply IPD-MA to examine the possible influence of baseline depression severity on its efficacy. CBT is by far the best-researched and widely practised form of psychotherapy for depression.<sup>7,8</sup> We chose pill placebo as the control condition in this IPD-MA mainly for two reasons. First, traditional control conditions used in psychotherapy trials, such as waiting list, no treatment or treatment as usual, are too heterogeneous and may be affected by publication bias.<sup>9,10</sup> Second, and more importantly, the pill-placebo condition will control for non-specific placebo effects including expectation, attention and support and will offer a common comparison condition for both antidepressants and psychotherapies. Our hypothesis was that increases in baseline severity would be associated with increases in the advantages of CBT over pill placebo among adult patients with major depression.

## Method

### Selection of studies and data extraction

The eligibility criteria for the present IPD-MA were as follows:

- (a) RCT;
- (b) the participants were adult patients with major depressive disorder, diagnosed according to any operational diagnostic criteria including the Research Diagnostic Criteria, DSM-III, DSM-III-R, DSM-IV, DSM-5 or ICD-10;
- (c) the intervention was face-to-face CBT in individual or group format. We defined CBT as a psychotherapy focused on the impact of a patient's dysfunctional thoughts on his/her current behaviour and future functioning and aimed at evaluating, challenging and modifying such dysfunctional beliefs (cognitive restructuring). We included two subtypes thereof, namely one in which cognitive restructuring is the core element and one in which it is an important component but in which at least two other components such as behavioural activation, social skills training, relaxation or coping skills also have a prominent role;<sup>11</sup>
- (d) the non-specific control was pill placebo. The trial could have other active or non-active intervention arms.

We conducted a systematic and cumulative search of all RCTs of psychotherapies for depression ([www.evidencebasedpsychotherapies.org](http://www.evidencebasedpsychotherapies.org)). The searches included PubMed, PsycInfo, EMBASE and the Cochrane Central Register of Controlled Trials using keywords indicative of psychological treatment and depression, and were supplemented by examination of published meta-analyses of psychological treatments for depression, up to January 2014. An updated search was conducted on 16 January 2015 with PubMed and the Cochrane Central Register. We identified the relevant studies satisfying our eligibility criteria from this database and then asked the primary authors of all the identified studies to contribute individual-participant data

for our primary and secondary outcomes (see below) as well as any possible effect moderators that each study may have measured. Study quality was assessed using four criteria from the Cochrane Collaboration's risk of bias tool for random sequence generation, concealment of random allocation, masking of assessors and complete outcome reporting. When the information provided in the publications was not clear, we sought information from the original study authors.

### Outcomes

Our pre-specified primary outcome measure was change in the Hamilton Rating Scale for Depression (HRSD)<sup>12</sup> scores from baseline to the last study time point at which valid comparison between CBT and pill placebo could be made. Two studies employed the pill-placebo arm only through week 8 of a 16-week study,<sup>13,14</sup> in which case we were able to compare CBT against pill placebo up to that time point only. Our secondary outcome was change in the Beck Depression Inventory (BDI).<sup>15</sup>

### Analyses

We first verified the integrity of the individual-participant data by comparing their descriptive statistics against the published figures. We conducted IPD-MA to estimate the difference in symptom change scores between the CBT and pill-placebo arms and to examine the influence of the baseline symptom severity on this difference. Because the original studies used different versions of the HRSD and BDI with different item questions and with different numbers of items, such continuous outcomes were standardised within each study, that is, the individual patients' raw scores at baseline and at end-point were divided by the study-specific standard deviations of end-point scores. After this standardisation of scores from different scales, when we graphically presented the relationship between the scores at baseline and the changes in scores, we subtracted the grand mean of all the studies' standardised scores from individual standardised scores in order to anchor the horizontal and vertical axes in the figure. The relationships between these standardised baseline symptom severity scores and the differences in symptom change scores were examined as the interaction term between baseline severity and treatment in the mixed models. When this interaction term was not statistically significant, we dropped the interaction term and estimated the main effect of treatment. The mixed models included trial effects as random effects throughout.

Missing data in HRSD and BDI scores were handled by the multiple imputation method using study, treatment, age, gender, minority status, marital status, employment, education, age at onset, diagnosis of depression subtype, presence of previous episodes, length of current episode, comorbid anxiety disorders, past substance use disorders, length of treatment and HRSD or BDI scores at baseline and at the last study time point as predictors. Multiple imputation is one of the currently recommended alternatives to the long-used last-observation-carried-forward (LOCF) method.<sup>16</sup> It has some advantages over hierarchical linear modelling as well as LOCF because auxiliary variables that are not included in the final analysis model can be used in the imputation model.<sup>17</sup>

We also conducted sensitivity analyses (a) excluding studies that continued the pill-placebo arm only through the mid-point of acute treatment, (b) excluding studies that administered CBT in a group format, (c) including patients diagnosed with dysthymia or minor depression at baseline, (d) employing the multiple imputation using the predictors other than treatment, and (e) restricted to only those participants who completed the treatment.

All reported  $P$ -values are two-tailed and  $P < 0.05$  was chosen as the threshold for statistical significance. An academic statistician (S.T.) conducted all statistical analyses using SAS version 9.3. This study was conducted in accordance with PRISMA and PRISMA-IPD guidelines.<sup>18</sup>

## Results

### Studies and participants

We found five RCTs that satisfied our eligibility criteria (online Fig. DS1).<sup>13,14,19–21</sup> One study<sup>21</sup> included some patients with minor depression or dysthymia: the latter patients were excluded in our primary analyses but were included in a sensitivity analysis to give a wider range of baseline severity. All the primary authors and their collaborators agreed to contribute to the present study by providing individual-level data for the primary and secondary outcomes as well as possible effect moderators that they had measured. The data from the NIMH Treatment of Depression Collaborative Program<sup>19</sup> were in the public domain and available for inclusion.

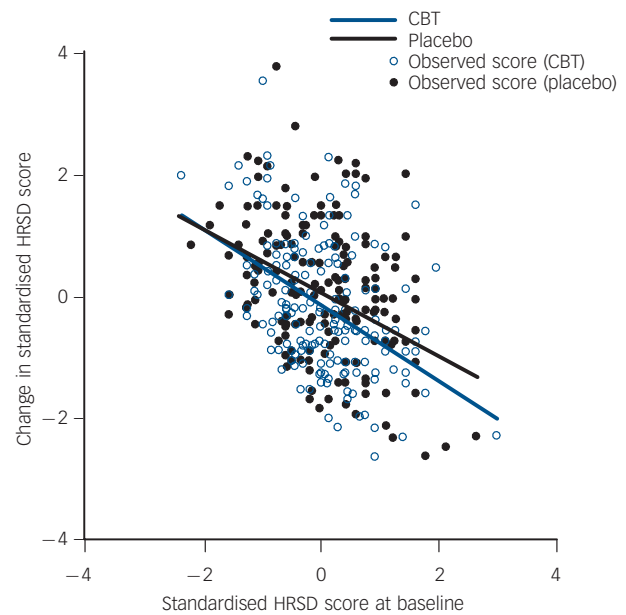
Table 1 lists the characteristics of the five RCTs that compared CBT against pill placebo. We obtained individual-patient data for 509 patients with major depression and 46 patients with minor depression and dysthymia. The mean age of the patients with major depression participating in these trials was in the 30s or 40s, with a preponderance of women. In addition to CBT and pill-placebo arms, the comparison of interest in the present study, all the studies had an antidepressant medication arm and three studies also included a different psychotherapy from CBT. One study<sup>21</sup> used the original 17-item version of HRSD, one study<sup>20</sup> the 21-item version, and the remaining three studies<sup>13,14,19</sup> employed the modified HRSD with additional items for atypical symptoms. Two studies<sup>19,20</sup> used the original BDI and three studies<sup>13,14,21</sup> used the second version.

All the studies were rated at low risk of bias for random sequence generation, allocation concealment, masking and incomplete outcome reporting, except for one study whose risk of bias was rated unclear as to allocation concealment but low for all the other aspects.<sup>14</sup> Raters masked to treatment assignment administered the HRSD in all studies. As a self-report measure, patients knew whether they were receiving CBT or were in a 'medication' condition when they completed the BDI; however, they were masked as to whether they were taking an active medication or a pill placebo.

### Baseline severity and symptom change in CBT and pill placebo

Figure 1 depicts observed and estimated changes in standardised HRSD scores in the CBT or pill-placebo arms. The end-point standard deviations used to standardise the HRSD score were 5.8 for DeRubeis *et al.*,<sup>13</sup> 6.6 for Dimidjian *et al.*,<sup>14</sup> 6.3 for Elkin *et al.*,<sup>19</sup> 6.2 for Hegerl *et al.*<sup>21</sup> and 7.3 for Jarrett *et al.*,<sup>20</sup> and we used the grand means of the standardised scores from all the studies, 3.2 for baseline scores and 1.3 for change scores, as subtrahends in scaling the axes in Fig. 1. The individual participant-level meta-analysis revealed that the differences in changes in HRSD between CBT and pill-placebo arms were not influenced by baseline HRSD severity. The slopes in the CBT and placebo arms were estimated to be  $-0.615$  (95% CI  $-0.839$  to  $-0.392$ ,  $P < 0.01$ ) and  $-0.526$  (95% CI  $-0.733$  to  $-0.319$ ,  $P < 0.01$ ), respectively and the interaction term, i.e. the difference between the slopes, was  $-0.089$  ( $-0.311$  to  $0.133$ ,  $P = 0.43$ ).

The interaction between baseline severity and treatment condition was also not significant with change in BDI as the



**Fig. 1** Observed (dots) and estimated (lines) changes in standardised scores on Hamilton Rating Scale for Depression (HRSD) in the cognitive-behavioural therapy (CBT) and pill-placebo arms.

dependent variable. The slopes in the CBT and pill-placebo arms were estimated to be  $-0.541$  (95% CI  $-0.724$  to  $-0.358$ ,  $P < 0.01$ ) and  $-0.615$  (95% CI  $-0.793$  to  $-0.438$ ,  $P < 0.01$ ), respectively and the interaction term was  $0.074$  ( $-0.175$  to  $0.323$ ,  $P = 0.56$ ). Removing the higher-order non-significant interaction terms from the models, the IPD-MA revealed that the standardised mean difference (SMD) between CBT and pill-placebo arms was  $-0.220$  (95% CI  $-0.419$  to  $-0.022$ ,  $P = 0.03$ ,  $I^2 = 0\%$ ) for the HRSD and  $-0.046$  (95% CI  $-0.264$  to  $0.172$ ,  $P = 0.68$ ,  $I^2 = 0\%$ ) for the BDI (Table 2).

### Sensitivity analyses

We also conducted sensitivity analyses to examine the robustness of our primary analyses by (a) excluding DeRubeis *et al.*<sup>13</sup> and Dimidjian *et al.*'s<sup>14</sup> studies that had the pill-placebo arm only up to the mid-point of the study, (b) excluding Hegerl *et al.*'s study<sup>21</sup> that administered group-format CBT, (c) including patients diagnosed with dysthymia or minor depression at baseline, (d) multiple imputation using the predictors other than treatment and (e) restricting the sample to participants who completed the treatment only (Table 3). The interaction terms were all statistically non-significant. The treatment effects in models in which the interaction term was excluded were statistically significant in the sensitivity analyses (c) and (e) above but not in (a), (b) or (d), likely because of reduced sample sizes or to residual bias related to missing data. However, the estimated effects ranged between  $-0.16$  and  $-0.27$ , and in all instances their 95% confidence intervals overlapped with those from the primary analysis ( $-0.22$ , 95% CI  $-0.42$  to  $-0.02$ ).

## Discussion

### Main findings

We identified five pill-placebo-controlled randomised studies of CBT conducted to date, obtained their individual-level data (509 patients with major depression), and conducted an IPD-MA to examine the influence of baseline depression severity on

**Table 1** Characteristics of the included five randomised controlled trials (RCTs) and their participants included in the current meta-analysis

Study	Key eligibility criteria for the original study	Randomisation arms in the original studies	Characteristics of the participants included in the present IPD-MA	Effect size of CBT over pill placebo according to the original publications and used in a previous aggregate data meta-analysis <sup>27</sup>
DeRubeis <i>et al.</i> <sup>13</sup> (2005)	Major depression with baseline modified HRSD-17 <sup>a</sup> score $\geq 20$	CBT ( $n = 60$ , 16 weeks: 50 min sessions twice weekly for the first 4 weeks, once or twice weekly for the next 8 weeks, and once weekly for the final 4 weeks) Paroxetine ( $n = 120$ , 16 weeks: once weekly for the first 4 weeks, biweekly thereafter; initial session lasted 30–45 min, and 20 min thereafter) Pill placebo ( $n = 60$ , 8 weeks: same as paroxetine)	Diagnosis: major depression ( $n = 120$ ) Mean age: 41.0 Women, %: 59.2% Baseline modified HRSD-17 <sup>a</sup> mean: 24.1 (s.d. = 3.4, range 20–36)	–0.44 ( $P = 0.09$ ) from HLM
Dimidjian <i>et al.</i> <sup>14</sup> (2006)	Major depression with baseline modified HRSD-17 <sup>a</sup> score $\geq 14$	CBT ( $n = 45$ , 16 weeks: maximum of 24 50 min sessions, with sessions generally held twice weekly for the first 8 weeks and once weekly for the next 8 weeks) Behavioural activation ( $n = 43$ , 16 weeks: same as CBT) Paroxetine ( $n = 100$ , 16 weeks: once weekly for the first 4 weeks, biweekly thereafter; initial session lasted 30–45 min, and 30 min thereafter) Pill placebo ( $n = 53$ , 8 weeks: same as paroxetine)	Diagnosis: major depression ( $n = 98$ ) Mean age: 39.8 Women, %: 72.5% Baseline modified HRSD-17 <sup>a</sup> mean: 20.6 (s.d. = 4.5, range 14–34)	–0.40 (ns) from completers' data
Elkin <i>et al.</i> <sup>19</sup> (1989)	Major depression with baseline modified HRSD-17 <sup>a</sup> score of $\geq 14$	CBT ( $n = 59$ , 16 weeks: 16 to 20 50 min sessions) IPT ( $n = 61$ , 16 weeks: same as CBT) Imipramine ( $n = 57$ , 16 weeks: 16–20 sessions; the initial session lasted 45–60 min and 20–30 min thereafter) Pill placebo ( $n = 62$ , 16 weeks: same as imipramine)	Diagnosis: major depression ( $n = 121$ ) Mean age: 35.0 Women, %: 71.9% Baseline modified HRSD-17 <sup>a</sup> mean: 20.7 (s.d. = 4.0, range 14–32)	–0.32 (ns) from LOCF data
Hegerl <i>et al.</i> <sup>21</sup> (2010)	Minor depression, dysthymia or major depression with mild to moderate severity with baseline HRSD-17 score of 8–22	CBT ( $n = 61$ , 10 weeks: 9 weekly 90 min group sessions) Sertraline ( $n = 83$ , 10 weeks: 6 sessions) Pill placebo ( $n = 83$ , 10 weeks: same as sertraline) Guided self-help ( $n = 59$ , 10 weeks: same as CBT) Patients' choice (patient could choose either CBT or sertraline) ( $n = 82$ )	Diagnosis: major depression ( $n = 98$ ), minor depression or dysthymia ( $n = 46$ ) Mean age: 47.9 Women, %: 70.4% Baseline HRSD-17 mean: 16.2 (s.d. = 4.4, range 5–27) for major depression; mean: 15.5 (s.d. = 4.6, range 3–23) for minor depression or dysthymia	–0.34 ( $P = 0.31$ ) from completers' data <sup>b</sup>
Jarrett <i>et al.</i> <sup>20</sup> (1999)	Major depression with atypical features with baseline HRSD-21 score of $\geq 14$	CBT ( $n = 36$ , 10 weeks: 20 twice-weekly sessions) Phenelzine ( $n = 36$ , 10 weeks: 11 sessions) Pill placebo ( $n = 36$ , 10 weeks: same as phenelzine)	Diagnosis: major depression ( $n = 72$ ) Mean age: 40.1 Women, %: 66.7% Baseline HRSD-21 mean: 17.9 (s.d. = 3.3, range 14–28)	–0.53 ( $P < 0.01$ ) from LOCF data

IPD-MA, individual-participant data meta-analysis; CBT, cognitive-behavioural therapy; HRSD, Hamilton Rating Scale for Depression; HLM, hierarchical linear modelling; IPT, interpersonal psychotherapy; ns, non-significant; LOCF, last-observation-carried-forward.

a. Modified by including items for hypersomnia, hyperphagia and weight gain.

b. Lynch *et al.*<sup>27</sup> did not include this study, which was published after their systematic review. We included this reported effect size here for comparison.

**Table 2** Treatment effects of cognitive-behavioural therapy (CBT) over pill placebo on different outcome measures

	<i>P</i> for interaction	<i>P</i> for main effect excluding the interaction term	Standardised mean difference (main effect) <sup>a</sup> (95% CI)
Hamilton Rating Scale for Depression change	0.43	0.03	−0.220 (−0.419 to −0.022)
Beck Depression Inventory change	0.56	0.68	−0.046 (−0.264 to 0.172)

a. Estimated by random-effects models after multiple imputation without the interaction term.

**Table 3** Sensitivity analyses

	<i>P</i> for interaction	<i>P</i> for main effect excluding the interaction term	Standardised mean difference (main effect) <sup>a</sup> (95% CI)
Excluding two trials that continued the pill-placebo arm only through the mid-point of acute treatment and in which the comparison between CBT and placebo was therefore possible only up to the middle of the study <sup>13,14</sup>	0.21	0.12	−0.217 (−0.494 to 0.060)
Excluding a trial that administered CBT in group format <sup>21</sup> and limiting to CBT in individual format	0.67	0.08	−0.191 (−0.407 to 0.024)
Including minor depression and dysthymia	0.43	0.02	−0.230 (−0.431 to −0.030)
Multiple imputation using predictors other than treatment	0.34	0.13	−0.158 (−0.364 to 0.048)
Complete-case analysis	0.39	0.01	−0.266 (−0.480 to −0.053)

CBT, cognitive-behavioural therapy.  
a. Estimated by random-effects models without the interaction term after multiple imputation except for the complete-case analysis.

its efficacy. There was no statistically significant influence of baseline depression severity, measured either with HRSD or BDI, on subsequent differential symptom change between CBT and pill placebo. The findings, thus, did not support our initial hypothesis that the efficacy of CBT over pill placebo would increase as baseline depression severity increased, when it was measured either by patient report or clinician ratings.

Finding that baseline severity has little influence on the efficacy of CBT can be said to be largely in accord with the previous literature, using study-level data<sup>4,22</sup> or individual-participant data.<sup>6</sup> Although a handful of studies that examined subgroup differences between low- *v.* high-severity groups have found evidence for moderation (with differences relative to pill placebo emerging only among patients with more severe depression) that was only the case for more purely behavioural interventions or interpersonal psychotherapy.<sup>4</sup> The IPD-MA that did find moderation by baseline severity concluded that the magnitude of interaction was small and that patients who were less severely depressed could derive as much benefit from low-intensity psychotherapies as did individuals who were more severely depressed.<sup>6</sup> Taken together, the evidence to date regarding the comparison of CBT *v.* pill placebo suggests that we can expect as much benefit from CBT across a wide range of baseline depression severity.

Excluding the non-significant interaction term and compared with pill placebo, CBT led to greater symptom reduction, on average, by an SMD of −0.22 (95% CI −0.42 to −0.02, *P* = 0.03) on HRSD. The SMD in terms of BDI was −0.05 and was not statistically significant (95% CI −0.26 to 0.17, *P* = 0.68). The HRSD was the prespecified primary outcome in this IPD-MA, as it was in all the included studies.<sup>13,14,19–21</sup> Generally, observer-rating scales including HRSD are said to be more sensitive to change than self-rating scales such as BDI.<sup>23,24</sup> Specifically in all the included studies, the HRSD was rated by assessors masked to all the treatments whereas the BDI was self-rated by participants who were masked to the distinctions in medications (active *v.* placebo) but not to those between CBT *v.* medications. It must also be pointed out that BDI was more often missing than HRSD in the original data-set (19.5% *v.* 13.7%).

### Comparison with findings from other studies

The SMD of 0.22 found in the primary analysis of this meta-analysis is smaller than the previously estimated SMD of CBT over waiting-list conditions, the most common control condition in the psychotherapy trials, such as 0.82<sup>25</sup> or 0.85.<sup>8</sup> However, it must be remembered that the magnitude of effect is dependent not only on the effects of treatment but also on the effects of the control condition against which the therapy is compared. Waiting-list control conditions are known to lead to greater effect estimates than other control conditions, including clinical management that is accompanied by the provision of medication placebos.<sup>10,23</sup> In the control arms in four of the five trials in our data-set<sup>13,14,19,20</sup> the pharmacotherapists explicitly followed the clinical management protocol originally developed for the NIMH Treatment of Depression Collaborative Program<sup>24</sup> and in the remaining study<sup>21</sup> the psychiatrists' practices were in accord with this protocol, so that these pill-placebo conditions have included strong supportive components. Furthermore, the effect size estimates obtained in previous comparisons of psychotherapies with various control conditions may have been inflated by the inclusion of poor-quality studies<sup>26</sup> and by publication bias.<sup>9</sup> Our choice of the pill-placebo condition, which controls for the non-specific effects of expectation, attention and support, our comprehensive and systematic search of the relevant literature, and the high quality of the included studies including masked assessment of the outcome should have mitigated these problems.

Second, our estimate of the CBT–pill-placebo difference was also smaller than that reported in a previous study-level meta-analysis of pill-placebo-controlled trials. Lynch *et al*<sup>27</sup> reported an SMD of 0.41 (95% CI 0.21–0.61) between CBT and pill placebo from a study-level aggregate data meta-analysis of four of the five studies included in our IPD-MA.<sup>13,14,19,20</sup> This meta-analysis took effect size estimates for the included studies according to the original publications (see Table 1). One of the estimates was based on hierarchical linear modelling (HLM),<sup>13</sup> two others on the LOCF data<sup>19,20</sup> (although results from random regression were available from one<sup>20</sup> but were not used), and another on completers' data.<sup>14</sup> In the arms of the included studies, 15–64% of the randomised patients dropped out per arm and therefore

did not provide data at the end-point assessment. How drop-out was handled in the respective studies would be expected to influence estimates of the CBT *v.* pill-placebo difference. Although in the original publications, the trial authors may have used statistical approaches, including HLM<sup>14,21</sup> that are currently considered appropriate or were deemed to be so at the time they were published, the systematic review authors were obliged to extract data in the format amenable to meta-analysis, as detailed above. However, as methods have progressed today, neither the LOCF nor the completers' data would be considered the standard method to summarise and analyse data with missing values.<sup>16</sup> By contrast, we were able to meta-analyse the five studies with individual-participant data, which allowed us to use multiple imputation and test for effects using a mixed model. It is increasingly recognised that IPD-MA has many advantages, albeit with its own difficulties,<sup>28</sup> and some authors consider it to be the gold standard in evidence synthesis.<sup>29</sup>

Many questions remain when we try to factor antidepressant pharmacotherapy into the comparison between CBT and pill placebo. First, our estimated standardised effect size of 0.22 appears smaller than the publication-bias-free estimate of 0.31 of newer antidepressants over pill placebo.<sup>30</sup> However, it is problematic to compare these estimates directly with each other, as they were obtained from different original investigations. Moreover, our estimate for CBT *v.* pill-placebo difference is based on IPD-MA using multiple imputation, whereas that of the antidepressant *v.* pill-placebo difference derives from an aggregate data meta-analysis from published and unpublished trials, many of which used LOCF. Second, when directly compared in a study-level meta-analysis, the difference between CBT and second-generation antidepressants was not statistically significant.<sup>31</sup> A recent study applying IPD-MA to 1466 patients from 14 out of 24 identified relevant studies comparing CBT and pharmacotherapies revealed a small difference in effect size (SMD = 0.11,  $P = 0.03$ ) but no significant differences in response or remission rates (odds ratio (OR) = 1.24,  $P = 0.12$  and OR = 1.18,  $P = 0.22$ , respectively).<sup>32</sup> Lastly, the independence of the effect of CBT for depression from the baseline severity, as shown in the present analysis, is also in contrast to some claims to the contrary with regard to the effect of antidepressant drugs.<sup>33</sup> There is collateral evidence that the effect of antipsychotic drugs is also moderated by baseline psychotic severity.<sup>34</sup> However, the question of the influence of baseline symptom severity on antidepressant medications' efficacy is hardly closed,<sup>35</sup> and hence neither is the issue if such influences may differ between antidepressant medication and psychotherapies.

## Limitations

Our study has limitations. First, detection of an interaction effect requires a large sample size. The present findings require replication with a larger sample of trials designed to test the interaction. The trials from which the data were extracted were not designed to test this hypothesis, but do provide a 'first look' into this question using all available data. Given the wide confidence interval around zero for the interaction coefficient in the present study, it is possible that our finding represents a type II error. Although none of the sensitivity analyses were suggestive of possible interaction effects, future studies may nonetheless still reveal an interaction that we failed to detect in the data available to us. It is disappointing that there were only 5 studies that compared CBT with pill placebo, whereas more than 20 studies compared it with antidepressant medications<sup>32</sup> and hundreds of studies have compared antidepressants with pill placebo.<sup>36</sup> Second, we were not able to examine important potential moderators such as therapist competence or other patient characteristics including comorbid personality disorder because such data were not consistently available in the obtained data-set although such

covariates were requested in the source data from parent trials. Third, as the included studies used different versions of HRSD or BDI, we were obliged to standardise these severity measures using their means and standard deviations. This was unavoidable in order to synthesise different metrics across studies. Lastly, in combining data, we also ignored some design features from the parent studies that may or may not prove to be important in future understanding of antidepressant treatment effects (for example exclusive focus on subtypes of major depression in some studies, such as atypical depression,<sup>20</sup> primary care settings;<sup>21</sup> inclusion of a 'non-specific treatment run-in' prior to randomisation;<sup>19,20</sup> differential lengths of acute phase treatment; or inclusion of active psychotherapy arms other than CBT in the study). In view of today's growing emphasis on the importance of the totality of evidence<sup>37</sup> and hence systematic reviews including IPD-MAs,<sup>29</sup> researchers may benefit by making concerted efforts to measure core variables in their original research and then by contributing to the common database.<sup>38</sup>

## Implications

Clinical implications of the present study may be as follows. First of all, patients and their clinicians can expect as much benefit from CBT for major depression across its wide range of baseline severity from mild through severe. Second, the benefit that can be expected was an effect size of 0.22, from which it can be calculated that the number needed to treat (NNT) is 12 for typical cases of major depression, for whom the expected placebo response rates may be between 30 and 50%.<sup>39</sup> This would compare favourably with the NNT of nine that can be expected for antidepressants with an effect size of 0.31 over placebo.<sup>30</sup> The difference in NNTs to obtain one more response with CBT or antidepressant medications than with pill placebo is therefore small enough to allow value judgements and preferences of individual patients to play a major role in treatment decision-making.

There are several research implications. We have been fortunate in having been able to identify and assemble all the relevant randomised evidence on the topic. We need to foster further collaborations in the field of psychotherapy research. Similar collaboration is burgeoning and is welcome in all fields of medicine.<sup>38</sup> It is also expected that such collaboration after completion of individual studies will increase the research community's awareness of the need for greater collaboration before individual studies by agreeing on common core variables to be assessed in the studies.

**Toshi A. Furukawa**, MD, PhD, Departments of Health Promotion and Human Behavior and of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan; **Erica S. Weitz**, MA, Department of Clinical Psychology and EMGO Institute for Health and Care Research, VU University Amsterdam, The Netherlands; **Shiro Tanaka**, PhD, Department of Pharmacoepidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan; **Steven D. Hollon**, PhD, Department of Psychology, Vanderbilt University, Nashville, Tennessee, USA; **Stefan G. Hofmann**, PhD, Department of Psychological and Brain Science, Boston University, Massachusetts, USA; **Gerhard Andersson**, PhD, Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden and Department of Clinical Neuroscience, Psychiatry Section, Karolinska Institutet, Stockholm, Sweden; **Jos Twisk**, PhD, Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Amsterdam, The Netherlands; **Robert J. DeRubeis**, PhD, Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania, USA; **Sona Dimidjian**, PhD, Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, USA; **Ulrich Hegerl**, MD, PhD, **Roland Mergl**, PhD, Department of Psychiatry and Psychotherapy, University of Leipzig, Leipzig, Germany; **Robin B. Jarrett**, PhD, Department of Psychiatry, The University of Texas Southwestern Medical Center, Dallas, Texas, USA; **Jeffrey R. Vittengl**, PhD, Department of Psychology, Truman State University, Kirksville, Missouri, USA; **Norio Watanabe**, MD, PhD, Departments of Health Promotion and Human Behavior and of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan; **Pim Cuijpers**, PhD, Department of Clinical Psychology and EMGO Institute for Health and Care Research, VU University Amsterdam, The Netherlands

**Correspondence:** Toshi A. Furukawa, MD, PhD, Departments of Health Promotion and Human Behavior and of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Yoshida Konoe-cho, Sakyo-ku, Kyoto, 606-8501 Japan. Email: furukawa@kuhp.kyoto-u.ac.jp

First received 13 May 2016, final revision 6 Sep 2016, accepted 10 Sep 2016

## Funding

This study was supported in part by Grant-in-Aid for Challenging Exploratory Research from Japan Society for the Promotion of Science (JSPS) to T.A.F. and S.T. (No. 26670314) and by Health and Labour Sciences Research Grant to T.A.F. (H25-Seishin-Ippan-002). S.D.H. is supported by NIMH grant MH60713 and MH01697. S.G.H. receives support from NIH/NCCIH (R01AT007257), NIH/NIMH (R01MH099021, R34MH099311, R34MH086668, R21MH102646, R21MH101567, K23MH100259) and the Department of the Army for work unrelated to the studies reported in this article. R.J.D. is supported by NIMH grant MH60998. N.W. has research funds from the Japanese Ministry of Health Labor and Welfare and the Japanese Ministry of Education, Science and Technology. The funders had no role in the study design, the collection, analysis and interpretation of data, the writing of the report, or the decision to submit the article for publication.

## References

- National Institute for Health and Care Excellence. *Depression: The Treatment and Management of Depression in Adults (partial update of NICE clinical guideline 23)*. NICE, 2009.
- American Psychiatric Association. Practice guideline for the treatment of patients with major depressive disorder (third edition). *Am J Psychiatry* 2010; **167** (suppl): 1–152.
- Cuijpers P, Turner EH, Mohr DC, Hofmann SG, Andersson G, Berking M, et al. Comparison of psychotherapies for adult depression to pill placebo control groups: a meta-analysis. *Psychol Med* 2013: 1–11.
- Driessen E, Cuijpers P, Hollon SD, Dekker JJ. Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? A meta-analysis. *J Consult Clin Psychol* 2010; **78**: 668–80.
- Simmonds MC, Higgins JP. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. *Stat Med* 2007; **26**: 2982–99.
- Bower P, Kontopantelis E, Sutton A, Kendrick T, Richards DA, Gilbody S, et al. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. *BMJ* 2013; **346**: f540.
- Chen P, Furukawa TA, Shinohara K, Honyashiki M, Imai H, Ichikawa K, et al. Quantity and quality of psychotherapy trials for depression in the past five decades. *J Affect Disord* 2014; **165**: 190–5.
- Barth J, Munder T, Gerger H, Nuesch E, Trelle S, Znoj H, et al. Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. *PLoS Med* 2013; **10**: e1001454.
- Cuijpers P, Smit F, Bohlmeijer E, Hollon SD, Andersson G. Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. *Br J Psychiatry* 2010; **196**: 173–8.
- Furukawa TA, Noma H, Caldwell DM, Honyashiki M, Shinohara K, Imai H, et al. Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis. *Acta Psychiatr Scand* 2014; **130**: 181–92.
- Cuijpers P, van Straten A, Andersson G, van Oppen P. Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. *J Consult Clin Psychol* 2008; **76**: 909–22.
- Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960; **23**: 56–62.
- DeRubeis RJ, Hollon SD, Amsterdam JD, Shelton RC, Young PR, Salomon RM, et al. Cognitive therapy vs medications in the treatment of moderate to severe depression. *Arch Gen Psychiatry* 2005; **62**: 409–16.
- Dimidjian S, Hollon SD, Dobson KS, Schmaling KB, Kohlenberg RJ, Addis ME, et al. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *J Consult Clin Psychol* 2006; **74**: 658–70.
- Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psych* 1961; **4**: 561–71.
- Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med* 2012; **367**: 1355–60.
- National Research Council. *The Prevention and Treatment of Missing Data in Clinical Trials*. National Academies Press, 2010.
- Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, et al. Preferred reporting items for systematic review and meta-analyses of individual participant data: the PRISMA-IPD Statement. *JAMA* 2015; **313**: 1657–65.
- Elkin I, Shea MT, Watkins JT, Imber SD, Sotsky SM, Collins JF, et al. National Institute of Mental Health Treatment of Depression Collaborative Research Program. General effectiveness of treatments. *Arch Gen Psychiatry* 1989; **46**: 971–82.
- Jarrett RB, Schaffer M, McIntire D, Witt-Browder A, Kraft D, Risser RC. Treatment of atypical depression with cognitive therapy or phenelzine: a double-blind, placebo-controlled trial. *Arch Gen Psychiatry* 1999; **56**: 431–7.
- Hegerl U, Hautzinger M, Mergl R, Kohlen R, Schutze M, Scheunemann W, et al. Effects of pharmacotherapy and psychotherapy in depressed primary-care patients: a randomized, controlled trial including a patients' choice arm. *Int J Neuropsychopharmacol* 2010; **13**: 31–44.
- Cuijpers P, Turner EH, Mohr DC, Hofmann SG, Andersson G, Berking M, et al. Comparison of psychotherapies for adult depression to pill placebo control groups: a meta-analysis. *Psychol Med* 2014; **44**: 685–95.
- Cunningham JA, Kypri K, McCambridge J. Exploratory randomized controlled trial evaluating the impact of a waiting list control design. *BMC Med Res Methodol* 2013; **13**: 150.
- Fawcett J, Epstein P, Fiester SJ, Elkin I, Atry JH. Clinical management–imipramine/placebo administration manual. NIMH Treatment of Depression Collaborative Research Program. *Psychopharmacol Bull* 1987; **23**: 309–24.
- Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. *Clin Psychol Rev* 2006; **26**: 17–31.
- Cuijpers P, van Straten A, Bohlmeijer E, Hollon SD, Andersson G. The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. *Psychol Med* 2010; **40**: 211–23.
- Lynch D, Laws KR, McKenna PJ. Cognitive behavioural therapy for major psychiatric disorder: does it really work? A meta-analytical review of well-controlled trials. *Psychol Med* 2010; **40**: 9–24.
- Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Eval Health Prof* 2002; **25**: 76–97.
- Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010; **340**: c221.
- Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008; **358**: 252–60.
- Amick HR, Gartlehner G, Gaynes BN, Forneris C, Asher GN, Morgan LC, et al. Comparative benefits and harms of second generation antidepressants and cognitive behavioral therapies in initial treatment of major depressive disorder: systematic review and meta-analysis. *BMJ* 2015; **351**: h6019.
- Weitz ES, Hollon SD, Twisk J, van Straten A, Huibers MJ, David D, et al. Baseline depression severity as moderator of depression outcomes between cognitive behavioral therapy vs pharmacotherapy: an individual patient data meta-analysis. *JAMA Psychiatry* 2015; **72**: 1102–9.
- Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA* 2010; **303**: 47–53.
- Furukawa TA, Levine SZ, Tanaka S, Goldberg Y, Samara M, Davis JM, et al. Initial severity of schizophrenia and efficacy of antipsychotics: participant-level meta-analysis of 6 placebo-controlled studies. *JAMA Psychiatry* 2015; **72**: 14–21.
- Gibbons RD, Hur K, Brown CH, Davis JM, Mann JJ. Benefits from antidepressants: synthesis of 6-week patient-level outcomes from double-blind placebo-controlled randomized trials of fluoxetine and venlafaxine. *Arch Gen Psychiatry* 2012; **69**: 572–9.
- Gartlehner G, Hansen RA, Morgan LC, Thaler K, Lux L, Van Noord M, et al. Comparative benefits and harms of second-generation antidepressants for treating major depressive disorder: an updated meta-analysis. *Ann Intern Med* 2011; **155**: 772–85.
- Ioannidis JP, Karassa FB. The need to consider the wider agenda in systematic reviews and meta-analyses: breadth, timing, and depth of the evidence. *BMJ* 2010; **341**: c4875.
- Institute of Medicine Committee on Strategies for Responsible Sharing of Clinical Trial Data. *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risks*. National Academic Press, 2015.
- Furukawa TA. From effect size into number needed to treat. *Lancet* 1999; **353**: 1680.

