Initial severity of major depression and efficacy of new generation antidepressants: Individual-participant data meta-analysis

RUNNNING TITLE: Baseline severity and efficacy of antidepressants

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
furukawa@kuhp.kyoto-u.ac.jp

Kazushi Maruo, PhD*
Department of Biostatistics, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan
kazushi.maruo@gmail.com

Hisashi Noma, PhD
Department of Data Science, The Institute of Statistical Mathematics, Tokyo, Japan
noma@ism.ac.jp

Shiro Tanaka, PhD
Department of Clinical Biostatistics, Kyoto University Graduate School of Medicine / School of Public Health, Kyoto, Japan
tanaka.shiro.8n@kyoto-u.ac.jp

Hissei Imai, 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
ihits@hotmail.com

Kiyomi Shinohara, MD, PhD
Department of Psychiatry, University of Oxford, Oxford, UK
Departments of Health Promotion and Human Behavior and of Clinical Epidemiology, Kyoto University Graduate School of Medicine / School of Public Health, Kyoto, Japan
shinohara.kiyomi.75r@st.kyoto-u.ac.jp

Kazutaka Ikeda, PhD
Addictive Substance Project, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan
ikeda-kz@igakuken.or.jp

Shigeto Yamawaki, MD, PhD
Academic-Industrial Cooperation Office, Hiroshima University, Hiroshima, Japan
yamawaki@hiroshima-u.ac.jp

Stephen Z. Levine, PhD
Department of Community Mental Health, Faculty of Social Welfare and Health Sciences, University of Haifa, Haifa, Israel
levine.sz@gmail.com

Yair Goldberg, PhD
Department of Statistics, Faculty of Social Sciences, University of Haifa, Haifa, Israel
ygoldberg@stat.haifa.ac.il

Stefan Leucht, MD, PhD
Department of Psychiatry and Psychotherapy, Technische Universität München, Munich, Germany
Stefan.Leucht@tum.de

Andrea Cipriani, MD, PhD
Department of Psychiatry, University of Oxford, Oxford, UK
andrea.cipriani@psych.ox.ac.uk

* These authors contributed equally to the study.

Corresponding author: Toshi A. Furukawa, 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, Japan 606-8501
Email: furukawa@kuhp.kyoto-u.ac.jp

ABSTRACT

Objective: The role of baseline severity as effect modifier in various psychiatric disorders is a topic of controversy and of clinical import. The present study aims to examine if baseline severity modifies the efficacy of various antidepressants for major depression through individual participant data (IPD) meta-analysis.

Method: We identified all placebo-controlled, double-blind randomised trials of new generation antidepressants in the acute phase treatment of major depression conducted in Japan and requested their IPD through the Public-Private Partnerships (PPPs) between the relevant academic societies and the pharmaceutical companies. The effect modification by baseline depression severity was examined through six increasingly complex competing mixed effects models for repeated measures.

Results: We identified eleven eligible trials and obtained IPD from six, which compared duloxetine, escitalopram, mirtazapine, paroxetine or bupropion against placebo (total n=2464). The best-fitting model revealed that the interaction between baseline severity and treatment was not statistically significant (coefficient=-0.04, 95% confidence interval: -0.16 to 0.08, p=0.49). Several sensitivity analyses confirmed the robustness of the findings.

Conclusion: We may expect as much benefit from antidepressant treatments for mild, moderate or severe major depression. Clinical practice guidelines will need to take these findings into consideration.

Keywords:

Depression; Antidepressives; Meta-analysis

Summations:

- Individual participant data meta-analysis found no influence of baseline depression severity on the efficacy of new generation antidepressants.
- Across the spectrum of baseline severity, we can expect similar magnitude of improvement among patients with major depression through their acute phase treatment.
- Clinical practice guidelines which currently provide differential therapeutic guidance according to the baseline disorder severity will need to take the current

findings into consideration.

Considerations:

- We were able to obtain IPD from six of the eleven eligible trials.
- All were phase II or III trials and the participants were highly selected patients with major depression consenting to placebo-controlled trials.

INTRODUCTION

Matching treatments with individual patients beyond the traditional diagnostic categorisation is receiving renewed interest (1, 2). Individuals with psychiatric disorders vary considerably in their responses to treatments even within the same diagnostic entity. Identifying robust effect modifiers for these patients would aid considerably in clinical decision making.

The single best studied effect modifier in psychiatry is initial symptom severity. Older studies used to examine study-level mean symptom severity and treatment response: it is increasingly recognised that such studies suffer from several critical weaknesses including decreased power, use of different definitions and statistical analyses across studies, inconsistent handling of missing data and the ecological fallacy in which relationships observed at the group level may not reflect the true relationships at the individual level. Only individual participant data meta-analyses (IPD-MA) can overcome these shortcomings and detect effect modification by examining the statistical interaction between treatment and baseline severity in the model. Strong evidence is fast accumulating that baseline severity is an effect modifier for schizophrenia (3), autistic disorder (4) and mania (5). All found that the greater the initial severity, the greater the difference between the drug and the placebo after treatment.

For major depressive disorder, however, while the representative two national guidelines in USA and UK provide differential therapeutic guidance according to baseline disorder severity (6, 7), the role of baseline severity is controverted.

Four studies have so far reported analyses based on IPD examining the influence of baseline depression severity on efficacy of antidepressants. The first study pooled results from 15 placebo-controlled studies conducted at the authors' center (329 patients) and found a statistically significant interaction between treatment and baseline severity in analysis of variance (8). The second study conducted IPD-MA on six studies (718 patients) comparing antidepressants and placebo and found that the superiority of medication over placebo increased as baseline depression severity rose (9). Subsequently, however, a much larger IPD-MA (37 trials, 8477 patients) found no influence of baseline severity on treatment efficacy (10). In another IPD-MA (34 trials, 10737 patients) the trial-level analysis suggested a significant interaction between treatment (drug or placebo) and baseline severity: however this was a case of ecological fallacy, because the interaction was not statistically significant when patient-level data were used (11).

Unfortunately, however, three of these four studies (8, 9, 11) used last-observation-

carried-forward data, a method no longer considered to be appropriate for handling missing data (12). Moreover, two of these studies have included patients or interventions that do not exactly match the clinical question of the efficacy of antidepressants in major depressive disorder. The second study (9) included a study with minor depression, a condition known to be unresponsive to antidepressants (13): this one study targeting minor depression (14) had a distinctively low mean baseline depression score (approximately 14 on 17-item Hamilton Rating Scale for Depression (HRSD)) in comparison with the other studies (approximately 20-25) and the observed smaller benefit for the less depressed patients may therefore be confounded by the presence of patients with minor depression. The fourth study (11) included trials examining quetiapine, which is usually not regarded as an antidepressant. Only Gibbons et al (10) applied a mixed effects model to account for missing data while taking advantage of repeated measurements. The last study, however, included studies with fluoxetine and venlafaxine only, and the generalisability of their findings remains to be elucidated.

We have planned IPD analyses of placebo-controlled antidepressant trials conducted in Japan as one of the projects within the Public-Private Partnerships (PPPs) between the International College of Neuropsychopharmacology (CINP), the Japanese Society of Neuropsychopharmacology and the pharmaceutical industry in Japan with interests in the central nervous system. The overall aim of the PPPs is to bring together and analyse the extant datasets owned individually by the involved parties in order to scrutinise the response patterns and their predictors among patients with major depression in greater details.

Aims of the study

The present study aims to examine the influence of baseline symptom severity on the differences in change scores between various antidepressants and placebo, based on a set of IPD from double-blind placebo-controlled randomised controlled trials in Japan. We limited the trial eligibility to those conducted in Japan in order to minimise possibilities of missing the relevant trials and to maximise the likelihood to obtain their IPD.

METHODS

This systematic review has been registered at PROSPERO (registration number CRD42017055912). The reporting follows the PRISMA extension guideline for individual participant data meta-analysis (15).

Search strategy and selection criteria

The eligibility criteria for trials to be included in the present IPD-MA were as follows:

i) Double-blind randomised controlled trial;

ii) Acute phase treatment of adults diagnosed as major depressive disorder according to standard operationalised diagnostic criteria;

iii) The intervention was a new generation antidepressant implemented as monotherapy, delivered within the licensed dose range (16);

iv) The control was pill placebo;

v) The trials were conducted in Japan.

We identified all such trials in the recently published comprehensive network meta-analysis of 21 antidepressants (16). In this network meta-analysis we searched the Cochrane Central Register of Controlled Trials, CINAHL, Embase, LILACS database, MEDLINE, MEDLINE In-Process, PsycINFO, AMED, the UK National Research Register, and PSYNDEX from the date of their inception to Jan 8, 2016, with no language restrictions. The electronic search was supplemented with manual searches for published, unpublished and ongoing RCTs in the drug-approval agencies including the Pharmaceuticals and Medical Devices Agency (PMDA) in Japan. We also undertook searches for published, unpublished and ongoing studies in a range of research registries including company websites. Studies were identified using search terms for depression (depress* or dysthymi* or adjustment disorder* or mood disorder* or affective disorder or affective symptoms) combined with the list of generic and commercial names of the antidepressants. After identifying the relevant trials conducted in Japan, we made further inquiries of experts in this field in Japan, and asked all the identified pharmaceutical companies to provide their protocols, codebooks and individual-level data.

There were 11 eligible trials examining nine different antidepressants from six pharmaceutical companies in Japan. All were phase II or III trials aiming at regulatory approval. Of these, four companies have agreed to provide the requested data representing six trials examining five antidepressants: three companies have provided

anonymised individual-level data for four trials examining duloxetine, escitalopram, mirtazapine and paroxetine (17-20), and one provided access to such data for two trials examining bupropion and paroxetine (21, 22) at their dedicated internet portal.
As it was possible to run all the detailed analyses only with a dataset directly available to us and not at a remote portal, we chose the first four trials as our primary analyses. We used the latter two trials to examine the generalisability of our primary findings.
We used all data that were included in the regulatory submission and the clinical study report as such.

Risk of bias of included studies

Two independent reviewers assessed the risk of bias of the included studies in accordance with the Cochrane Handbook for Systematic Reviews of Interventions (23). Any disagreement was resolved through discussion and, where necessary, in consultation with the other review team members.

Statistical analyses

We conducted IPD-MA to investigate the relationship between baseline symptom severity and subsequent symptom change in the comparison of antidepressants versus placebo in the acute treatment of major depressive disorder.
Four trials used 17-item HRSD (24, 25) as the primary outcome, while two used Montgomery-Asberg Depression Rating Scale (MADRS) (26). We converted the latter into the former by using the conversion algorithm based on the item response theory (27). The validity of this conversion was tested in a sensitivity analysis.
We used the approach developed by Hedeker and Gibbons, and applied a three-level mixed-effects model repeated measures analysis with restricted maximum likelihood estimation (3-5, 10, 28). The levels accounted for the data structure such that level 1 represented visit, level 2 the participant, and level 3 the trial. The following competing models with increasing complexity were tested: model 1, visit, treatment, and the two-way visit*treatment interaction; model 2, model 1 plus all the two-way interactions between visit, treatment and baseline; model 3, model 2 plus the three-way interaction of visit*treatment*baseline. For all models, we included baseline severity as covariate (29) and treated visit as categorical variable: individuals, trials and drugs were treated as random effects. These models were tested unadjusted and adjusted for confounders (i.e. age and sex). The model with the smallest Bayesian Information Criterion (BIC) was

chosen as the most parsimonious (30). The effect of interest was the interaction between treatment and baseline score.

All reported p values are two-tailed and $p<0.05$ was chosen as the threshold for statistical significance. We used PROC MIXED in SAS version 9.4 and RevMan version 5.3

Sensitivity analyses

We conducted the following sensitivity analyses to examine the robustness of the primary findings.

(i) By applying the best-fitting model found in the primary analyses to the two further trials to ascertain if the findings were replicable.

(ii) By including only measurement time points at which all studies reported the outcomes.

(iii) By excluding studies in which we converted MADRS scores into HRSD scores.

(iv) By using 6-item HRSD (31) which has been shown to be more sensitive to change than 17-item HRSD (32, 33)

Investigation of heterogeneity

We assessed the heterogeneity of the estimates of the treatment*baseline interaction across the six included studies by Q-test and I-squared.

Data availability bias

In order to test out if the included studies were a biased subsample of the total eligible sample (34), we examined the subgroup heterogeneity between the six trials whose IPD were available and the remaining five whose IPD were not, using the aggregate data meta-analysis.

RESULTS

Characteristics of the included trials

Table 1 tabulates the characteristics of the six included studies. One study (22) included an arm that administered the drug below the licensed dose range; this arm was excluded from the following analyses. The included participants were mainly in their 30s and 40s, slightly more women were involved, and their average baseline HRSD-17scores were around 22 or 23. The trials lasted between 6-8 weeks.

Table S1 shows the risk of bias of the six included studies. All the studies were rated at low or unclear risk of bias in the assessed domains.

Primary analyses

The best fitting mixed effects model was model 2, unadjusted for confounders (Table 2). In this model, the baseline depression severity by treatment interaction was not statistically significant (coefficient=-0.04, 95%CI: -0.16 to 0.08, p=0.49), where the negative coefficient would denote that the higher the baseline score, the greater the superiority of the drug over the placebo. Table S2 in the supporting information tabulates the coefficients for all the fixed effects in the best fitting mixed effects model.

Table 3 shows the estimated least-squares mean differences between the antidepressants and the placebo according to model 2, and Figure depicts scatterplots of all observed scores, superimposed with regression lines for the antidepressants and placebo arms at week 8. Figure S1 shows similar scatterplots at every visit. The antidepressants are statistically significantly superior to the placebo since the 4th through the 8th week, and the estimated regression lines are essentially parallel throughout these visits.

Sensitivity analyses

Table 4 shows the parameter estimates for the interaction between treatment and baseline score in the four sensitivity analyses. None of the sensitivity analyses was suggestive of significant interaction.

Examination of heterogeneity

There was no indication of substantial heterogeneity among the six included trials in the estimated interaction coefficients (chi-squared=7.71, df=5, p=0.17; I-squared=35%). (Figure S2)

Data availability bias

The test for subgroup differences between the included six studies and the non-included five studies was not statistically significant (chi-squared=0.71, df=1, p=0.40; I-squared=0%). (Figure S3)

DISCUSSION

Individual participant data meta-analysis of six trials (n=2464) comparing five new generation antidepressants against placebo in Japan showed no evidence of effect modification by baseline depression severity. Several sensitivity analyses confirmed the primary findings.

The lack of effect modification by baseline severity in depression is not only concordant with the similar analyses for fluoxetine and venlafaxine by Gibbons et al (10) but also for cognitive-behaviour therapy (CBT) for major depression. Furukawa et al (35) analysed IPD of all randomised trials comparing CBT against pill placebo and found that baseline depression severity did not influence the difference in changes in depression scores between CBT and pill placebo. Bower et al (36) examined low intensity interventions including self-help booklets and internet programs for minor and major depression and detected a statistically significant interaction between severity and treatment effect; however, the interaction was small and they concluded that both less and more severe patients could derive similar benefit from such low intensity interventions.

By contrast, our findings for major depression are in stark contrast with those for schizophrenia, autism or mania. One possible explanation for this discrepancy is that depression is more heterogeneous than schizophrenia, autism or mania, leading to grater sources of heterogeneity in treatment effects which may have masked the influence of baseline severity. A second reason could be the differences in effect sizes of the interventions for these disorders. The effect size for schizophrenia positive symptoms was approximately 0.6 (3), that for negative symptoms 0.4 (3), that for autism irritability 1.2 (4) and that for mania 0.6 (5), while for depression it was statistically significant but was only around 0.3 in Gibbons et al's meta-analysis (10), 0.2 in the case of CBT (35) and, in our analyses, also between 0.2 and 0.3 for the examined antidepressants (0.20 on 17-item HRSD and 0.27 on 6-item HRSD). A third possible explanation is the differences in the scores that can be expected for the healthy people: fully recovered people would score zero, or close to zero, on the symptom scales of positive or negative symptoms of schizophrenia, agitation in autism, or manic symptoms in mania, whereas even remitted people would often score between 5 and 10 on HAMD (37). Lastly, the differences in adverse effects may also partially explain the discrepancies between depression on one hand and schizophrenia, autism or mania on the other, as adverse effects due to

antidepressants or psychotherapies to treat the former disorder are allegedly less severe than those due to antipsychotics to treat the latter disorders and the latter adverse effects would affect both mildly and severely ill people to more or less the same degree thus diminishing possible benefits of treatments for the mildly ill people.

There are several weaknesses to the present study. First, for logistic reasons we limited the eligible trials to those conducted in Japan. Even then we were able to obtain IPD from only 6 out of 11 eligible trials. However, there was no evidence of data availability bias in terms of the effect sizes of the included versus the not included studies. It must also be noted that we were able to include five different new generation antidepressants in our analyses. Second, subsequent to the above eligibility criteria, the participants were limited to highly selected patients consenting to phase II and III trials in Japan. This means that our sample did not include very severe, suicidal patients or treatment-refractory patients. In addition, while we were able to include several new generation antidepressants, we could not include tricyclic antidepressants or monoamine oxidase inhibitors which would be administered in severe and treatment-refractory patients. The present results can therefore not rule out a possibility that patients with such very severe depressive symptoms might potentially derive a higher benefit from some antidepressants than those typically included in phase II or III trials of new generation antidepressants. The generalisability of our findings to less restrictive samples in the "psychiatric real-world" therefore needs be examined in more pragmatic trials with broader eligibility criteria in and outside Japan.

The main strength of our analysis is the use of individual patient data, with a sample size of 2464 patients, whose baseline 17-item HRSD scores ranged up to 42. The sensitivity analyses confirmed the robustness of our findings. In conjunction with Gibbons et al (10), it must be emphasised that both of the analyses with higher methodological rigor in accounting for missing data through the mixed effects models concluded that there was no appreciable influence of baseline depression severity on the efficacy of various antidepressants over placebo.

Clinically, the implications of the current findings may be summarised as follows. Patients would benefit equally through the whole spectrum of severity so long as they suffer from major depression. The myth of specifically smaller benefit of antidepressants for the milder spectrum of the disorder in comparison with its severer spectrum must now be expelled. Given the wide clinical implications of the present findings, it is worthwhile to examine their generalisability, i.e. if baseline severity modifies the efficacy

of interventions of variable effect sizes in less restricted samples of patients with major depression, in other disorders in psychiatry, and outside psychiatry.

Acknowledgements

This study has been conducted within the Public-Private Partnerships (PPPs) between the International College of Neuropsychopharmacology (CINP), the Japanese Society of Neuropsychopharmacology, and pharmaceutical companies in Japan including GSK, Meiji Seika, Mochida, MSD, Pfizer, and Shionogi. This study was supported in part by a grant-in-aid from Japan Agency for Medical Research and Development (AMED) to TAF (grant number JP17dk0307072) and to SY (grant number JP17dm0107093: Strategic Research Program for Brain Sciences). GSK, Meiji Seika, Mochida and Shionogi provided permission to use their data. These entities had no role in study design, data collection, data analysis, data interpretation, or writing of the report. KM had full access to all the data in the study and TAF had final responsibility for the decision to submit for publication.

Declaration of interest

TAF has received lecture fees from Janssen, Meiji, Mitsubishi-Tanabe, MSD and Pfizer and research support from Mitsubishi-Tanabe and Mochida; HN has received lecture fees from Boehringer Ingelheim and Kyowa Hakko Kirin, and research support from Kyowa Hakko Kirin and GSK; KI has received a lecture fee from Otsuka; SZL received honoraria or research support or consultancy fees or travel support from Shire Pharmaceuticals, F. Hoffmann-La Roche and Eli Lilly; SL has received honoraria for consulting from LB Pharma, Lundbeck, Otsuka, TEVA, Geodon Richter, Recordati, LTS Lohmann, Boehringer Ingelheim, and for lectures from Janssen, Lilly, Lundbeck, Otsuka, SanofiAventis and Servier; AC received personal fees from Accord Healthcare, all outside the submitted work. All the other authors report no conflict of interest.

REFERENCES

1. COLLINS FS, VARMUS H. A new initiative on precision medicine. N Engl J Med. 2015 Feb 26;372:793-5.
2. JAMESON JL, LONGO DL. Precision medicine--personalized, problematic, and promising. N Engl J Med. 2015 Jun 4;372:2229-34.
3. FURUKAWA TA, LEVINE SZ, TANAKA S, et al. Initial severity of schizophrenia and efficacy of antipsychotics: Participant-level meta-analysis of 6 placebo-controlled studies. JAMA psychiatry. 2015 Jan 1;72:14-21.
4. LEVINE SZ, KODESH A, GOLDBERG Y, et al. Initial severity and efficacy of risperidone in autism: Results from the RUPP trial. European Psychiatry. 2016;32:16-20.
5. SAMARA MT, GOLDBERG Y, LEVINE SZ, et al. Initial symptom severity of bipolar I disorder and the efficacy of olanzapine: a meta-analysis of individual participant data from five placebo-controlled studies. Lancet Psychiatry. 2017 Nov;4:859-67.
6. AMERICAN PSYCHIATRIC ASSOCIATION. Practice guideline for the treatment of patients with major depressive disorder (third edition). American Psychiatric Association. Am J Psychiatry. 2010;167:1-152.
7. NICE. Depression: the treatment and management of depression in adults (partial update of NICE clinical guideline 23). London: National Institute for Clinical Excellence; 2009.
8. KHAN A, BRODHEAD AE, KOLTS RL, BROWN WA. Severity of depressive symptoms and response to antidepressants and placebo in antidepressant trials. J Psychiatr Res. 2005 Mar;39:145-50.
9. FOURNIER JC, DERUBEIS RJ, HOLLON SD, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. JAMA. 2010 Jan 6;303:47-53.
10. 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 Jun;69:572-9.
11. RABINOWITZ J, WERBELOFF N, MANDEL FS, MENARD F, MARANGELL L, KAPUR S. Initial depression severity and response to antidepressants v. placebo: patient-level data analysis from 34 randomised controlled trials. Br J Psychiatry. 2016 May 19.
12. LITTLE RJ, D'AGOSTINO R, COHEN ML, et al. The prevention and treatment of missing data in clinical trials. N Engl J Med. 2012 Oct 4;367:1355-60.
13. BARBUI C, CIPRIANI A, PATEL V, AYUSO-MATEOS JL, VAN OMMEREN M. Efficacy of antidepressants and benzodiazepines in minor depression: systematic review and meta-analysis. The British journal of psychiatry : the journal of mental science. 2011 Jan;198:11-6, sup 1.
14. BARRETT JE, WILLIAMS JW, JR., OXMAN TE, et al. Treatment of dysthymia and minor depression in primary care: a randomized trial in patients aged 18 to 59 years. The Journal of family practice. 2001 May;50:405-12.
15. STEWART LA, CLARKE M, ROVERS M, et al. Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD Statement. JAMA. 2015 Apr 28;313:1657-65.
16. CIPRIANI A, FURUKAWA TA, SALANTI G, et al. Comparative efficacy and acceptability of first- and second-generation antidepressants in the acute treatment of major depressive disorder: A network meta-analysis. The Lancet. 2018.
17. HIRAYASU Y. A dose-response and non-inferiority study evaluating the efficacy and safety of escitalopram in patients with major depressive disorder: a placebo- and paroxetine-controlled, double-blind, comparative study. Rinsho Seishin Yakuri (Japanese Journal of Clinical Psychopharmacology). 2011;14:883-99.
18. HIRAYASU Y. A dose-response study of escitalopram in patients with major depressive disorder: a placebo-controlled, double-blind study. Rinsho Seishin Yakuri (Japanese Journal of Clinical Psychopharmacology). 2011;14:871-82.
19. HIGUCHI T, MURASAKI M, KAMIJIMA K. Clinical evaluation of duloxetine in the treatment of major depressive disorder: placebo- and paroxetine-controlled double-blind comparative study. Japanese Journal of Clinical Psychopharmacology [in Japanese]. 2009;12:1613-34.
20. KINOSHITA T. A double-blind, placebo-controlled study of a new antidepressant, mirtazapine, in depressed patients. Japanese Journal of Clinical Psychopharmacology [in Japanese].

2009;12:289-306.
21. HIGUCHI T, HONG JP, JUNG HY, WATANABE Y, KUNITOMI T, KAMIJIMA K. Paroxetine controlled-release formulation in the treatment of major depressive disorder: a randomized, double-blind, placebo-controlled study in Japan and Korea. Psychiatry Clin Neurosci. 2011 Dec;65:655-63.
22. KOSHINO Y, BAHK WM, SAKAI H, KOBAYASHI T. The efficacy and safety of bupropion sustained-release formulation for the treatment of major depressive disorder: a multi-center, randomized, double-blind, placebo-controlled study in Asian patients. Neuropsychiatr Dis Treat. 2013;9:1273-80.
23. HIGGINS JP, ALTMAN DG, GOTZSCHE PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.
24. TABUSE H, KALALI A, AZUMA H, et al. The new GRID Hamilton Rating Scale for Depression demonstrates excellent inter-rater reliability for inexperienced and experienced raters before and after training. Psychiatry Res. 2007 Sep 30;153:61-7.
25. HAMILTON M. A rating scale for depression. J Neurol Neurosurg Psychiatry. 1960 Feb;23:56-62.
26. MONTGOMERY SA, ASBERG M. A new depression scale designed to be sensitive to change. Br J Psychiatry. 1979 Apr;134:382-9.
27. CARMODY TJ, RUSH AJ, BERNSTEIN I, et al. The Montgomery Asberg and the Hamilton ratings of depression: a comparison of measures. Eur Neuropsychopharmacol. 2006 Dec;16:601-11.
28. HEDEKER D, GIBBONS RD. Longitudinal Data Analysis. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2006.
29. BORM GF, FRANSEN J, LEMMENS WA. A simple sample size formula for analysis of covariance in randomized clinical trials. J Clin Epidemiol. 2007 Dec;60:1234-8.
30. SCHWARZ G, GIDEON E. Estimating the dimension of a model. Annals of Statistics 1978;6:461-4.
31. BECH P, GRAM LF, DEIN E, JACOBSEN O, VITGER J, BOLWIG TG. Quantitative rating of depressive states. Acta Psychiatr Scand. 1975 Mar;51:161-70.
32. TIMMERBY N, ANDERSEN JH, SONDERGAARD S, OSTERGAARD SD, BECH P. A Systematic Review of the Clinimetric Properties of the 6-Item Version of the Hamilton Depression Rating Scale (HAM-D6). Psychother Psychosom. 2017;86:141-9.
33. OSTERGAARD SD. Do not blame the SSRIs: blame the Hamilton Depression Rating Scale. Acta neuropsychiatrica. 2017 Mar 1:1-3.
34. NEVITT SJ, MARSON AG, DAVIE B, REYNOLDS S, WILLIAMS L, SMITH CT. Exploring changes over time and characteristics associated with data retrieval across individual participant data meta-analyses: systematic review. BMJ. 2017 Apr 05;357:j1390.
35. FURUKAWA TA, WEITZ ES, TANAKA S, et al. Initial severity of depression and efficacy of cognitive-behavioural therapy: individual-participant data meta-analysis of pill-placebo-controlled trials. Br J Psychiatry. 2017 Mar;210:190-6.
36. BOWER P, KONTOPANTELIS E, SUTTON A, et al. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. BMJ. 2013;346:f540.
37. FURUKAWA TA, AKECHI T, AZUMA H, OKUYAMA T, HIGUCHI T. Evidence-based guidelines for interpretation of the Hamilton Rating Scale for Depression. J Clin Psychopharmacol. 2007 Oct;27:531-4.

Supporting information

Additional Supporting Information may be found in the online version of this article.

Figure S1. Observed and estimated changes in the 17-item Hamilton Rating Scale for Depression scores at each visit during the acute treatment of major depression

Figure S2. Heterogeneity in the estimates of treatment*baseline interaction coefficients among the six included studies

Figure S3. Heterogeneity in treatment efficacy estimates for the included and the non-included studies (Examination of data availability bias)

Table S1. Risk of bias of the included studies

Table S2. Coefficients for the fixed effects in the best fitting mixed effects model

Table 1. Trial and participant characteristics of the six included trials

	Trial characteristics		Participants characteristics						Number of observations						
	Drug tested	Phase	Key inclusion criteria	N	Age, Mean (Range)	Sex, No of women (%)	Baseline severity, Mean (Range)	No allocated to placebo (%)	Week1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 8
Derivation set	Duloxetine (40 mg/d, 60 mg/d) Paroxetine (flexible 20-40 mg/d) (19)	III	HRSD-17≥19, Depressed mood≥2, Placebo run-in after inclusion	438	38.3 (20-64)	198 (45.2)	20.4(7-35) [a]	145 (33.1)	438	423	413	405	0	391	0
	Escitalopram (10 mg/d, 20 mg/d) (18)	II	HRSD-17≥18, Depressed mood≥2	297	34.6 (20-64)	152 (51.2)	22.4(18-40)	100 (33.7)	296	285	0	277	0	265	259
	Escitalopram (10 mg/d, 20 mg/d) Paroxetine (flexible 20-40 mg/d) (17)	III	MADRS≥22, CGI-S≥4, Current episode≥4 weeks	484	36.2 (20-63)	255 (52.7)	19.9(8-40) [b]	124 (25.6)	483	468*	455*	451*	0	437*	423
	Mirtazapine (15 mg/d, 30 mg/d, 45 mg/d) (20)	III	HRSD-17≥18	263	39.7 (20-75)	132 (50.2)	22.6(18-36)	69 (26.2)	262	255	243	228	218	216	0
Subtotal				1482	37.1 (20-75)	737 (49.7)	21.0(7-40)	438 (29.6)	1479	1431	1111	1361	218	1309	682

Validation set	Paroxetine CR (25 mg/d, 50 mg/d) Paroxetine (20 mg/d, 40 mg/d) (21)	III	HRSD-17≥20, Depressed mood≥2	412	36.4 (20-70)	226 (54.9)	22.6(20-35)	171 (41.5)	399	395	393	379	0	367	360
	Bupropion SR (300 mg/d) (22)	III	HRSD-17≥20, CGI-S≥4, Current episode ≥8 weeks but <24 months	374	37.7 (20-64)	206 (55.1)	27.9(17-42) [c]	186 (49.7)	367*	345*	0	327*	0	318*	312*
Subtotal				786	37.0 (20-70)	432 (55.0)	25.2(17-42)	357 (45.4)	766	740	393	706	0	685	672
Total				2268	37.1 (20-75)	1169 (51.5)	22.5(7-42)	795 (35.0)	2245	2171	1504	2067	218	1994	1354

CGI-S: Clinical Global Impression Severity, HRSD-17: 17-item Hamilton Rating Scale for Depression, MADRS: Montgomery-Asberg Depression Rating Scale

*: HRSD-17 scores converted from MADRS

[a] There were patients with baseline HRSD-17 scores below 19 because this trial included a one-week single-blind placebo-run in between the subject eligibility check and the randomisation.

[b] The eligibility threshold was determined by MADRS while the baseline severity was measured by HRSD-17.

[c] There was one patient with baseline HRSD-17 score below the baseline eligibility threshold of 20 in the dataset provided: we used all data that were included in the regulatory submission and the clinical study report as such.

Table 2. Statistical significance of the fixed effects in the mixed effects models for the derivation set

Fixed effects	Not adjusted			Adjusted		
	Model 1	Model 2*	Model 3	Model 1	Model 2	Model 3
Allocation	0.012	0.976	0.782	0.012	0.910	0.721
Week	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Treatment×Visit	<0.001	<0.001	0.403	<0.001	<0.001	0.402
Baseline	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Treatment×Baseline	.	0.492	0.686	.	0.552	0.751
Baseline×Visit	.	<0.001	<0.001	.	<0.001	<0.001
Treatment×Baseline×Visit	.	.	0.782	.	.	0.782
Sex	.	.	.	0.025	0.029	0.029
Age	.	.	.	0.544	0.520	0.518
BIC	42914.4	42890.3	42907.7	42917.2	42893.3	42910.8

*: Best-fitting model according to Bayes Information Criterion

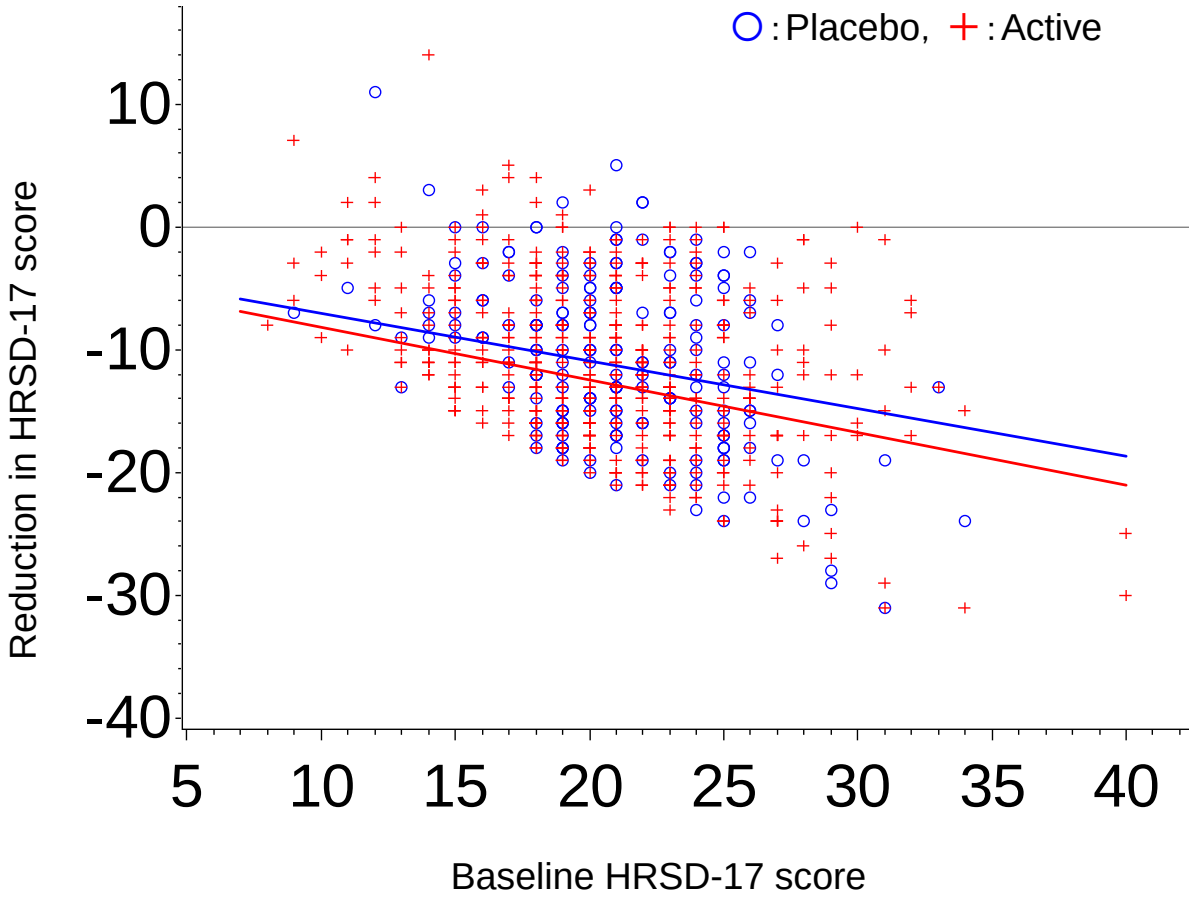
Table 3. Estimated least-squares mean differences between antidepressants and placebo

Visit	LS mean differences	95% CI	p value
Week 1	-0.11	-0.81 to 0.59	0.734
Week 2	-0.58	-1.28 to 0.12	0.096
Week 3	-0.73	-1.46 to 0.01	0.052
Week 4	-1.05	-1.75 to -0.34	0.007
Week 5	-1.23	-2.44 to -0.03	0.045
Week 6	-1.16	-1.87 to -0.45	0.003
Week 8	-1.62	-2.43 to -0.81	<0.001

Table 4. Parameter estimates for the interaction between treatment and baseline score in the three sensitivity analyses

Sensitivity analyses	Coefficient	95% CI	p value
Applying the model to the validation set	-0.03	-0.20 to 0.13	0.692
Including only visits with full outcome measurements	-0.04	-0.16 to 0.08	0.494
Excluding studies with converted HRSD values	0.00	-0.13 to 0.13	0.981
Using 6-item HRSD	0.04	-0.09 to 0.17	0.550

Figure. Observed and estimated changes in the 17-item Hamilton Rating Scale for Depression scores following 8-week acute treatment of major depression



Lines indicate the expected reduction in HRSD score with placebo (blue line) or antidepressants (red line) at 8 weeks on the vertical axis for corresponding baseline score on the horizontal axis. Each dot represents an individual patient either on placebo or antidepressants, with the observed reduction at 8 weeks on the vertical axis and the baseline score on the horizontal axis.

HRSD-17: 17-item Hamilton Rating Scale for Depression