Exploratory analyses of effect modifiers in the antidepressant treatment of major depression: Individual-participant data meta-analysis of 2803 participants in seven placebo-controlled randomized trials

RUNNING TITLE: Effect modifiers of the antidepressant pharmacotherapy

Hisashi Noma, PhD

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

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

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

Departments of Health Promotion and Human Behavior and of Clinical Epidemiology,
Kyoto University Graduate School of Medicine / School of Public Health, Kyoto, Japan
kiyomi.wb3@gmail.com

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

Kazutaka Ikeda, PhD

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

Shigeto Yamawaki, MD, PhD

Brain, Mind and Kansei Sciences Research Center, H, Hiroshima University, Hiroshima,
Japan

yamawaki@hiroshima-u.ac.jp

Andrea Cipriani, MD, PhD

Department of Psychiatry, University of Oxford, Oxford, UK

andrea.cipriani@psych.ox.ac.uk

*Corresponding author: Toshi A. Furukawa

Department of Health Promotion and Human Behavior,

Kyoto University Graduate School of Medicine / School of Public Health

Yoshida Konocho, Sakyo-ku, Kyoto 606-8501 JAPAN

Email: furukawa@kuhp.kyoto-u.ac.jp

Tel: +81-75-753-9491

ABSTRACT

Background: It is clinically important to know who are likely to respond more or less to antidepressants. However, meaningful effect modifiers (variables associated with differential response depending on the treatment) are yet to be identified.

Methods: We conducted individual participant data (IPD) meta-analysis and meta-regression to explore effect modifiers in placebo-controlled antidepressant trials conducted so far in Japan.

Results: We obtained access to IPD from seven placebo-controlled trials comparing bupropion, duloxetine, escitalopram, mirtazapine, paroxetine or venlafaxine with placebo in the acute phase treatment of major depression (total n=2803). The higher the guilt subscale score of the baseline Hamilton Rating Scale for Depression (HRSD), the greater the difference in reduction in depression severity between the antidepressants and placebo at week 6, while the older the current age or the age at onset, the smaller the difference. At week 8, the guilt subscale score of HRSD and presence of suicidal ideation at baseline predicted greater, and the anhedonia subscale and insomnia subscale scores of HRSD and early response at week 2 predicted smaller, difference in reduction.

Limitations: Different studies measured different sets of baseline variables and we were able to analyze only a limited set of candidate variables for effect modification.

Conclusion: Age, age at onset, several HRSD subscales including guilt, anhedonia and insomnia, presence of suicidal ideation at baseline and early response are potential effect modifiers for response to antidepressants in the acute phase antidepressant treatment of major depression. Future trials need to measure these and additional variables in concerted efforts to enable matching of treatments with individual characteristics in depression.

Keywords:

Depression; Antidepressive agents; Placebo; Individual participant data meta-analysis

INTRODUCTION

Since antiquity clinicians have aspired to match treatments with individuals' characteristics. Search for solid evidence for such matching characteristics in psychiatry, however, has seen little success so far (Cuijpers and Christensen, 2016; Holsboer, 2008; Ozomaro et al., 2013; Simon and Perlis, 2010).

Variables that may influence an individual's response to a treatment can be classified into two categories: effect modifiers (effect moderators) and prognostic factors (Furukawa et al., 2018). The former refers to baseline variables that predict differential response to treatments, while the latter refers to those that predict the overall course regardless of the treatments. In other words, effect modifiers affect the relative measures of treatment effect typically quantified as an odds ratio or a standardized mean difference, while prognostic factors affect the absolute response in the control condition usually expressed as a control event rate or an endpoint mean score. We need to identify the former in order to individualize the treatments and match treatments with individual patients.

Statistical methods to examine the effect modification are now reasonably well established (Kraemer et al., 2008; Willke et al., 2012). Currently, however, it is rare to have sufficient evidence to identify and replicate effect modifiers as one needs a very large sample size to detect a clinically significant interaction to establish effect modification (Kessler et al., 2017). Individual participant data (IPD) meta-analysis pooling several randomized controlled trials (RCTs) offers a hopeful approach to attain such power (Simmonds and Higgins, 2007). Unfortunately, however, different trials often measure different sets of baseline variables and, even when they measure the same construct, may use slightly different scales, definitions and/or categorizations (Rogozinska et al., 2017).

Needed then are concerted hypothesis-generating approaches to detect possible effect modifiers based on the currently available body of evidence, which then can inform future RCTs to enable more robust replication of identified effect modifiers and ultimately to lead to more powerful and cost-effective trial designs (Kraemer, 2013; Kraemer et al., 2006; Ondra et al., 2016).

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. In this study we would like to use this important database to explore and identify potential effect modifiers for response to antidepressants among patients with major depression.

METHODS

We followed the PRISMA extension guideline for individual participant data meta-analysis (Stewart et al., 2015). This systematic review is registered at PROSPERO (registration number CRD42017055912). The present study had been defined in the protocol of the main project before commencement.

Data source

We used the dataset from seven randomized controlled trials (RCTs) comparing mirtazapine, escitalopram, duloxetine, paroxetine controlled release, bupropion slow release, venlafaxine extended release with placebo or paroxetine in the treatment for major depression (Higuchi et al., 2011; Higuchi et al., 2016; Higuchi et al., 2009; Hirayasu, 2011a, b; Kinoshita, 2009; Koshino et al., 2013) (Table 1). We obtained individual participants data from pharmaceutical companies through the Public-Private Partnerships with the International College of Neuropsychopharmacology and the Japanese Society of Neuropsychopharmacology after complete anonymization. All were phase II or III trials aiming at regulatory approval. Of these, five companies have agreed to provide the requested data representing seven trials examining six antidepressants: three companies have provided anonymized individual-level data for four trials examining duloxetine, escitalopram, mirtazapine and paroxetine (Higuchi et al., 2009; Hirayasu, 2011a, b; Kinoshita, 2009), and two provided access to such data for three trials examining bupropion, paroxetine and venlafaxine (Higuchi et al., 2011; Higuchi et al., 2016; Koshino et al., 2013) at their respective internet portals. We used all data that were included in the regulatory submission and the clinical study report.

Two independent reviewers assessed the risk of bias of the included studies in accordance with the Cochrane Handbook for Systematic Reviews of Interventions (Higgins and Green, 2011). Any disagreement was resolved through discussion and, where necessary, in consultation with the other review team members.

Candidates of effect modifiers

Previous studies suggest many candidates for prognostic factors (variables associated with response regardless of the treatment) and for effect modifiers (variables associated with differential response depending on the treatment) in the treatment of depression (Kessler et al., 2017). In the original protocol we had planned to examine 18 variables encompassing demographics, life history, catamnesis, current symptomatology or therapeutic process based on the literature review (CRD42017055912, https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=55912) but in the current study were only able to analyze the following due the data availability in the seven trials: age, sex, age at onset, number of previous episodes, duration of current episode (weeks), suicide ideation, and early response. All the trials administered the 17-item Hamilton Rating Scale for Depression (HRSD) (Hamilton, 1960; Tabuse et al., 2007) at baseline. The individual item scores of HRSD constituted five subscales of anhedonia, guilt, bodily symptoms, appetite and insomnia (Furukawa et al., 2005). We were obliged to dichotomize the number of previous episodes as ≥ 3 times or not because this was how the information was coded in one trial while the other trials reported the number as a continuous variable. After patients with serious suicidal ideation had been excluded from the trials as per the protocols, the residual suicidal ideation was assessed with Mini International Neuropsychiatric Interview (MINI) (Sheehan et al., 1998). Because MINI was administered only in two trials, we were able to conduct the analyses for suicidal ideation only for these two trials (Hirayasu, 2011a, b). The early response was defined as more than 20% reduction of HRSD score within two weeks from the start of the trials (Szegedi et al., 2009).

Statistical analyses

We conducted IPD meta-analysis to examine the relationship between each candidate variable and the differences in change scores between the antidepressants and placebo in the acute treatment of major depressive disorder.

Five trials used 17-item HRSD (Hamilton, 1960; Tabuse et al., 2007) as the primary outcome, while two used Montgomery-Asberg Depression Rating Scale (MADRS) (Montgomery and Asberg, 1979). We converted the latter into the former by using the conversion algorithm based on the item response theory (Carmody et al., 2006).

At the internet portals used by the two companies, we could have access to individual-level data separately and could only download the summary statistics obtained by

analysis of each trial. Thus, it was impossible to run the straightforward one step IPD meta-analysis using all the datasets directly. Therefore, we adopted two step IPD meta-analysis for combining the results of regression analyses involving interaction terms for assessing effect modifications (Noma et al., 2019). At first, we conducted linear mixed-effects analysis using the restricted maximum likelihood estimation. The within-participant correlations were modelled by random effects. We modelled the differences in change of HRSD scores as the outcome variable, and modelled visit, treatment, the candidate covariate, all the two-way interactions of the three variables, and the three-way interaction of visit by treatment by baseline covariate as the explanatory variables. We treated visit as categorical variable (1st, 2nd, 3rd, 4th, 5th, 6th, 8th weeks). Then, we synthesized the estimated regression coefficients of the linear mixed effects models jointly using the multivariate meta-analysis methods. **The procedure takes account of the clustering within each trial.**

As the primary interest here is the interactions between treatment and the candidate covariates, we estimated regression slopes of the treatment group and the placebo group at 6th and 8th weeks based on the regression coefficients and their covariance matrix estimates using the parametric bootstrap. In addition, we estimated the difference of the two regression slopes and assessed the significance of the difference using the bootstrap test (Efron and Tibshirani, 1993). This bootstrap p-value corresponds to the interaction test of the treatment-by-covariate at the 6th and 8th weeks.

All reported p values are two-tailed and $p < 0.05$ was chosen as the threshold for statistical significance. We used R version 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria) and the *mvmeta* package for the multivariate meta-analysis (Gasparrini, 2018).

RESULTS

Characteristics of the included trials

Table 1 tabulates the study characteristics of the seven included studies. One study (Koshino et al., 2013) included an arm that administered the drug below the licensed dose range; this arm was excluded from the **subsequent** analyses. Excluding this arm, the seven RCTs included 2803 participants. Table 2 shows the demographic and clinical characteristics of the participants. The included participants were mainly in their 30s and 40s, slightly more women were involved, and their average baseline HRSD scores were around 22 or 23. The trials lasted between six to eight weeks. **The risk of bias of the included studies is shown in Table 3: all the studies were rated at low or unclear risk**

of bias in the assessed domains.

Evaluations of effect modifications

The estimated regression slopes of the antidepressant and placebo groups are presented in Table 4 (week 6) and Table 5 (week 8).

All of the candidate variables except for the episode frequency were significantly associated with at least one of the outcomes at either 6 or 8 weeks, i.e., these variables would have some prognostic values regardless of the antidepressant treatments. For example, the guilt subscale score positively predicted the change in depression severity, defined as baseline score – week 6/8 score, for both the antidepressant and the placebo groups: in other words, the greater the baseline guilt, the greater the reduction in symptom severity both in the antidepressant group and the placebo group.

When we examined the differences of the regression slopes between the two groups, there were several significant variables that indicated interactions with antidepressant treatments. The estimated difference of those variables can be interpreted as follows:

At week 6, the guilt subscale score positively predicted the difference in change in depression severity between the antidepressant and the placebo groups, defined as change on antidepressant – change on placebo: in other words, the greater the baseline guilt, the greater the difference in change in depression severity between the antidepressant and the placebo, in favor of the antidepressant group. One point increase in the guilt subscale score predicted greater superiority of the drugs over placebo by 0.18 (95% confidence interval [CI]: 0.04 to 0.33). By contrast, the greater the current age or the age at onset, the less efficacious the antidepressants are over placebo: for every 10-year increase in age or in age at onset, the superiority of the antidepressants over placebo will be smaller by 0.49 (0.07 to 0.92) or 0.56 (0.15 to 0.97) points in HRSD, respectively. At week 8, one point increase in the guilt subscale score meant greater difference in change between antidepressants and placebo by 0.15 (0.01 to 0.30) point, while one point increase in the anhedonia subscale score or on the insomnia subscale score meant smaller difference by 0.16 (0.04 to 0.28) points and 0.34 (0.12 to 0.56) points in terms of HRSD, respectively. If suicidal ideation was present at baseline or early response was observed at week 2, the difference between the antidepressants and placebo would be greater by 2.53 (0.50 to 4.54) and smaller by 1.25 (0.42 to 2.08) points, respectively.

DISCUSSION

The individual participant data meta-analyses and meta-regression analyses of 2803 patients with major depression have identified current age, age at onset, several HRSD subscale scores, suicidal ideation and early response as potential effect modifiers in their 6- to 8-week acute phase treatment with antidepressants in comparison with placebo. Increasing age, age at onset, anhedonia and insomnia subscale scores, and presence of early response predicted smaller superiority of antidepressants over placebo, while guilt subscale score and suicidal ideation at baseline predicted greater superiority, either at week 6 or week 8.

Age, age at onset, anhedonia, insomnia, current suicidality and early response have often been found to be prognostic factors (Kessler et al., 2017): this study confirmed that they not only predicted the symptom change regardless of the treatments but also influenced the difference in symptom changes between antidepressants and placebo. For example, it is only logical that early response predicts better absolute response later in the course of the treatment (Furukawa et al., in press; Henkel et al., 2009; Tadic et al., 2010): however, in this study we found that, while early response was observed similarly frequently in both antidepressant and placebo groups (57.5% vs 56.0%, Table 2) and predicted greater reduction in depression severity both in the antidepressant and placebo groups (statistically significantly positive regression coefficients for change in depression severity in each group both at week 6 and week 8, Tables 4 and 5), it was also a negative predictor for the difference in change scores between the two groups: when early response was present, the superiority of the antidepressant over placebo was smaller by 0.43 points (95%CI: 1.18 to -0.33) at week 6 and by 1.25 points (0.42 to 2.08) at week 8.

We were able to examine the influence of suicidal ideation at baseline only in two studies where this was measured in detail with MINI (n=781). Because patients with serious/imminent suicidality had been excluded in these placebo-controlled antidepressant trials, the findings indicate that residual suicidality still was an important effect modifier, such that when present the superiority of antidepressants over placebo was greater by 1.16 points (-0.84 to 3.16) at week 6 and by 2.53 points (0.50 to 4.54) at week 8. The findings are corroborated by the statistically significant moderating influence in the same direction by the HRSD guilt subscale which covered items for guilt, suicide, agitation, psychic anxiety and insight, at week 6 and 8.

The magnitude of effect modification for some variables may appear clinically small. For

example, the influence of age or age of onset was approximately 0.5 point in HRSD total score for every 10 years. However, in the included trials, the age of the participants ranged from 20s through 70s, and the superiority of the antidepressant over placebo would be reduced by about two points on HRSD at week 6 if the patients are in their 60s rather than in their 20s. The estimated regression coefficient for the guilt subscale was approximately 0.15. However, the guilt subscale score ranged between 0 to 12: a 10-point difference would mean a difference of 1.5 points in terms of HRSD total score reduction after 6 or 8 weeks. Because the standard deviation of HRSD is typically about 6 to 7 in the antidepressant trials (Cipriani et al., 2018), 1.5 or 2-point difference would translate into a difference in effect sizes by 0.2 to 0.3.

There are several weaknesses in the current analyses. First, it is to note that different covariates emerged as statistically significant or not, depending on whether we examined the outcomes at week 6 or week 8. **HRSD guilt subscale was the only variable found to be consistently significant both at week 6 and week 8.** When we examine their coefficients, however, their 95% confidence intervals overlapped greatly. Even with close to 3000 participants in the dataset, our sample size may not have been sufficient for the accurate estimation of effect modifications (interactions). **These inconsistencies highlight the need for larger and better-coordinated datasets to be collected through concerted efforts.** Secondly, we examined effect modifiers and prognostic factors between antidepressants and placebo, and not between different antidepressants. The included antidepressants had different mechanisms of actions or receptor/transporter binding profiles. We need data from head-to-head trials comparing different antidepressants in order to examine effect modification among different antidepressants. Lastly, models incorporating prognostic factors and effect modifiers are essentially prediction models and therefore need external validation to be trustworthy and be considered applicable in the clinical practices (Moons et al., 2015; Moons et al., 2014; Riley et al., 2016). **Our sample size was not sufficient to consistently identify strong prognostic factors or effect modifiers, let alone to allow external validation. The clinical applicability of the current study findings is therefore limited.**

The current study findings therefore remain exploratory. As such, however, they have implications for future research. As we have observed in this IPD meta-analysis, **we were able to examine only some of the pre-specified potential effect modifiers identified from the literature, the episode frequency was coded differently across trials, and the MINI semi-structured interview was conducted only in two out of seven trials.** Such body of

evidence is hard to harmonize and synthesize, **resulting in loss of power to detect meaningful variables even when IPD are available.** The current study has suggested several variables that are worthy to be measured in the future trials in depression. Only such concerted data collection and data sharing can lead to replicable prediction models, which ultimately need to be tested in a large randomized trial to show their added values.

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, MSD, Pfizer 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

HN has received lecture fees from Boehringer Ingelheim, ASKA and Kyowa Hakko Kirin, and research support from Kyowa Hakko Kirin and GSK, outside the submitted work. TAF reports personal fees from Meiji, Mitsubishi-Tanabe, MSD and Pfizer and a grant from Mitsubishi-Tanabe, outside the submitted work; TAF has a patent 2018-177688 pending. HI has received lecture fees from Mitsubishi-Tanabe outside the submitted work. KI has received lecture fees from Otsuka, Sumitomo Dainippon, Vistagen, and Atheneum Partners, outside the submitted work. AC received personal fees from Accord Healthcare, outside the submitted work. All the other authors report no conflict of interest.

REFERENCES

- Carmody, T.J., Rush, A.J., Bernstein, I., Warden, D., Brannan, S., Burnham, D., Woo, A., Trivedi, M.H., 2006. The Montgomery Asberg and the Hamilton ratings of depression: a comparison of measures. *Eur. Neuropsychopharmacol.* 16, 601-611.
- Cipriani, A., Furukawa, T.A., Salanti, G., Chaimani, A., Atkinson, L.Z., Ogawa, Y., Leucht, S., Ruhe, H.G., Turner, E.H., Higgins, J.P.T., Egger, M., Takeshima, N., Hayasaka, Y., Imai, H., Shinohara, K., Tajika, A., Ioannidis, J.P.A., Geddes, J.R., 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., Christensen, H., 2016. Are personalised treatments of adult depression finally within reach? *Epidemiol Psychiatr Sci*, 1-3.
- Efron, B., Tibshirani, R.J., 1993. *An Introduction to Bootstrap*. Chapman & Hall, London.
- Furukawa, T.A., Efthimiou, O., Weitz, E.S., Cipriani, A., Keller, M.B., Kocsis, J.H., Klein, D.N., Michalak, J., Salanti, G., Cuijpers, P., 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 meta-regression. *Psychother. Psychosom.* 87, 140-153.
- Furukawa, T.A., Kato, T., Shinagawa, Y., Miki, K., Fujita, H., Tsujino, N., Kondo, M., Inagaki, M., Yamada, M., in press. Prediction of remission in pharmacotherapy of untreated major depression: development and validation of multivariable prediction models. *Psychol. Med.*
- Furukawa, T.A., Streiner, D.L., Azuma, H., Higuchi, T., Kamijima, K., Kanba, S., Ozaki, N., Aoba, A., Murasaki, M., Miura, S., 2005. Cross-cultural equivalence in depression assessment: Japan-Europe-North American study. *Acta Psychiatr. Scand.* 112, 279-285.
- Gasparrini, A., 2018. *mvmteta: Multivariate and Univariate Meta-Analysis and Meta-Regression: R package version 0.4.11*.
- Hamilton, M., 1960. A rating scale for depression. *J. Neurol. Neurosurg. Psychiatry* 23, 56-62.
- Henkel, V., Seemuller, F., Obermeier, M., Adli, M., Bauer, M., Mundt, C., Brieger, P., Laux, G., Bender, W., Heuser, I., Zeiler, J., Gaebel, W., Mayr, A., Moller, H.J., Riedel, M., 2009. Does early improvement triggered by antidepressants predict response/remission? Analysis of data from a naturalistic study on a large sample of inpatients with major depression. *J. Affect. Disord.* 115, 439-449.
- Higgins, J.P., Green, S., 2011. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]* Available from www.cochrane-handbook.org.
- Higuchi, T., Hong, J.P., Jung, H.Y., Watanabe, Y., Kunitomi, T., Kamijima, K., 2011. 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.* 65, 655-663.
- Higuchi, T., Kamijima, K., Nakagome, K., Itamura, R., Asami, Y., Kuribayashi, K., Imaeda, T., 2016. A randomized, double-blinded, placebo-controlled study to evaluate the efficacy and safety of venlafaxine extended release and a long-term extension study for patients with major depressive disorder in Japan. *Int. Clin. Psychopharmacol.* 31, 8-19.
- Higuchi, T., Murasaki, M., Kamijima, K., 2009. 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]* 12, 1613-1634.
- Hirayasu, Y., 2011a. 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)* 14, 883-899.
- Hirayasu, Y., 2011b. 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)* 14, 871-882.
- Holsboer, F., 2008. How can we realize the promise of personalized antidepressant medicines? *Nat. Rev. Neurosci.* 9, 638-646.
- Kessler, R.C., van Loo, H.M., Wardenaar, K.J., Bossarte, R.M., Brenner, L.A., Ebert, D.D., de Jonge, P., Nierenberg, A.A., Rosellini, A.J., Sampson, N.A., Schoevers, R.A., Wilcox, M.A., Zaslavsky, A.M., 2017. Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. *Epidemiol Psychiatr Sci* 26, 22-36.
- Kinoshita, T., 2009. A double-blind, placebo-controlled study of a new antidepressant, mirtazapine,

in depressed patients. *Japanese Journal of Clinical Psychopharmacology* [in Japanese] 12, 289-306.

Koshino, Y., Bahk, W.M., Sakai, H., Kobayashi, T., 2013. 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* 9, 1273-1280.

Kraemer, H.C., 2013. Discovering, comparing, and combining moderators of treatment on outcome after randomized clinical trials: a parametric approach. *Stat. Med.* 32, 1964-1973.

Kraemer, H.C., Frank, E., Kupfer, D.J., 2006. Moderators of treatment outcomes: clinical, research, and policy importance. *JAMA* 296, 1286-1289.

Kraemer, H.C., Kiernan, M., Essex, M., Kupfer, D.J., 2008. How and why criteria defining moderators and mediators differ between the Baron & Kenny and MacArthur approaches. *Health Psychol.* 27, S101-108.

Montgomery, S.A., Asberg, M., 1979. A new depression scale designed to be sensitive to change. *Br. J. Psychiatry* 134, 382-389.

Moons, K.G., Altman, D.G., Reitsma, J.B., Ioannidis, J.P., Macaskill, P., Steyerberg, E.W., Vickers, A.J., Ransohoff, D.F., Collins, G.S., 2015. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann. Intern. Med.* 162, W1-73.

Moons, K.G., de Groot, J.A., Bouwmeester, W., Vergouwe, Y., Mallett, S., Altman, D.G., Reitsma, J.B., Collins, G.S., 2014. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med.* 11, e1001744.

Noma, H., Maruo, K., Goshio, M., Levine, S.Z., Goldberg, Y., Leucht, S., Furukawa, T.A., 2019. Efficient two-step multivariate random effects meta-analysis of individual participant data for longitudinal clinical trials using mixed effects models. *BMC Med. Res. Methodol.* 19, 33.

Ondra, T., Jobjornsson, S., Beckman, R.A., Burman, C.F., Konig, F., Stallard, N., Posch, M., 2016. Optimizing Trial Designs for Targeted Therapies. *PLoS ONE* 11, e0163726.

Ozomaro, U., Wahlestedt, C., Nemeroff, C.B., 2013. Personalized medicine in psychiatry: problems and promises. *BMC Med.* 11, 132.

Riley, R.D., Ensor, J., Snell, K.I., Debray, T.P., Altman, D.G., Moons, K.G., Collins, G.S., 2016. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ* 353, i3140.

Rogozinska, E., Marlin, N., Thangaratinam, S., Khan, K.S., Zamora, J., 2017. Meta-analysis using individual participant data from randomised trials: opportunities and limitations created by access to raw data. *Evid. Based Med.* 22, 157-162.

Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatry* 59 Suppl 20, 22-33;quiz 34-57.

Simmonds, M.C., Higgins, J.P., 2007. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. *Stat. Med.* 26, 2982-2999.

Simon, G.E., Perlis, R.H., 2010. Personalized medicine for depression: can we match patients with treatments? *Am. J. Psychiatry* 167, 1445-1455.

Szegedi, A., Jansen, W.T., van Willigenburg, A.P., van der Meulen, E., Stassen, H.H., Thase, M.E., 2009. Early improvement in the first 2 weeks as a predictor of treatment outcome in patients with major depressive disorder: a meta-analysis including 6562 patients. *J. Clin. Psychiatry* 70, 344-353.

Tabuse, H., Kalali, A., Azuma, H., Ozaki, N., Iwata, N., Naitoh, H., Higuchi, T., Kanba, S., Shioe, K., Akechi, T., Furukawa, T.A., 2007. The new GRID Hamilton Rating Scale for Depression demonstrates excellent inter-rater reliability for inexperienced and experienced raters before and after training. *Psychiatry Res.* 153, 61-67.

Tadic, A., Helmreich, I., Mergl, R., Hautzinger, M., Kohnen, R., Henkel, V., Hegerl, U., 2010. Early improvement is a predictor of treatment outcome in patients with mild major, minor or subsyndromal depression. *J. Affect. Disord.* 120, 86-93.

Willke, R.J., Zheng, Z., Subedi, P., Althin, R., Mullins, C.D., 2012. From concepts, theory, and evidence of heterogeneity of treatment effects to methodological approaches: a primer. *BMC Med. Res. Methodol.* 12, 185.

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

Trial	Trial characteristics					Number of observations						
	Drug tested	Phase	Key inclusion criteria	N	No allocated to placebo (%)	Week1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 8
(Higuchi et al., 2009)	Duloxetine (40 mg/d, 60 mg/d) Paroxetine (flexible 20-40 mg/d)	III	HRSD-17≥19, Depressed mood≥2, Placebo run-in after inclusion	438	145 (33.1)	438	423	413	405	0	391	0
(Hirayasu, 2011b)	Escitalopram (10 mg/d, 20 mg/d)	II	HRSD-17≥18, Depressed mood≥2	297	100 (33.7)	296	285	0	277	0	265	259
(Hirayasu, 2011a)	Escitalopram (10 mg/d, 20 mg/d) Paroxetine (flexible 20-40 mg/d)	III	MADRS≥22, CGI-S≥4, Current episode≥4 weeks	484	124 (25.6)	483	468*	455*	451*	0	437*	423
(Kinoshita, 2009)	Mirtazapine (15 mg/d, 30 mg/d, 45 mg/d)	III	HRSD-17≥18	263	69 (26.2)	262	255	243	228	218	216	0
(Higuchi et al., 2011)	Paroxetine CR (25 mg/d, 50 mg/d) Paroxetine (20 mg/d, 40 mg/d)	III	HRSD-17≥20, Depressed mood≥2	412	171 (41.5)	399	395	393	379	0	367	360

(Koshino et al., 2013)	Bupropion SR (300 mg/d)	III	HRSD-17 $\geq$ 20, CGI-S $\geq$ 4, Current episode $\geq$ 8 weeks but <24 months	374	186 (49.7)	367*	345*	0	327*	0	318*	312*
(Higuchi et al., 2016)	Venlafaxine ER (75 mg/d, flexible 75-225 mg/d)	III	MADRS $\geq$ 26, CGI-S $\geq$ 4, Current episode $\geq$ 90 days, Recurrent episode $\geq$ 28 days, QIDS-J $\geq$ 16	535	184 (34.4)	521	516	502	499	0	489	530
<i>Total</i>				<i>2803</i>	<i>979 (34.9)</i>	<i>2399</i>	<i>1874</i>	<i>1551</i>	<i>1788</i>	<i>218</i>	<i>1728</i>	<i>1572</i>

CGI-S: Clinical Global Impression Severity, HRSD-17: 17-item Hamilton Rating Scale for Depression, MADRS: Montgomery-Asberg Depression Rating Scale, QIDS-J: Quick Inventory of Depressive Symptomatology

*: HRSD-17 scores converted from MADRS according to Carmody et al (2006)

Table 2. Demographic characteristics of the participants of seven included trials.

		No. of trials	Antidepressant group (N=1824)	Placebo group (N=979)
Baseline severity (according to HRSD-17)		7	22.2 (7, 41)	22.9 (7, 42)
Age		7	37.2 (20, 76)	37.6 (20, 75)
Female sex		7	940 / 1824 (51.5%)	497 / 979 (50.8%)
Age at onset		7	33.9 (10.1, 75.6)	34.2 (12.1, 74.0)
Episode frequency (≥ 3 or not)		7	236 / 1824 (12.9%)	108 / 979 (11.0%)
Episode duration (weeks)		7	41.7 (0.6, 670)	40.3 (0.7, 680)
Early response ($\geq 50\%$ reduction at 2 weeks or not)		7	1041 / 1812 (57.5%)	547 / 977 (56.0%)
HRSD subscales	Anhedonia (score range: 0-14)	7	7.82 (2, 13)	8.02 (3, 14)
	Guilt (score range: 0-18)	7	5.50 (0, 12)	5.47 (0, 12)
	Bodily symptoms (score range: 0-10)	7	3.95 (0, 9)	4.10 (0, 9)
	Appetite (score range: 0-4)	7	1.35 (0, 4)	1.47 (0, 4)
	Insomnia (score range: 0-6)	7	3.52 (0, 6)	3.56 (0, 6)
Suicide ideation (present or not according to MINI) *		2	259 / 557 (46.5%)	115 / 224 (51.3%)

Quantitative data are given as mean (range), and categorical data are given as number/total (%). Missing data are excluded.

HRSD: Hamilton Rating Scale for Depression, MINI: Mini International Neuropsychiatric Interview (Sheehan et al., 1998)

*: available in two trials (Hirayasu, 2011a, b).

Table 3. Risk of bias of the included studies

Study	Sequence generation	Allocation concealment	Blinding participant and personnel	Blinding assessor	Selective reporting bias	Attrition bias
Higuchi 2009	Low risk of bias	Low risk of bias	Stated but not tested	Stated but not tested	Low risk of bias	Low risk of bias
Hirayasu 2011a	Unclear risk of bias	Unclear risk of bias	Stated but not tested	Stated but not tested	Low risk of bias	Low risk of bias
Hirayasu 2011b	Unclear risk of bias	Unclear risk of bias	Stated but not tested	Stated but not tested	Low risk of bias	Low risk of bias
Kinoshita 2009	Unclear risk of bias	Unclear risk of bias	Stated but not tested	Stated but not tested	Low risk of bias	Unclear risk of bias
Higuchi 2011	Unclear risk of bias	Unclear risk of bias	Stated but not tested	Unclear risk of bias	Low risk of bias	Low risk of bias
Koshino 2013	Unclear risk of bias	Low risk of bias	Stated but not tested	Stated but not tested	Low risk of bias	Unclear risk of bias
Higuchi 2016	Unclear risk of bias	Unclear risk of bias	Stated but not tested	Stated but not tested	Unclear risk of bias	Low risk of bias

Table 4. Estimated regression slopes of antidepressant and placebo groups, and their differences at Week 6.

		Change in depression severity in the antidepressant group			Change in depression severity in the placebo group			Difference in change between the antidepressant and the placebo groups			
		Slope	95% CI		Slope	95% CI		Estimate	95% CI		p-value
Baseline severity		0.550	0.495	0.606	0.582	0.506	0.658	−0.032	−0.127	0.063	0.503
Age		−0.013	−0.039	0.012	0.036	0.002	0.07	−0.049	−0.092	−0.007	0.048
Sex		−0.057	−0.468	0.352	−0.559	−1.114	−0.003	0.504	−0.187	1.197	0.153
Age at onset		−0.007	−0.032	0.018	0.049	0.016	0.082	−0.056	−0.097	−0.015	0.015
Episode frequency (≥3)		−0.104	−0.896	0.684	−1.071	−2.24	0.105	0.969	−0.445	2.376	0.179
Episode duration (weeks)		−0.007	−0.012	−0.002	−0.005	−0.013	0.002	−0.002	−0.011	0.007	0.711
Early response (≤ 2 weeks)		6.472	6.02	6.921	6.897	6.289	7.505	−0.425	−1.179	0.330	0.547
HRSD subscales	Anhedonia	0.035	−0.039	0.110	0.127	0.022	0.231	−0.092	−0.220	0.036	0.321
	Guilt	0.647	0.563	0.731	0.467	0.348	0.584	0.180	0.035	0.326	0.015
	Bodily symptoms	0.235	0.111	0.36	0.09	−0.089	0.267	0.145	−0.071	0.362	0.190
	Appetite	−0.295	−0.497	−0.093	−0.295	−0.567	−0.025	0.000	−0.338	0.336	0.999
	Insomnia	−0.469	−0.592	−0.344	−0.268	−0.432	−0.105	−0.201	−0.406	0.006	0.113
Suicide ideation		−0.027	−1.095	1.049	−1.194	−2.885	0.489	1.161	−0.841	3.158	0.254

95%CI: 95% confidence interval, HRSD: Hamilton Rating Scale for Depression

Table 5. Estimated regression slopes of antidepressant and placebo groups, and their differences at Week 8.

		Change in depression severity in the antidepressant group			Change in depression severity in the placebo group			Difference in change between the antidepressant and the placebo groups			
		Estimate	95% CI		Estimate	95% CI		Estimate	95% CI		p-value
Baseline severity		0.499	0.439	0.560	0.598	0.517	0.678	-0.098	-0.198	0.003	0.114
Age		-0.037	-0.065	-0.01	-0.002	-0.038	0.035	-0.035	-0.081	0.010	0.253
Sex		-0.796	-1.223	-0.37	-0.948	-1.526	-0.369	0.151	-0.569	0.868	0.680
Age at onset		-0.025	-0.052	0.002	0.017	-0.018	0.052	-0.042	-0.086	0.003	0.133
Episode frequency (≥ 3)		0.074	-0.78	0.922	-1.15	-2.454	0.15	1.233	-0.324	2.782	0.120
Episode duration (weeks)		-0.01	-0.015	-0.005	-0.006	-0.014	0.002	-0.004	-0.014	0.005	0.769
Early response (≤ 2 weeks)		5.288	4.79	5.785	6.536	5.874	7.196	-1.249	-2.077	-0.42	0.006
HRSD subscales	Anhedonia	-0.065	-0.136	0.006	0.095	-0.004	0.195	-0.16	-0.282	-0.038	0.020
	Guilt	0.515	0.43	0.602	0.366	0.245	0.486	0.149	0.001	0.296	0.048
	Bodily symptoms	-0.064	-0.191	0.064	-0.07	-0.246	0.105	0.006	-0.211	0.224	0.955
	Appetite	-0.398	-0.617	-0.181	-0.171	-0.45	0.109	-0.228	-0.581	0.125	0.411
	Insomnia	-0.590	-0.723	-0.457	-0.248	-0.422	-0.074	-0.343	-0.562	-0.123	0.005
Suicide ideation		0.235	-0.841	1.305	-2.286	-3.983	-0.589	2.526	0.499	4.540	0.014

95%CI: 95% confidence interval, HRSD: Hamilton Rating Scale for Depression