Published in final edited form as: *Int Clin Psychopharmacol.* 2018 Mar;33(2):66-72.
doi: 10.1097/YIC.0000000000000201

Efficacy of antidepressants over placebo is similar in two-armed versus three- or more-armed randomized placebo-controlled trials

Running head: Two-armed versus three- or more-armed antidepressants trials

Yusuke Ogawa[a], Toshi A. Furukawa[b], Nozomi Takeshima[a], Yu Hayasaka[a], Lauren Z. Atkinson[c], Shiro Tanaka[d], Andrea Cipriani[e], Georgia Salanti[f]

[a] Departments of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan
[b] Departments of Health Promotion and Human Behavior and of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan
[c] Oxford Centre for Human Brain Activity, Wellcome Centre for Integrative Neuroimaging, Department of Psychiatry, University of Oxford, Oxford, UK
[d] Department of Clinical Biostatistics, Kyoto University Graduate School of Medicine, Kyoto, Japan
[e] Department of Psychiatry, University of Oxford, Oxford, UK and Oxford Health NHS Foundation Trust, Warneford Hospital, Oxford, UK
[f] Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

Corresponding author
Toshi A. Furukawa, MD, PhD
Address: Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501 JAPAN
Email: furukawa@kuhp.kyoto-u.ac.jp
Tel: +81-75-753-9491

Conflicts of interest: We have the following potential conflicts of interest. TAF has received lecture fees from Eli Lilly, Janssen, Meiji, Mitsubishi-Tanabe, MSD and Pfizer and consultancy fees from Takeda Science Foundation. He has received research support from Mochida and Mitsubishi-Tanabe. NT has received lecture fees from Otsuka and Meiji. YH has received lecture fees from Yoshitomi. ST has received lecture fee from Kobe City, Astra-Zeneca, Taiho Pharmaceutical, and Ono Pharmaceutical. He has received consultation fee from the Pharmaceuticals and Medical Devices Agency, DeNA Life Science, and CanBus. He has received outsourcing fee from Public Health Research Foundation, Japan Breast Cancer Research Group, Satt, and Asahi Kasei Pharma. ST

Published in final edited form as: *Int Clin Psychopharmacol.* 2018 Mar;33(2):66-72.
doi: 10.1097/YIC.0000000000000201

has received grants from the Japan Agency for Medical Research and Development, the Japanese Ministry of Health Labor and Welfare, and the Japanese Ministry of Education, Science, and Technology. He engaged in a research project of the Japan Agency for Medical Research and Development. His wife had engaged in a research project of Bayer Yakuhin. ACi was expert witness for a patent issue about quetiapine extended release. All other authors report no conflict of interest.

Published in final edited form as: *Int Clin Psychopharmacol.* 2018 Mar;33(2):66-72.
doi: 10.1097/YIC.0000000000000201

Abstract

Objective

Previous studies reported effect sizes of antidepressants were larger in two- than in three- or more-armed ("multi-armed") randomized trials, where the probability to be allocated to placebo is smaller. However, these studies have not taken into account the publication bias, differences among antidepressants or covariance in multi-armed studies, or examined sponsorship bias.

Methods

We searched published and unpublished randomized controlled trials that compared placebo with 21 antidepressants for the acute treatment of major depression in adults. We calculated the ratio of odds ratios (ROR) of drug response over placebo in two-armed versus multi-armed trials for each antidepressant, and then synthesized RORs across all the included antidepressants using the multi-variate meta-analysis. Random effects model was used throughout.

Results

Two hundred fifty-eight trials (66 two-armed and 192 multi-armed trials; 80454 patients; 43.0% with unpublished data) were included in the present analyses. The pooled ROR for response of two-armed trials over multi-armed trials was 1.09 (95 %CI:

Published in final edited form as: *Int Clin Psychopharmacol.* 2018 Mar;33(2):66-72.
doi: 10.1097/YIC.0000000000000201

0.96 to 1.24). The ROR did not materially change between types of antidepressants, publication year or sponsorship.

Conclusion

The differences between two- versus multi-armed studies were much smaller than were suggested in previous studies and were not significant.

Key words: Systematic review; Meta-analysis; Antidepressants; Randomized controlled trial; Placebo-controlled trial; Trial design; Number of arms

2641 words

Published in final edited form as: *Int Clin Psychopharmacol.* 2018 Mar;33(2):66-72.
doi: 10.1097/YIC.0000000000000201

Introduction

Pharmacotherapy is the mainstay in today's treatment of major depression, and hundreds of randomized controlled trials (RCTs) of various antidepressants have been conducted so far to examine their efficacy (Furukawa *et al.*, 2016). Randomized, double-blind, placebo-controlled trials are required by regulatory agencies world-wide to obtain their approval for use with humans and are considered to be the gold standard for the evaluation of efficacy of antidepressants.

However, overestimation of drug efficacy in traditional placebo-controlled trials has been suggested when effect sizes (ESs) were compared between two-armed and three-armed RCTs. While the efficacy of the same antidepressant over placebo should not be different whether compared head-to-head against placebo or compared against another active drug along with placebo, the magnitude of the ES for antidepressants in three-armed RCTs was much smaller than those obtained in previous analyses that included two-armed trials (Greenberg *et al.*, 1992). These authors ascribed this difference to greater possibility of unblinding in two- versus multi-armed studies. Blinding may indeed be difficult to maintain in studies of psychotropic drugs because these drugs have characteristic side effects (Moncrieff *et al.*, 2004, Margraf *et al.*, 1991, Even *et al.*, 2000). When double-blindness is breached, drug efficacy over placebo would

probably be over-estimated (Leucht *et al.*, 2009).

Some reports have also suggested that antidepressant-placebo difference was negatively associated with the number of treatment arms (Khan *et al.*, 2004, Sinyor *et al.*, 2010, Papakostas and Fava, 2009). These authors implicated the role of expectancy which would lead to greater drug-placebo difference when the expectancy of receiving placebo is high.

All the above studies, however, have several problems. Firstly, previous meta-analyses have unfortunately often been subject to publication bias. Analysis of the trial data submitted to FDA as requirement of their submission process showed that only half of the phase II or III placebo-controlled trials had positive results, and most of the 'negative' trials had not been published (Turner *et al.*, 2008). The reported difference of ESs between two-armed and three-armed trials may be due to greater publication bias among the former, as the latter RCTs may be more likely to be published even when there is no significant difference between the antidepressant of interest and placebo because the publication can focus on the comparison between the two active drugs. Secondly, previous studies have generally assumed that ESs of antidepressants are the same among all antidepressants. However, it has been reported that they may be substantively different (Cipriani *et al.*, 2009). Therefore, intervention effects should be

Published in final edited form as: *Int Clin Psychopharmacol.* 2018 Mar;33(2):66-72.
doi: 10.1097/YIC.0000000000000201

examined and compared for each antidepressant separately. Thirdly, it has been demonstrated that an antidepressant appeared more effective when it was the new agent rather than the comparator, suggesting evidence of the so-called 'novelty effect' (Barbui *et al.*, 2004, Salanti *et al.*, 2010). The studies cited above (Greenberg *et al.*, 1992, Khan *et al.*, 2004, Sinyor *et al.*, 2010, Papakostas and Fava, 2009) have not taken this factor into account, so that the apparently bigger ES reported in two-armed studies might be due to 'novelty effect' of the agent which is more likely to be studied in two- rather than multi-armed trials when the agent is 'new' and when the trial is sponsored by the manufacturer of the drug.

The purpose of the present study is therefore to compare the odds ratios (OR) of antidepressants over placebo when examined in two-armed versus three- or more-armed (heretofore termed multi-armed) trials while taking into account possible differences among different antidepressants, based on a dataset compiled with as little publication bias as possible.

Methods

This is a secondary analysis of published and unpublished data from RCTs of antidepressants that was collected for GRISELDA, a multinational project to conduct

network meta-analyses of 21 new and old antidepressants for adult major depression. The details of the study methodology have been published (Furukawa *et al.*, 2016) and we hereby present its summary as relevant to this secondary analysis.

Criteria for considering studies for this review

All double-blind RCTs that compared placebo with the following selected first- and second-generation antidepressants as monotherapy for the acute phase treatment of depression were included: agomelatine, amitriptyline, bupropion, citalopram, clomipramine, desvenlafaxine, duloxetine, escitalopram, fluoxetine, fluvoxamine, levomilnacipran, milnacipran, mirtazapine, nefazodone, paroxetine, reboxetine, sertraline, trazodone, venlafaxine, vilazodone and vortioxetine. We included RCTs with patients aged 18 years or older, of both genders and with a primary diagnosis of unipolar major depression, diagnosed according to any standard operationalized diagnostic criteria.

Search methods for identification of studies

We searched Cochrane CENTRAL, CINAHL, EMBASE, LiLACS, MEDLINE, PSYCINFO, trial databases of the drug-approving agencies, trial registers and

Published in final edited form as: *Int Clin Psychopharmacol.* 2018 Mar;33(2):66-72.
doi: 10.1097/YIC.0000000000000201

homepages of pharmaceutical companies that market the included drugs up to Jan 8, 2016. The National Institute for Health and Care Excellence (UK) and the Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Germany) were also contacted. The reference lists of the identified RCTs and recent systematic reviews were checked. No language restriction was applied.

Data collection

Response to the treatment was defined as a reduction of at least 50% from baseline on the total score on Hamilton Rating Scale for Depression (HAM-D) (Hamilton, 1960), Montgomery-Asberg Depression Rating Scale (MADRS) (Montgomery and Asberg, 1979), or any other validated depression scale at the end of acute phase treatment. In the present review, acute treatment was defined as an 8-week treatment (Bauer *et al.*, 2002). If 8-week data were not available, we used data ranging between 4 to 12 weeks. When the number of responders was not reported but baseline mean and endpoint mean and standard deviation of the depression rating scales were provided, we calculated the number of responding patients employing a validated imputation method (Furukawa *et al.*, 2005).

Two researchers independently examined the titles and abstracts of all reports obtained

through the search strategy. Full articles of all the potentially eligible studies were then obtained and inspected by two review authors to identify trials meeting the review criteria. Data from each study were extracted into a structured data abstraction form independently by two researchers. Risk of bias were assessed for each included study using the Cochrane Collaboration 'risk of bias' tool (Higgins JP, 2011) by two independent researchers. Any disagreement was resolved through discussion or in consultation with a third member of the review team. Based on assessments of risks of bias for each domain, we quantified the overall risk of bias for each study as low risk if none of the domains was rated at high risk and three or fewer domains at unclear risk; as moderate risk if one domain was rated at high risk or none rated at high risk but four or more at unclear risk; or as high risk for all other cases.

Statistical analysis

For each antidepressant, we first estimated the overall odds ratios (ORs) of response between the antidepressant and placebo by synthesizing ORs from all two-armed or multi-armed comparisons by using the random effects model. We next estimated the ratios of odds ratios (RORs) and their variance of two-armed versus multi-armed trials for each antidepressant, and finally meta-analytically synthesized RORs across all the

Published in final edited form as: *Int Clin Psychopharmacol.* 2018 Mar;33(2):66-72.
doi: 10.1097/YIC.0000000000000201

included antidepressants using the random effects model. Random effects model was used throughout because of possible clinical heterogeneity across the included trials due to differences in clinical populations, drugs, and drug dosages. A summary ROR larger than 1 would mean that two-armed RCTs show larger intervention effects compared with placebo than multi-armed trials do. Because two or more antidepressants were involved in multi-armed studies, the summary RORs were correlated (the placebo arm is in common for such antidepressants) and we need to take account of these correlations; for example, the ROR for placebo versus agomelatine and the ROR for placebo versus paroxetine will be dependent because they include data from the same placebo arms in placebo vs agomelatine vs paroxetine trials. The synthesis of these RORs was therefore performed using a multivariate meta-analysis routine in R (rma.mv in the metafor package in R) after specifying the entire variance-covariance matrix (See appendix at the end of the article). We used Review Manager 5.3, Stata 14 and R to conduct the analyses.

We started to assess heterogeneity by visual inspection of the forest plots. We also calculated I^2 statistics (Higgins JP, 2011) and analyzed them on the basis of the Cochrane Handbook's recommendations (I^2 values of 0% to 40%: might not be important; 30% to 60%: may represent moderate heterogeneity; 50% to 90%: may

Published in final edited form as: *Int Clin Psychopharmacol.* 2018 Mar;33(2):66-72.
doi: 10.1097/YIC.0000000000000201

represent substantial heterogeneity; 75% to 100%: considerable heterogeneity).

Sensitivity analyses

In order to ascertain the robustness of our findings, we conducted the following sensitivity analyses.

1. By excluding studies at high risk of bias
2. By excluding studies where primary outcomes were imputed rather than reported
3. By using the fixed effect model instead of the random effects model

Subgroup analyses

We had a priori planned to conduct the following subgroup analyses.

1. numbers of arms in the multi-armed trials separately (three-armed, four-armed, and five-armed)
2. type of antidepressants (Tricyclic antidepressants (TCA) vs new generation antidepressants)
3. publication year (those published until the date of search, until 1990 (Greenberg *et al.*, 1992), and unpublished)
4. sponsorship (sponsored drug arms and non-sponsored drug arms in multi-armed

trials)

Results

Characteristics of included RCTs

Three-hundred-and-four placebo-controlled trials were identified by the electronic search. However, efficacy data were missing in 35 studies. There were no RCTs comparing milnacipran or clomipramine against placebo providing efficacy data. All placebo-controlled RCTs for fluvoxamine were three- or more-armed. We were therefore unable to calculate ROR for these three antidepressants. Altogether, 258 RCTs (80,454 patients) were finally included in the present analyses (Figure 1). Table 1 presents detailed characteristics for two-armed and multi-armed RCTs. Among the 258 RCTs included in this study, 66 (25.6%) were two-armed and 192 were multi-armed, including, 139 (53.9%) three-armed RCTs, 43 (16.7%) four-armed RCTs, and 10 (3.9%) five-armed RCTs. Median sample size of each active arm was 98.5 (first quartile, 43.5; third quartile, 158) for two-armed trials and 118.5 (first quartile, 66; third quartile, 157) for multi-armed trials. The median number of studies per antidepressant was 13.5 (range, 5 to 46). Figure 2 summarizes the risk of bias of the included studies. All in all, 46 studies were rated as being at low risk of bias, 214 at moderate risk of bias and 64 at

Published in final edited form as: *Int Clin Psychopharmacol.* 2018 Mar;33(2):66-72.
doi: 10.1097/YIC.0000000000000201

high risk of bias.

Differences in effect size between two-armed and multi-armed RCTs

Pooled response rates for the two treatment groups (antidepressants and placebo) were 45.8% and 31.4% in two-armed RCTs and 49.7% and 37.6% in multi-armed RCTs, respectively (Figure 3). There was no significant difference between two-armed and multi-armed RCTs in the OR of response between antidepressant and placebo (pooled ROR, 1.09; 95% CI 0.96 to 1.24) (Figure 4). The antidepressants are listed in the order of their approval. There was small to moderate heterogeneity in RORs across antidepressants (I^2 = 39.6%; 95% CI 0.0% to 65.6%). Because taking account of the covariance had little influence on the estimated ROR (the simple pooled ROR was 1.09 (95% CI 0.96 to 1.24, I^2=38.6%), the following sensitivity and subgroup analyses were conducted without accounting for the covariances due to multi-armed studies.

Sensitivity analyses

After exclusion of studies at high risk of bias, ROR was 1.06 (95%CI, 0.92 to 1.21; I^2 =34%). After exclusion of studies that imputed the number of responders, ROR was 1.06 (95% CI 0.90 to 1.25; I^2 =45%). Using the fixed effect model instead of the random effects

Published in final edited form as: *Int Clin Psychopharmacol.* 2018 Mar;33(2):66-72.
doi: 10.1097/YIC.0000000000000201

model, ROR was 1.09 (95% CI 0.99 to 1.19, I^2 =38.6%).

Subgroup analyses

The pooled ROR was 1.12 (95% CI 0.99 to 1.26; I^2 = 22%) for two-armed vs. three-armed RCTs, 1.03 (95% CI 0.87 to 1.22; I^2 = 33%) for two-armed vs. four-armed RCTs, and 1.10 (95% CI 0.84 to 1.43; I^2 = 36%) for two-armed vs. five-armed RCTs. ROR of TCA vs. placebo was 2.00 (95% CI 0.39 to 10.32) and that of new generation antidepressants vs. placebo was 1.09 (95% CI 0.96, 1.24; I^2 =41%). ROR was 1.08 (95% CI 0.93 to 1.25; I^2 =40%) based on the studies published up to the date of search (i.e. by excluding all unpublished studies), 2.34 (95% CI 0.57 to 9.66; I^2 =0%) based on the studies up to 1990 and 1.19 (95% CI 0.93 to 1.51; I^2 =0%) based on the studies which were not published. Similar results were obtained when the drug in multi-armed studies were marketed by the sponsor of the drug (ROR was 1.09; 95% CI 0.96 to 1.25; I^2 =32%) or when it was not (ROR was 1.07; 95% CI 0.90 to 1.28; I^2 =37%).

Discussion

The differences between the two- versus multi-armed studies were much smaller than found in previous studies and were not statistically significant. For this study we used

the data of the largest systematic review of antidepressants including 66 two-armed RCTs and 192 multi-armed RCTs, corresponding to 80,454 patients. Results of subgroup and sensitivity analyses did not alter this conclusion. RORs appeared larger for TCAs and for studies before 1990 but were not statistically significant either.

The differences between the previous studies and the present study may be explained as follows. First, the publication bias in our dataset is reduced as we were able to find unpublished information for 43.0% of the included studies through contacts with pharmaceutical companies and regulatory agencies. We were thus able to include the largest number of trials to date (258 trials), in comparison with 22 (Greenberg *et al.*, 1992), 52 (Khan *et al.*, 2004), 90 (Sinyor *et al.*, 2010) or 182 (Papakostas and Fava, 2009). Secondly, we employed the random effects model which produces wider 95% CI than the fixed effect model in the presence of heterogeneity. While overall the ORs tended to be bigger in two-armed studies than multi-armed ones, the differences did not reach statistical significance. We believe that our study had conducted a more methodologically rigorous synthesis by estimating the ROR for each antidepressant, and then meta-analytically pooling all the RORs of the included antidepressants, instead of assuming a common efficacy for all the included antidepressants. A sensitivity analysis employing the fixed effect model instead of the random effects

Published in final edited form as: *Int Clin Psychopharmacol.* 2018 Mar;33(2):66-72.
doi: 10.1097/YIC.0000000000000201

model confirmed the primary findings. Thirdly, the novelty effect (Barbui *et al.*, 2004, Salanti *et al.*, 2010) did not appear to be at play to explain the possible differences between two- versus multi-armed studies, because our subgroup analysis found little difference when the drug in multi-armed studies were marketed by the sponsor of the drug or when it was not.

Sinyor et al. (Sinyor *et al.*, 2010) showed that response rate for placebo was significantly higher in three-armed studies than in two-armed studies, so it is hard to show superiority of drugs in studies with more active treatment arms. While the placebo response rate in multi-armed studies was indeed larger than in two-armed studies in our dataset, so was the response rate on antidepressant drugs (Figure 3), resulting in the similar relative efficacy of drugs over placebo in both types of trials (Figure 4).

Our study has some limitations. We were unable to consider other trial and patient features that may have an impact on intervention effects, such as the difference of rating scales, countries and cultures, the proportion of melancholic depression, depression severity, and duration of the illness or the number of depressive episodes. Systematic differences in these characteristics between two- versus multi-armed studies might be playing a role, but we would need individual participant data to examine such effect modifiers. Moreover, given that the field of antidepressant trials in

Published in final edited form as: *Int Clin Psychopharmacol.* 2018 Mar;33(2):66-72.
doi: 10.1097/YIC.0000000000000201

the past has been prone to publication bias, we cannot completely rule out the possibility that some studies are still missing.

In summary, we found that intervention effects were not significantly different between two-armed and multi-armed RCTs. Our original hypotheses that possible breach of the double-blinding in antidepressant clinical trials or the lower expectancy for the active drug in two- rather than multi-armed trials would lead to overestimation of antidepressant efficacy was not borne out. Our results were different from those in the previous studies possibly because we appropriately took into account differences among different antidepressants through the random effects model and also because we were able to minimize the publication bias.

Acknowledgments

This work was supported by Japan Society for the Promotion of Science to YO (16K09033). ACi is supported by the NIHR Oxford cognitive health Clinical Research Facility. GS is a Marie Skłodowska-Curie fellow.

Published in final edited form as: *Int Clin Psychopharmacol.* 2018 Mar;33(2):66-72.
doi: 10.1097/YIC.0000000000000201

References

Barbui, C, Cipriani, A, Brambilla, P,Hotopf, M 2004. "Wish bias" in antidepressant drug trials? *J Clin Psychopharmacol,* 24**,** 126-30.

Bauer, M, Whybrow, PC, Angst, J, Versiani, M,Möller, HJ 2002. and World Federation of Societies Biological Psychiatry Task Force on Treatment Guidelines for Unipolar Depressive Disorders. World Federation of Societies of Biological Psychiatry (WFSBP) Guidelines for Biological Treatment of Unipolar Depressive Disorders, Part 1: Acute and continuation treatment of major depressive disorder. *World J Biol Psychiatry,* 3**,** 5-43.

Cipriani, A, Furukawa, TA, Salanti, G, Geddes, JR, Higgins, JP, Churchill, R*, et al.* 2009. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet,* 373**,** 746-58.

Even, C, Siobud-Dorocant, E,Dardennes, RM 2000. Critical approach to antidepressant trials. Blindness protection is necessary, feasible and measurable. *Br J Psychiatry,* 177**,** 47-51.

Furukawa, TA, Cipriani, A, Barbui, C, Brambilla, P,Watanabe, N 2005. Imputing response rates from means and standard deviations in meta-analyses. *Int Clin Psychopharmacol,* 20**,** 49-52.

Furukawa, TA, Salanti, G, Atkinson, LZ, Leucht, S, Ruhe, HG, Turner, EH*, et al.* 2016.

Comparative efficacy and acceptability of first-generation and second-generation antidepressants in the acute treatment of major depression: protocol for a network meta-analysis. *BMJ Open,* 6**,** e010919.

Greenberg, RP, Bornstein, RF, Greenberg, MD,Fisher, S 1992. A meta-analysis of antidepressant outcome under "blinder" conditions. *J Consult Clin Psychol,* 60**,** 664-9.

Hamilton, M 1960. A rating scale for depression. *J Neurol Neurosurg Psychiatry,* 23**,** 56-62.

Higgins JP, GS 2011. (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011] http://handbook-5-1.cochrane.org/ (accessed July 16 2017).

Khan, A, Kolts, RL, Thase, ME, Krishnan, KR,Brown, W 2004. Research design features and patient characteristics associated with the outcome of antidepressant clinical trials. *Am J Psychiatry,* 161**,** 2045-9.

Leucht, S, Corves, C, Arbter, D, Engel, RR, Li, C,Davis, JM 2009. Second-generation versus first-generation antipsychotic drugs for schizophrenia: a meta-analysis. *Lancet,* 373**,** 31-41.

Margraf, J, Ehlers, A, Roth, WT, Clark, DB, Sheikh, J, Agras, WS*, et al.* 1991. How "blind" are double-blind studies? *J Consult Clin Psychol,* 59**,** 184-7.

Moncrieff, J, Wessely, S,Hardy, R 2004. Active placebos versus antidepressants for depression. *Cochrane Database Syst Rev***,** CD003012.

Published in final edited form as: *Int Clin Psychopharmacol.* 2018 Mar;33(2):66-72.
doi: 10.1097/YIC.0000000000000201

Montgomery, SA,Asberg, M 1979. A new depression scale designed to be sensitive to change. *Br J Psychiatry,* 134**,** 382-9.

Papakostas, GI,Fava, M 2009. Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD. *Eur Neuropsychopharmacol,* 19**,** 34-40.

Salanti, G, Dias, S, Welton, NJ, Ades, AE, Golfinopoulos, V, Kyrgiou, M*, et al.* 2010. Evaluating novel agent effects in multiple-treatments meta-regression. *Stat Med,* 29**,** 2369-83.

Sinyor, M, Levitt, AJ, Cheung, AH, Schaffer, A, Kiss, A, Dowlati, Y*, et al.* 2010. Does inclusion of a placebo arm influence response to active antidepressant treatment in randomized controlled trials? Results from pooled and meta-analyses. *J Clin Psychiatry,* 71**,** 270-9.

Turner, EH, Matthews, AM, Linardatos, E, Tell, RA,Rosenthal, R 2008. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med,* 358**,** 252-60.

Published in final edited form as: *Int Clin Psychopharmacol.* 2018 Mar;33(2):66-72.
doi: 10.1097/YIC.0000000000000201

Figure legends

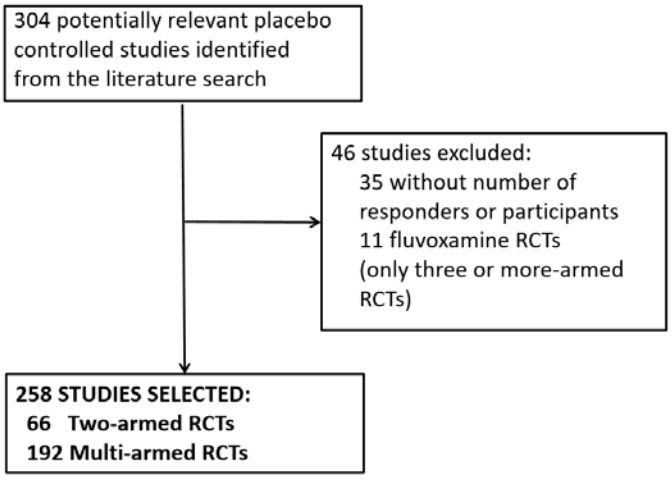

Figure. 1. Flow diagram; Abbreviations: RCTs: randomized controlled trials.

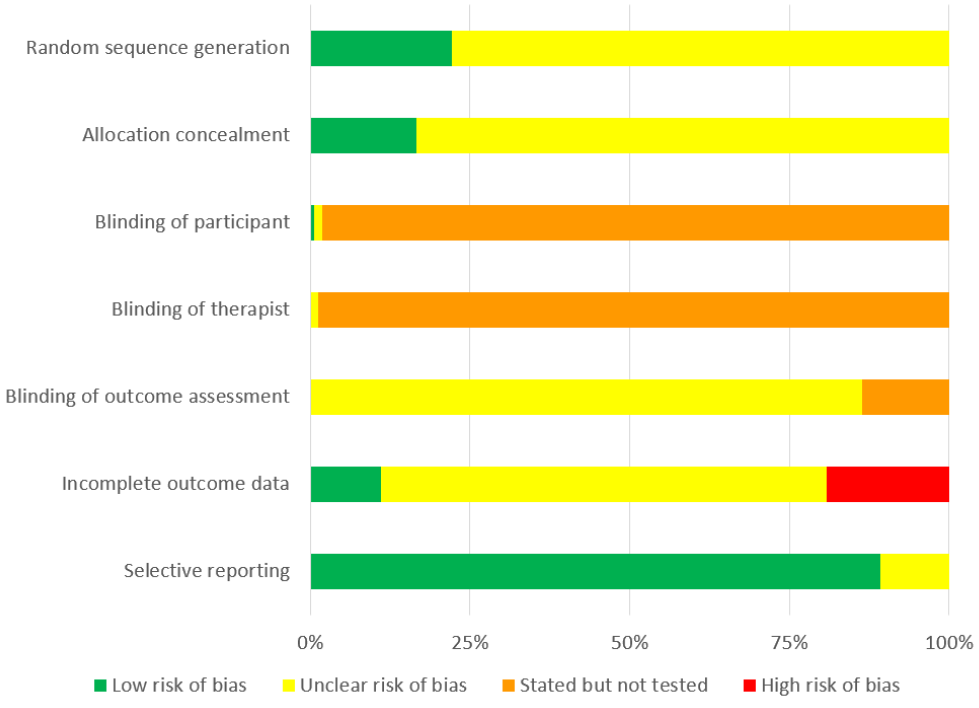

Figure 2. 'Risk of bias' graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.

Published in final edited form as: *Int Clin Psychopharmacol.* 2018 Mar;33(2):66-72.
doi: 10.1097/YIC.0000000000000201

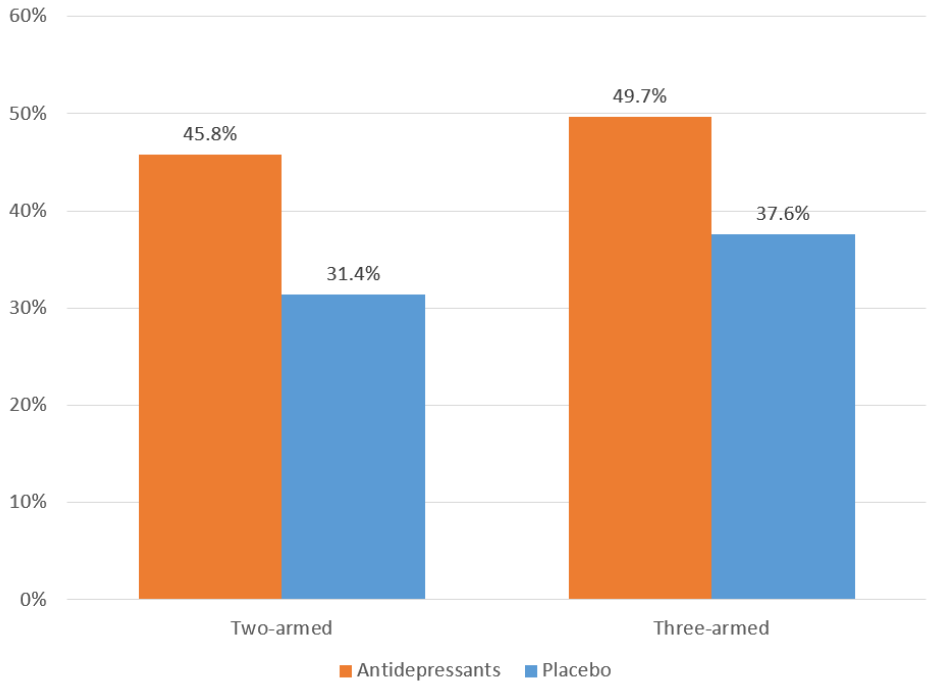

Figure 3. Antidepressant and placebo response rates

Published in final edited form as: *Int Clin Psychopharmacol.* 2018 Mar;33(2):66-72.
doi: 10.1097/YIC.0000000000000201

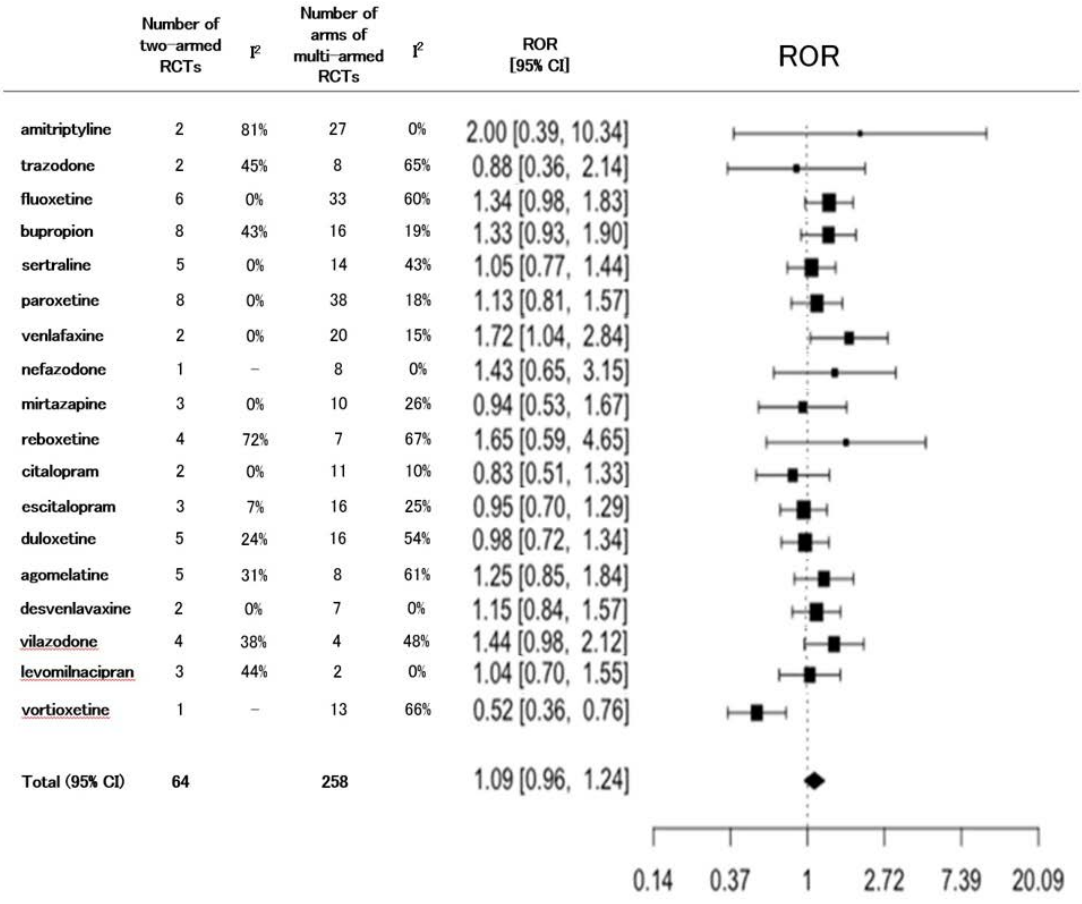

Figure 4. ROR between two-armed and multi-armed RCTs

The antidepressants are listed in the order of their approval.

Published in final edited form as: *Int Clin Psychopharmacol.* 2018 Mar;33(2):66-72.
doi: 10.1097/YIC.0000000000000201

Table 1. Characteristics of two-armed and three or more-armed RCTs

		Two-armed RCTs (n=66)	Multi-armed RCTs (n=192)
Number of RCTs, n(%)		2 armed: 66 (25.6%)	3 armed: 139 (53.9%) 4 armed: 43 (16.7%) 5 armed: 10 (3.9%)
Sample size per active arm, median (interquartile range)		98.5 (43.5, 158)	118.5 (66, 157)
Antidepressants examined (n of trials, n of participants)		amitriptyline (2, 176) trazodone (2, 794) fluoxetine (6, 1018) bupropion (8, 1531) sertraline (5, 1374) paroxetine (8, 734) venlafaxine (2, 290) nefazodone (1, 120) mirtazapine (3, 297) reboxetine (4, 368) citalopram (2, 358) escitalopram (3, 956) duloxetine (5, 1599) agomelatine (5, 1112) desvenlavaxine (2, 876) vilazodone (4, 1629) levomilnacipran (3, 1362) vortioxetine (1, 600)	(27, 3112) (8, 517) (33, 7431) (16, 4144) (14, 2775) (38, 8899) (20, 4895) (8, 1242) (10, 1450) (7, 2244) (11, 3428) (16, 5133) (16, 4673) (8, 3061) (7, 3503) (4, 1841) (2, 1292) (13, 5620)
Year of publication, n (%)	1979-1990	7 (11%)	24 (13%)
	1991-2000	17 (26%)	45 (23%)
	2001-2016	29 (44%)	79 (41%)
	unpublished	13 (20%)	44 (23%)

Published in final edited form as: *Int Clin Psychopharmacol.* 2018 Mar;33(2):66-72.
doi: 10.1097/YIC.0000000000000201

Appendix. Multivariate meta-regression to synthesize RORs

Consider there are n_A multi-arm trials (more than two arms) the involve drug A and n_B multi-arm trials the involve drug B. There are n trials with $n \leq n_B$ and $n \leq n_A$ that involve placebo (P) and drugs A and B. Because there are studies in common that, the summary meta-analytic treatment effect of A and B versus placebo OR^{AvP} and OR^{BvP} are correlated; the placebo arm is the same in these two estimates in the n trials in common. Consequently the two rations of odds-ratios

$$ROR^A = \frac{OR^{AvP}\ in\ two\ armed\ studies}{OR^{AvP}\ in\ n_A\ multi-armed\ studies}$$

$$ROR^B = \frac{OR^{BvP}\ in\ \ two\ armed\ studies}{OR^{BvP}\ in\ n_B\ multi-armed\ studies}$$

are correlated because their denominators re correlated. We need to estimate the covariance c($logROR^A, logROR^B$). Assuming a fixed effects model and that the study weights are known and fixed it is easy to show that

$$c(logROR^A, logROR^B) = c(logOR^{AvP}\ in\ multi-armed, logOR^{AvP}\ in\ \ multi-armed)$$

After some algebra it turns out that

$$c(logROR^A, logROR^B) = \frac{\sum_i^n w_i^A w_i^B \left(\frac{1}{S_i} + \frac{1}{F_i}\right)}{\sum_i^{n_A} w_i^A \sum_i^{n_B} w_i^B}$$

where

w_i^A: inverse of the variance of $log\ OR^{AvP}$ in the multi-arm study i

w_i^B: inverse of the variance of $log\ OR^{BvP}$ in the multi-arm study i

S_i: the number of successes in the placebo arm in the multi-arm study i

F_i: the number of failures in the placebo arm in the multi-arm study i

The synthesis of the RORs was performed using a multivariate meta-analysis routine in R (rma.mv in the metafor package) after specifying the entire variance-covariance matrix