

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

Yusuke Ogawa<sup>a</sup>, Toshi A. Furukawa<sup>a,b</sup>, Nozomi Takeshima<sup>a</sup>, Yu Hayasaka<sup>a</sup>, Lauren Z. Atkinson<sup>d,e</sup>, Shiro Tanaka<sup>c</sup>, Andrea Cipriani<sup>e,f</sup> and Georgina Salanti<sup>g</sup>

Previous studies have reported that effect sizes of antidepressants were larger in two-armed than in three-armed or more-armed (multiarmed) randomized trials, where the probability of being allocated to placebo is lower. However, these studies have not taken into account the publication bias, differences among antidepressants, or covariance in multiarmed studies, or examined sponsorship bias. We searched published and unpublished randomized-controlled trials that compared placebo with 21 antidepressants for the acute treatment of major depression in adults.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 multiarmed trials for each antidepressant, and then synthesized RORs across all the included antidepressants using the multivariate meta-analysis. A random-effects model was used throughout. Two hundred and fifty-eight trials (66 two-armed and 192 multiarmed trials; 80 454 patients; 43.0% with unpublished data) were included in the present analyses. The pooled ROR for response of two-armed trials over multiarmed trials was 1.09 (95% confidence interval: 0.96–1.24). The ROR did not materially change between types of antidepressants, publication year,

or sponsorship. The differences between two-armed versus multiarmed studies were much smaller than were suggested in previous studies and were not significant. *Int Clin Psychopharmacol* 33:66–72 Copyright © 2018 Wolters Kluwer Health, Inc. All rights reserved.

International Clinical Psychopharmacology 2018, 33:66–72

**Keywords:** antidepressants, meta-analysis, number of arms, placebo-controlled trial, randomized-controlled trial, systematic review, trial design

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

Correspondence to Toshi A. Furukawa, MD, PhD, Departments of Health Promotion and Human Behavior and of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan  
Tel: + 81 75 753 9491; fax: + 81 75 753 4641;  
e-mail: furukawa@kuhp.kyoto-u.ac.jp

Received 1 August 2017 Accepted 26 September 2017

## 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 worldwide 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. Although 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 the greater possibility of unblinding in two-armed versus multiarmed studies. Blinding may indeed be difficult to maintain in studies of psychotropic drugs because these drugs have characteristic side effects (Margraf *et al.*, 1991; Even *et al.*, 2000; Moncrieff *et al.*, 2004). When double-blindness is breached, drug efficacy over placebo would probably be overestimated (Leucht *et al.*, 2009).

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

All the above studies, however, have several problems. First, previous meta-analyses have unfortunately often been subject to publication bias. Analysis of the trial data submitted to Food and Drug Administration as a 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 in ESs between two-armed and three-armed trials may be because of 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. Second, previous studies have generally assumed that ESs of antidepressants are the same among all antidepressants. However, it has been reported that they may be markedly different (Cipriani *et al.*, 2009). Therefore, intervention effects should be examined and compared for each antidepressant separately. Third, it has been shown that an antidepressant appeared to be 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; Papakostas and Fava, 2009; Sinyor *et al.*, 2010) have not taken this factor into account so that the apparently larger ES reported in two-armed studies might be because of the ‘novelty effect’ of the agent, which is more likely to be studied in two-armed rather than multiarmed trials when the agent is ‘new’ and when the trial is sponsored by the manufacturer of the drug.

The aim of the present study is therefore to compare the odds ratios (ORs) of antidepressants over placebo when examined in two-armed versus three-armed or more-armed (heretofore termed multiarmed) trials while taking into account possible differences among different antidepressants on the basis of 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 were 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 thus, here, we 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-generation 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 sexes, 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 homepages of pharmaceutical companies that market the included drugs up to 8 January 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 the Hamilton Rating Scale for Depression (Hamilton, 1960), the Montgomery–Asberg Depression Rating Scale (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 and 12 weeks. When the number of responders was not reported, but the baseline mean and endpoint mean and SD of the depression rating scales were provided, we calculated the number of responding patients by using 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 that fulfilled the review criteria. Data from each study were extracted into a structured data abstraction form independently by two researchers. The risk of bias was assessed for each included study using the Cochrane Collaboration ‘risk of bias’ tool (Higgins and Green, 2011) by two independent researchers. Any disagreement was resolved through discussion or in consultation with a third member of the review team. On the basis of 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 ORs of response between the antidepressant and placebo by synthesizing ORs from all two-armed or multiarmed comparisons using the random-effects model. We next estimated the ratio of odds ratios (RORs) and their variance of

two-armed versus multiarmed trials for each antidepressant, and finally meta-analytically synthesized RORs across all the included antidepressants using the random-effects model. A random-effects model was used throughout because of possible clinical heterogeneity across the included trials because of 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 multiarmed trials do. Because two or more antidepressants were involved in multiarmed studies, the summary RORs were correlated (the placebo arm is the same in two estimates in the same trials in common) 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 versus agomelatine versus 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 Core Team; R Foundation for Statistical Computing, Vienna, Austria) after specifying the entire variance-covariance matrix (Appendix). We used Review Manager 5.3 (Nordic Cochrane Centre, Cochrane Collaboration, Copenhagen, Denmark), Stata 14 (StataCorp LP, College Station, Texas, USA), and R to carry out the analyses.

We started the assessment of heterogeneity by visual inspection of the forest plots. We also calculated  $I^2$  statistics (Higgins and Green, 2011) and analyzed them on the basis of the Cochrane Handbook's recommendations ( $I^2$  values of 0–40%: might not be important; 30–60%: may represent moderate heterogeneity; 50–90%: may represent substantial heterogeneity; 75–100%: considerable heterogeneity).

### Sensitivity analyses

To ascertain the robustness of our findings, we carried out 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 carry out the following subgroup analyses:

- (1) Numbers of arms in the multiarmed trials separately (three-armed, four-armed, and five-armed).
- (2) Type of antidepressants (tricyclic antidepressants versus 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 nonsponsored drug arms in multiarmed trials).

## Results

### Characteristics of the randomized-controlled trials included

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-armed or more-armed. Therefore, we could not calculate ROR for these three antidepressants. Altogether, 258 RCTs (80 454 patients) were finally included in the present analyses (Fig. 1). Table 1 presents detailed characteristics for two-armed and multiarmed RCTs. Among the 258 RCTs included in this study, 66 (25.6%) were two-armed and 192 were multiarmed, including, 139 (53.9%) three-armed RCTs, 43 (16.7%) four-armed RCTs, and 10 (3.9%) five-armed RCTs. The 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 multiarmed trials. The median number of studies per antidepressant was 13.5 (range: 5–46). Figure 2 summarizes the risk of bias of the studies included. All in all, 46 studies were rated as being at low risk of bias, 214 at moderate risk of bias, and 64 at high risk of bias.

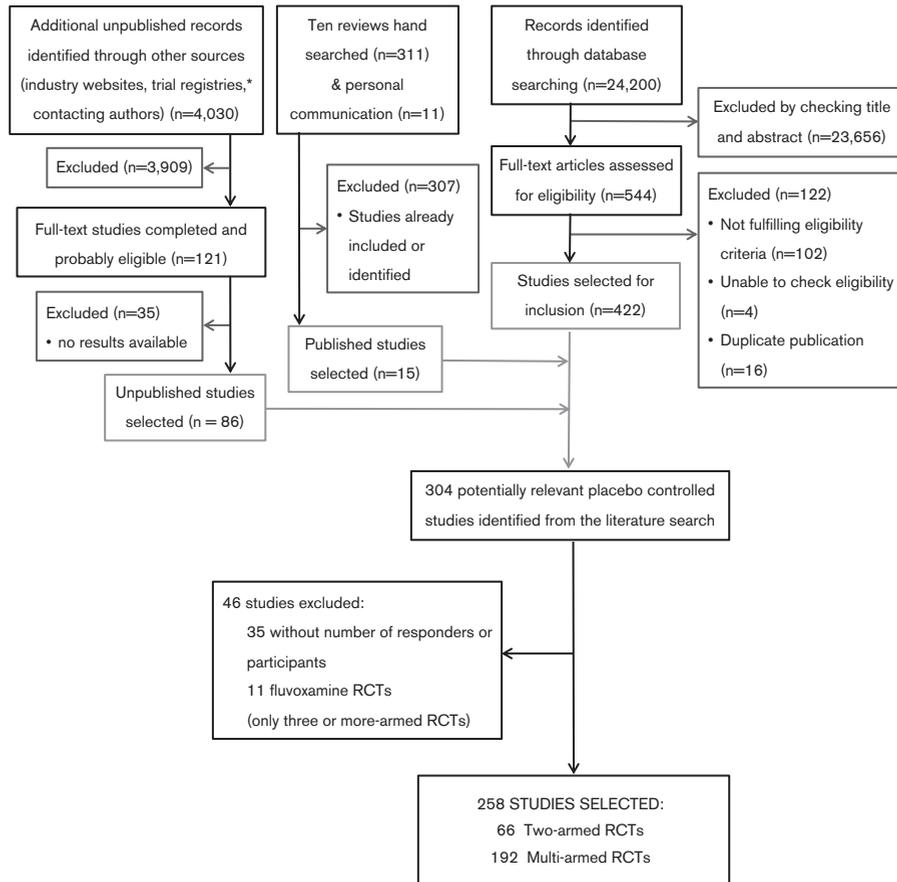
### Differences in effect size between two-armed and multiarmed randomized-controlled trials

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 multiarmed RCTs, respectively (Fig. 3). There was no significant difference between two-armed and multiarmed RCTs in the OR of response between antidepressant and placebo [pooled ROR: 1.09; 95% confidence interval (CI): 0.96–1.24] (Fig. 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: 0.0–65.6%). Because taking account of the covariance had little influence on the estimated ROR (the simple pooled ROR was 1.09 ( $I^2 = 38.6%$ ; 95% CI: 0.96–1.24), the following sensitivity and subgroup analyses were carried out without accounting for the covariances because of multiarmed studies.

### Sensitivity analyses

After exclusion of studies at high risk of bias, the ROR was 1.06 ( $I^2 = 34%$ ; 95% CI: 0.92–1.21). After exclusion of studies that imputed the number of responders, ROR was 1.06 ( $I^2 = 45%$ ; 95% CI: 0.90–1.25). Using the fixed-effect model instead of the random-effects model, ROR was 1.09 ( $I^2 = 38.6%$ ; 95% CI: 0.99–1.19).

Fig. 1



Flow diagram. RCT, randomized-controlled trial.

### Subgroup analyses

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

### Discussion

The differences between the two-armed versus multi-armed studies were much smaller than those 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 multiarmed RCTs, corresponding to 80 454 patients. The results of subgroup and sensitivity analyses did not alter this conclusion. RORs appeared larger for tricyclic antidepressants and for studies before 1990, but were not statistically significant either.

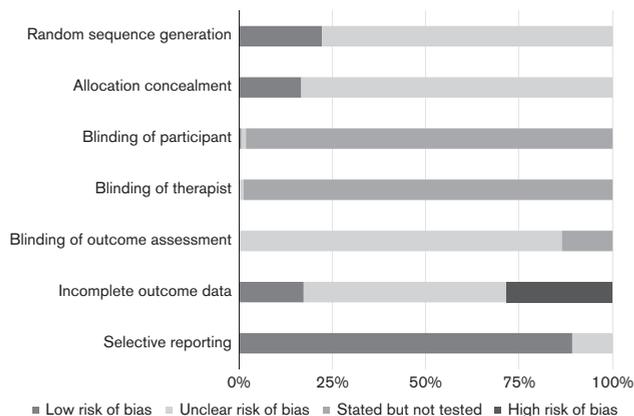
The differences between the previous studies and the present study may be explained as follows: first, the publication bias in our dataset is reduced as we could find unpublished information for 43.0% of the included studies through contacts with pharmaceutical companies and regulatory agencies. We could thus 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). Second, we used the random-effects model, which produces wider 95% CI than the fixed-effect model in the presence of heterogeneity. Although the overall the ORs tended to be higher in two-armed studies than multi-armed ones, the differences did not reach statistical

**Table 1 Characteristics of two-armed and three-armed or more-armed randomized-controlled trials**

	Two-armed RCTs (n = 66)	Multiarmed RCTs (n = 192)
Number of RCTs [n (%)]	Two-armed: 66 (25.6)	Three-armed: 139 (53.9) Four-armed: 43 (16.7) Five-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	27, 3112
Trazodone	2, 794	8, 517
Fluoxetine	6, 1018	33, 7431
Bupropion	8, 1531	16, 4144
Sertraline	5, 1374	14, 2775
Paroxetine	8, 734	38, 8899
Venlafaxine	2, 290	20, 4895
Nefazodone	1, 120	8, 1242
Mirtazapine	3, 297	10, 1450
Reboxetine	4, 368	7, 2244
Citalopram	2, 358	11, 3428
Escitalopram	3, 956	16, 5133
Duloxetine	5, 1599	16, 4673
Agomelatine	5, 1112	8, 3061
Desvenlafaxine	2, 876	7, 3503
Vilazodone	4, 1629	4, 1841
Levomilnacipran	3, 1362	2, 1292
Vortioxetine	1, 600	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)

RCT, randomized-controlled trial.

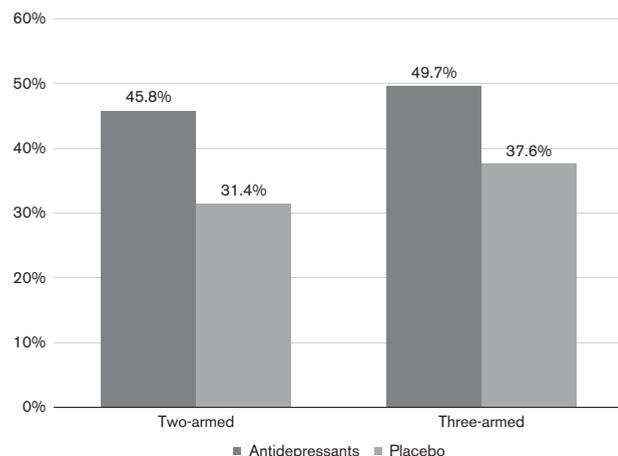
**Fig. 2**



'Risk of bias' graph: review authors' judgments of each risk of bias item presented as percentages across all included studies.

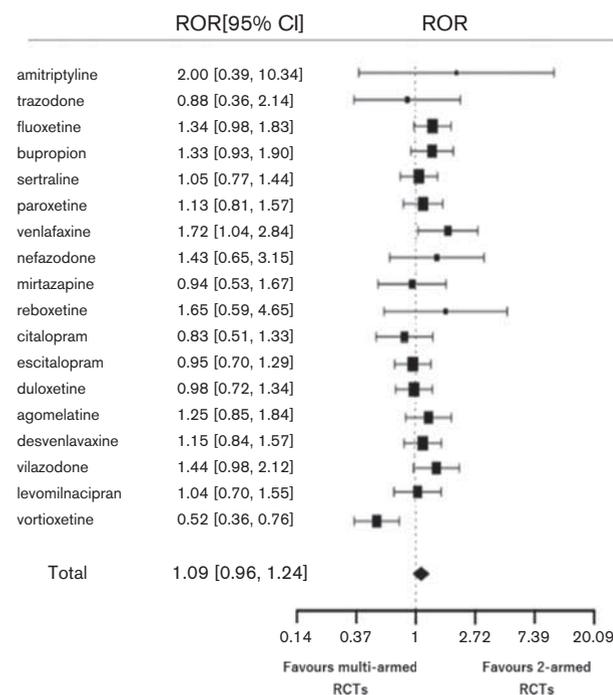
significance. We believe that our study had performed a more methodologically rigorous synthesis by estimating the ROR for each antidepressant, and then meta-analytically pooling all the RORs of the included antidepressants, instead of assuming a common efficacy for all the antidepressants included. A sensitivity analysis using the fixed-effect model instead of the random-effects

**Fig. 3**



Antidepressant and placebo response rates.

**Fig. 4**



Ratio of odds ratios (ROR) between two-armed and multiarmed randomized-controlled trials (RCTs). The antidepressants are listed in the order of their approval. CI, confidence interval.

model confirmed the primary findings. Third, 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-armed versus multiarmed studies because our subgroup analysis found little difference when the drugs in multiarmed studies were marketed by the sponsor of the drug or when they were not.

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

Our study has some limitations. We could not consider other trial and patient features that may have an impact on intervention effects, such as the difference in 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-armed versus multiarmed studies might have played a role, but we would need individual participant data to examine such effect modifiers. Moreover, given that the field of antidepressant trials in 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 multiarmed 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-armed rather than multiarmed trials would lead to over-estimation 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 could minimize the publication bias.

### Acknowledgements

This work was supported by the Japan Society for the Promotion of Science to Y.O. (16K09033). A.C. is supported by the NIHR Oxford cognitive health Clinical Research Facility. G.S. is a Marie Skłodowska-Curie fellow.

### Conflicts of interest

T.A.F. 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. N.T. has received lecture fees from Otsuka and Meiji. Y.H. has received lecture fees from Yoshitomi. S.T. has received lecture fees from Kobe City, Astra-Zeneca, Taiho Pharmaceutical, and Ono Pharmaceutical. He has received consultation fees from the Pharmaceuticals and Medical Devices Agency, DeNA Life Science, and CanBus. He has received outsourcing fees from Public Health Research Foundation, Japan Breast Cancer Research Group, Satt, and

Asahi Kasei Pharma. S.T. 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 Yakuin. A.C. was expert witness for a patent issue about quetiapine extended release. For the remaining authors there are no conflicts of interest.

### Appendix: multivariate meta-regression to synthesize RORs

Consider that there are  $n_A$  multiarm trials (more than two arms) that involve drug A and  $n_B$  multiarm trials that involve drug B. There are also  $n$  multi-arm trials that involve both drugs A and B; these studies contribute correlated data to the estimation of  $OR^{AVP}$  and  $OR^{BVP}$ . Consequently, the two ratios of odds ratios

$$ROR^A = \frac{OR^{AVP} \text{ in two-armed studies}}{OR^{AVP} \text{ in } n_A \text{ multiarmed studies}},$$

$$ROR^B = \frac{OR^{BVP} \text{ in two-armed studies}}{OR^{BVP} \text{ in } n_B \text{ multiarmed studies}},$$

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

$$c(\log ROR^A, \log ROR^B) =$$

$$c(\log OR^{AVP} \text{ in multiarm studies};$$

$$\log OR^{AVP} \text{ in multiarm studies}).$$

Consequently:

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

where  $w_i^A$  is the inverse of the variance of  $\log OR^{AVP}$  in the multiarm study  $i$ ;  $w_i^B$  is the inverse of the variance of  $\log OR^{BVP}$  in the multiarm study  $i$ ;  $S_i$  is the number of successes in the placebo arm in the multiarm study  $i$ ;  $F_i$  is the number of failures in the placebo arm in the multiarm 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.

### References

Barbui C, Cipriani A, Brambilla P, Hotopf M (2004). 'Wish bias' in antidepressant drug trials? *J Clin Psychopharmacol* 24:126-130.

- Bauer M, Whybrow PC, Angst J, Versiani M, Möller HJ, World Federation of Societies Biological Psychiatry Task Force on Treatment Guidelines for Unipolar Depressive Disorders (2002). 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* 373:746–758.
- 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 'blinded' conditions. *J Consult Clin Psychol* 60:664–669.
- Hamilton M (1960). A rating scale for depression. *J Neurol Neurosurg Psychiatry* 23:56–62.
- Higgins JP, Green S (2011). *Cochrane Handbook for Systematic Reviews of Interventions*, version 5.1.0. Available at: <http://handbook-5-1.cochrane.org/> [Accessed 16 July 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* 161:2045–2049.
- 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* 59:184–187.
- Moncrieff J, Wessely S, Hardy R (2004). Active placebos versus antidepressants for depression. *Cochrane Database Syst Rev* 1:CD003012.
- Montgomery SA, Asberg M (1979). A new depression scale designed to be sensitive to change. *Br J Psychiatry* 134:382–389.
- 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* 29:2369–2383.
- 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* 71:270–279.
- 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* 358:252–260.