

EBMH Perspective

Is placebo response in antidepressant trials rising or not? A re-analysis of datasets to conclude this long-lasting controversy

Toshi A. Furukawa¹, Andrea Cipriani², Stefan Leucht³, Lauren Z. Atkinson², Yusuke Ogawa¹, Nozomi Takeshima¹, Yu Hayasaka¹, Anna Chaimani^{4,5,6}, Georgia Salanti⁷

¹ Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine / School of Public Health, Kyoto, Japan

² Department of Psychiatry, University of Oxford, Oxford, UK; Oxford Health NHS Foundation Trust, Oxford, UK

³ Department of Psychiatry and Psychotherapy, Technische Universität München, Klinikum rechts der Isar, Munich, Germany

⁴ Paris Descartes University, Paris, France

⁵ INSERM, UMR1153 Epidemiology and Statistics, Sorbonne Paris Cité Research Center (CRESS), METHODS Team, Paris, France

⁶ Cochrane France, Paris, France

⁷ Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

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

Abstract

It had long been believed that placebo response rates in antidepressant trials have been increasing and that they were responsible for rising numbers of so-called failed antidepressant trials. Two recent systematic reviews examined this issue and reached completely opposite findings. Furukawa and colleagues in a paper published in 2016 found that the placebo response rates are stable since 1991 and the apparent increase up to 2000 was confounded by changes in trial design features. By contrast, Khan and colleagues more recently concluded that placebo response rates had grown steadily in the past 30 years. The two reviews differed in the datasets they used, definitions of placebo response, and statistical analyses. In this perspective article we examined if such differences were responsible for the two reviews' contrasting conclusions. Our re-analyses confirmed our previous results. We found that in any dataset and for any placebo response definition, there was no increase in placebo response over the years when the analysis was adjusted for the confounders related to study design features or when it was limited to studies published after 1990s. We conclude that placebo response in antidepressant trials has remained stable for the past 25 years, during which time the large majority of the studies have come to share similar design features.

There is ongoing debate whether placebo response in antidepressant trials has been increasing over the years. We recently examined this issue based on 252 published and unpublished double-blind randomized controlled trials of 20 antidepressants conducted between 1978 and 2015, and concluded (i) response rates on placebo have remained steady since 1991, and (ii) there was apparent increase up to 2000 but this trend became non-significant when adjusted for length of trial, number of study centres and use of fixed dosing ¹.

Khan et al ² re-examined this issue by studying 85 trials contained in FDA reviews for 16 antidepressants (1987-2013) and found that the magnitude of placebo response has been steadily growing in the past 30 years. To these authors' merit, they also examined the time trend in the efficacy estimates of antidepressants, namely the difference in symptom reduction between antidepressants and placebo. In this EBMH Perspective, we will focus on exploring the reasons for contradicting conclusions between the two research articles regarding the time trends in placebo response.

There are several important differences between these two studies which may explain their contradictory findings.

- (i) Furukawa et al ¹ included all published and unpublished studies, while Khan et al ² included studies reported in FDA reviews only. However comprehensive it may be, the former dataset cannot be exempt from risk of publication bias, while the latter apparently suffers from some selective reporting bias as the FDA reviews sometimes did not provide details especially for "failed" studies which then cannot be included in the analyses.
- (ii) The outcome was the proportion of responders, defined as showing 50% or greater reduction from baseline, in the former review ¹, while it was the percentage of symptom reduction from baseline in the latter study ². Both outcomes included imputations according to the last-observation-carried-forward method for study participants that provided at least one post-baseline measurement before dropping out of the study. In the former approach participants without post-baseline measurements were assumed to be non-responders. Such missing data rate was small (on average 3.6% across all studies) and previous studies have shown that this technique provides robust results ³. The latter approach ignores such early dropouts

completely but has the advantage of potentially being more sensitive because it is a continuous outcome.

Khan et al suggested that these two factors might explain the discrepancies between the two studies. However, there are some other differences as well.

- (iii) In the statistical analyses of the relationship between placebo response and year, Furukawa et al ¹ utilized meta-regression to take account of the different sample sizes of the included studies, while Khan et al ² used linear regression without weighting for precision of each study. Moreover, because there is no guarantee that the relationship between placebo response and year was linear, Furukawa et al ¹ studied the structural break point.
- (iv) Furukawa et al ¹ defined the year as the year of study completion or, when this was not reported, the year of publication. The year variable was considered missing when neither was available. Khan et al ² defined year as the FDA approval year of the drug in. This apparently risks some inaccuracies because studies for each drug must have been conducted several and variable years ahead of the drug approval year.
- (v) Khan et al ² examined three additional covariates: duration of the trial, number of trial arms and dosing schedule. Furukawa et al ¹ examined 14 covariates including the above three, as pre-specified in the study protocol for the systematic review ⁴, and found that length of trial, number of study centres and use of fixed dosing impacted on the placebo response.

In this reappraisal, we will examine whether the above differences can explain the discrepancies in the findings of the two studies and thereby elucidate the true underlying relationship between placebo response and year. First, we will examine the differences in the datasets. Second, after determining the dataset with possibly the least publication bias, we will examine the two primary outcomes, namely the proportion of responders and the percentage reduction in symptoms. Last, we will perform some sensitivity analyses to see whether our final findings are robust.

Dataset

In order to prepare the common dataset so that the ensuing analyses will be interpretable, we decided to concentrate on the studies contained in the FDA reviews of the 16 antidepressants that Khan et al

had included. In comparing the eligible studies in Furukawa et al ¹ and Khan et al ², we made the following observations.

- (i) There were six studies that Khan et al ² included in their study but Furukawa et al ¹ did not. One study may not have been eligible because 27% of the participants did not suffer from major depression (Study 86141 of citalopram). Five studies were excluded in Furukawa et al ¹ because the active drugs were administered outside the range of the FDA-licensed doses (Studies 223, 304, 308, 320 of desvenlafaxine and 85027 of mirtazapine). The latter studies might not provide valid estimates of placebo response as it would be harder to maintain double-blindness in studies using extreme high doses; by contrast, it may be argued that including such studies would provide a better estimate of placebo response across antidepressant trials including “failed” trials.

On the other hand, Khan et al ² seem to have missed 14 studies that are mentioned in the FDA reviews (Studies 84023, 84062, 003-023 of mirtazapine; CN104-002, 030A2-0004, CN104-045, CN104-054 of nefazodone; PAR-274 of paroxetine; SER 103, SER 310 of sertraline; 304, 12541A of vortioxetine).

- (ii) Furukawa et al ¹ had more outcome data, whether proportion of responders or percentage of symptom reduction, than Khan et al ² for the same studies. The latter restricted the data to be extracted from FDA reviews which often did not report details of trials (e.g. baseline or endpoint scores) especially when the trials were “negative”. In contrast the former review searched for all published and unpublished data for the trials by searching Cochrane Central Register of Controlled Trials, CINAHL, EMBASE, LiLACS, MEDLINE, PSYCINFO, regulatory agencies' websites including FDA, EMA and PMDA, several drug companies' websites, and international registers of clinical trials. When there were discrepancies in the reported outcome between published and unpublished reports, Furukawa et al ¹ prioritized the unpublished data according to the pre-specified review protocol to minimize the risk of selective reporting ⁴.

We judged that the primary dataset that is least susceptible to study publication bias and selective outcome reporting bias should therefore contain: all the studies included in Khan et al ², except for

86141 of citalopram, plus the 14 studies that they may have missed, with the outcome information supplemented from all available published and unpublished sources in addition to the FDA documents. The year of the study was defined as the year of study completion, the year of study publication or the year of drug approval from FDA, where available in this order. All in all, there were 98 placebo arms (9440 participants) representing 16 antidepressants (study year range: 1985-2012). We will however run a sensitivity analysis to see if the findings would be different if we restricted the included studies and data to those by Khan et al ².

Analyses

Figure 1 shows the secular changes in (i) proportion of responders and (ii) percentage of symptom reduction, meta-analytically synthesized for each year between 1985 and 2012.

Conventional meta-regression of the year on these two outcomes revealed that the year was a statistically significant predictor of placebo response (OR=1.01, 95%CI: 1.00 to 1.02, p=0.025 for proportion of responders, beta=0.004, 95%CI: 0.002 to 0.006, p<0.001 for percentage of symptom reduction).

Furukawa et al ¹ suggested, however, that the apparent increase may be confounded by the increase in trial length, in the number of multi-centred trials and in the use of fixed dosing schedule over the years. We therefore re-ran the above meta-regressions by adjusting for these three covariates: the influence of the year was no longer significant (OR=1.01, 95%CI: 0.99 to 1.02, p=0.38 and beta=0.001, 95%CI: -0.002 to 0.004, p=0.59, respectively): in both instances, the increase in multi-centered trials was the strongest confounder.

Furukawa et al ¹ further suggested that the placebo response has stabilized after 1991. The present dataset was too small to examine the structural break point so we simply examined if the meta-regression still revealed a significant influence of the year if we limited the included studies to those after 1991. For both outcomes, the year was no longer statistically significant (OR=1.00, 95%CI: 0.98 to 1.01, p=0.73 and beta=0.003, 95%CI: -0.0003 to 0.005, p=0.08, respectively) even without taking into account the other covariates. These trends after 1990 are clearly visible in Figure 1.

Sensitivity analyses

We ran several sensitivity analyses to confirm Khan et al ² 's findings by (i) restricting the dataset to what they used, and (ii) using simple linear regression without weighting for study precision.

The meta-regression of the year on the percentage of symptom reduction in Khan's dataset revealed a statistically significant influence of the study year ($p < 0.001$). However, when adjusted for the length of trial, number of study centres and dosing schedule, it was no longer significant ($p = 0.28$). The year was also non-significant when we limited the studies to those after 1991 ($p = 0.17$).

Using linear regression as in Khan et al, the results were similar: significant in the crude analysis ($p < 0.001$) but no longer so when adjusted for the confounders ($p = 0.28$) or when limited to studies after 1991 ($p = 0.16$).

Conclusion

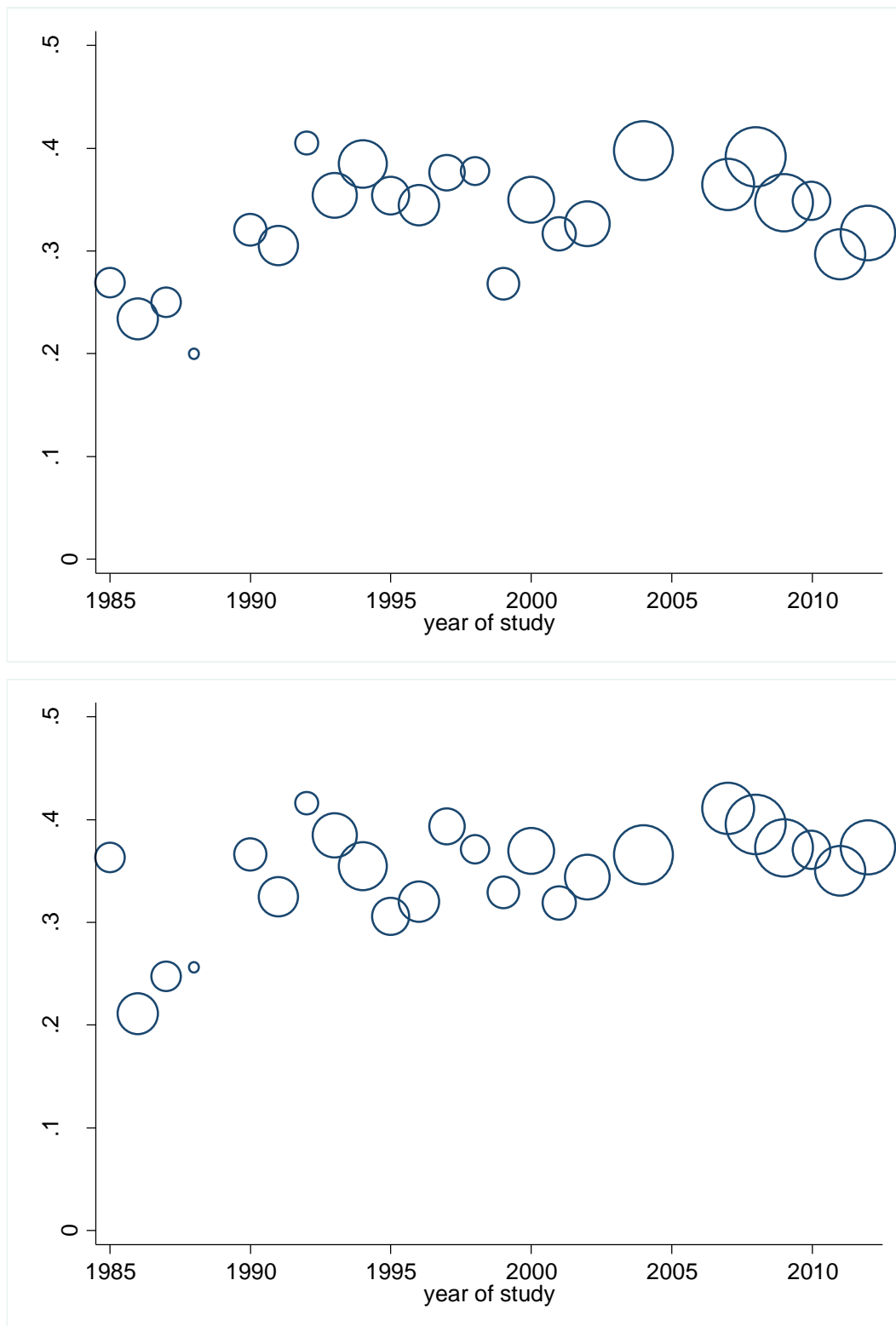
The opposing conclusions regarding the secular trend in placebo response in antidepressant trials reached by Furukawa et al ¹ and Khan et al ² were not due to the differences in the datasets they used (all available published and unpublished studies vs FDA reviews), the definition of response they examined (proportion of responders vs percentage reduction in symptoms) or the analytical methods they used (weighted meta-regression vs simple linear regression). Regardless of these differences, there was apparent increase in the placebo response over the years, which however became non-detectable when adjusted for several confounders notably the number of study centres (which Khan et al ² did not examine) and when limited to studies since 1990s. We conclude that the placebo response in antidepressant trials has remained stable for the past 25 years, during which time the large majority of the studies have come to share similar design features.

References

1. Furukawa TA, Cipriani A, Atkinson LZ, et al. Placebo response rates in antidepressant trials: a systematic review of published and unpublished double-blind randomised controlled studies. *Lancet Psychiatry* 2016; **3**(11): 1059-66.

2. Khan A, Mar KF, Faucett J, Schilling SH, Brown WA. Has the rising placebo response impacted antidepressant clinical trial outcome? Data from the US Food and Drug Administration 1987-2013. *World Psychiatry* 2017; **16**: 181-92.
3. Spineli LM, Higgins JP, Cipriani A, Leucht S, Salanti G. Evaluating the impact of imputations for missing participant outcome data in a network meta-analysis. *Clin Trials* 2013; **10**(3): 378-88.
4. Furukawa TA, Salanti G, Atkinson LZ, et al. 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* 2016; **6**(7): e010919.

Figure 1. Secular changes in proportion of responders (top) and percentage of symptom reduction (bottom)



The size of the bubble is proportional to the sample size.