

# **Revisiting placebo response rates in antidepressant trials: a systematic review of published and unpublished double-blind studies**

Toshi A. Furukawa\*, MD, Departments of Health Promotion and Human Behavior and of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan

Andrea Cipriani\*, PhD, Department of Psychiatry, University of Oxford, Oxford, UK

Lauren Z. Atkinson, MSc, Department of Psychiatry, University of Oxford, Oxford, UK

Stefan Leucht MD, Department of Psychiatry and Psychotherapy, Technische Universität München, Klinikum rechts der Isar, Munich, Germany

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

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

Yu Hayasaka, MD, Department of Health Promotion and Human Behavior, Kyoto 〓 University Graduate School of Medicine/School of Public Health, Kyoto, Japan

Anna Chaimani, PhD, Department of Hygiene and Epidemiology, University of Ioannina, Greece

Georgia Salanti, PhD, Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland; Department of Hygiene and Epidemiology, University of Ioannina, Ioannina, Greece.

\*Contributed equally

Correspondence to:

Dr Andrea Cipriani, Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford, OX3 7JX, UK. Email: [andrea.cipriani@psych.ox.ac.uk](mailto:andrea.cipriani@psych.ox.ac.uk)

## SUMMARY

**Background:** Previous studies have shown that placebo response rates in antidepressant trials have been increasing over the years since 1970s. However, these studies have been based on outdated or limited datasets and used inappropriate statistical methods.

**Methods:** We conducted a systematic review of published and unpublished placebo-controlled double-blind randomised controlled trials of first- and second-generation antidepressants for acute treatment of major depression in adults (update: January 2016). The log-transformed proportions of response, defined as 50% or greater reduction in depression severity from baseline, were meta-analytically synthesized and the structural break point in their secular changes was examined. The influence of the study year and other trial and patient characteristics on the response rates was examined through meta-regression.

**Findings:** We identified 252 placebo-controlled trials (26,324 patients on placebo) between 1978 and 2015. There was a structural break in 1991, and the average placebo response rates in antidepressant trials have remained constant in the range between 35-40% since then (relative risk (RR) 1.00, 95% CI 0.97 to 1.03, for every five-year increase). The reported influence of the study year before 2000 was replicated (RR 1.10, 95% CI 1.05 to 1.15) but became non-significant (RR=1.04, 95% CI 0.99 to 1.09) when controlled for trial methodological factors such as the study duration and the number of study centres.

**Interpretation:** Contrary to the widely-held belief, the average placebo response rates in antidepressant trials have been staying the same for more than 25 years. This new evidence will have an impact on the interpretation of the scientific literature and the future of psychopharmacology, both from a clinical and methodological point of view.

**Funding:** Japan Society for Promotion of Science; Great Britain Sasakawa Foundation

## **Research in context**

### ***Evidence before this study***

It has been accepted knowledge that placebo response rates in antidepressant trials have been increasing over the past decades since 1970s. This has been associated with the increasing number of so-called “failed” antidepressant trials, and many authors have examined and debated factors behind this relationship such as depressive symptoms’ severity, inflation of baseline score and the use of placebo run-in phases in the trial design. Unfortunately, previous reviews have been based on incomplete datasets or used inappropriate statistical methods (for instance, the relationship between the proportion of placebo responders and the year of publication was examined as correlation coefficients or linear regressions without taking into account the precision of estimated proportion in each included study and erroneously assuming normal distribution for proportions).

### ***Added value of this study***

We have carried out a systematic and comprehensive search for published and unpublished randomised trials of first- and second-generation antidepressants for acute treatment of adult patients with major depression (update: January 2016), and identified a dataset of 252 placebo-controlled double-blind studies since 1978 (overall, 26,324 participants randomised to placebo). In our large dataset, there was a structural break in 1991, and the average placebo response rate

in antidepressant trials has basically remained constant in the range of 35-40% since 1991 (relative risk (RR) 1.00, 95%CI 0.97 to 1.03, for every five year). We were able to replicate the reported increase in placebo response rates before 2000 (RR 1.10, 95% CI 1.05 to 1.15, for every five year). However, we found that this analysis was confounded by methodological factors (such as shorter duration of trials and preponderance of single-center studies), which were frequent in the very early trials and became less often employed since 1990s. Controlling for these variables, the relationship between placebo response rate and study year was no longer significant.

#### ***Implications of all the available evidence***

Placebo response rates in antidepressant trials have, on average, stayed constant over the past 25 years. This is completely new evidence, which will have an impact on the interpretation of the scientific literature and the future psychopharmacology, with both clinical and methodological implications. In terms of study design, some important factors should be reconsidered. Appropriately-timed assessments and multi-center design are necessary for trials to provide results directly relevant and applicable to the real-world clinical practice setting. Trialists need to employ design characteristics to make their trials as clinically relevant as possible.

## INTRODUCTION

It has been widely accepted that placebo response rates in antidepressant trials have been increasing over the past 30 years. The first systematic study on this topic revealed that among the 75 placebo-controlled antidepressant trials up to 2000, there was a positive correlation ( $r=0.45$ , 95% Confidence Interval (CI) 0.25 to 0.61) between the proportion of responders on placebo and the year of publication.<sup>1</sup> A very similar relationship was found among paediatric antidepressant trials ( $r=0.64$ , 95% CI 0.02 to 0.91).<sup>2</sup> Other studies examined pre-post raw change scores<sup>3,4</sup> or their effect sizes<sup>5,6</sup> in the placebo arms and confirmed a significant association between the thus-measured placebo response and the year of study publication.

As the increasing placebo response was suspected of contributing to increasing numbers of so-called “failed” antidepressant trials,<sup>7,8</sup> many authors explored factors behind this relationship in the hope of finding measures to curtail this increase. Walsh et al. found that the length of the trial and the minimum depression severity required at baseline also influenced the placebo response rates in adult antidepressant trials; however, when jointly entered into multiple linear regression, only the year of publication remained significantly associated with the response to placebo.<sup>1</sup> Age, sex, use of placebo run-in or co-medication did not show significant association either.<sup>1</sup> The number of recruiting sites and the number of randomized patients were identified as important factors associated with placebo response among paediatric trials.<sup>2</sup> Other patient or

study characteristics suggested in the literature to date include baseline severity,<sup>3,6</sup> number of arms or probability of being allocated to placebo,<sup>9,10</sup> dosing schedule,<sup>11</sup> length of trial,<sup>3,12</sup> and inflation of baseline severity<sup>13</sup>. However, all these empirical studies have methodological problems. Many of them are based on samples that not only are now out of date but also are biased, because they included only published trials, often only in English language.<sup>1,3,6,12,14</sup> Other reviews included data from the US Food and Drug Administration, which thus precluded publication bias but which at the same time necessarily limited the number and scope of available studies.<sup>4,5,11</sup> Moreover, in terms of statistical analysis, previous studies examined the relationship between the proportion of placebo responders and the year of publication as correlation coefficients or linear regressions without taking into account the precision of estimated proportion in each included study and erroneously assuming normal distribution for proportions.<sup>1,2,14</sup>

We have therefore conducted a comprehensive and systematic search for published and unpublished double-blind placebo-controlled trials about first- and second-generation antidepressants (the search is part of an update of a network meta-analysis about antidepressants in major depression<sup>15</sup>). This report focuses on the placebo response rates and, using the largest dataset so far, aimed at addressing the following questions: (i) whether the believed increase in placebo response rates have persisted up to 2015, and if not, (ii) whether

we can replicate the previous findings <sup>1</sup>, and (iii) what patient and study characteristics may influence the studied relationships.

## **METHODS**

### ***Sources of data and criteria for review***

We included all double-blind randomised controlled trials (RCTs) comparing one of the following active drugs with placebo in the acute treatment of major depression: agomelatine, amitriptyline, bupropion, citalopram, clomipramine, desvenlafaxine, duloxetine, escitalopram, fluoxetine, fluvoxamine, levomilnacipran, milnacipran, mirtazapine, nefazodone, paroxetine, reboxetine, sertraline, trazodone, venlafaxine, vilazodone and vortioxetine. We have included all the second generation antidepressants licensed in Europe, US, Australia and Japan, and, of the older agents, we have selected the two tricyclics included in the WHO list of essential medicine (amitriptyline and clomipramine), plus trazodone and nefazodone because they have very distinct effect and tolerability profiles.<sup>15</sup> RCTs with patients aged 18 or older, of both sexes and with a primary diagnosis of unipolar major depression according to standard operationalised diagnostic criteria were included. We searched the relevant electronic databases including the Cochrane CENTRAL, CINAHL, EMBASE, LILACS, MEDLINE, MEDLINE In-Process and PSYCINFO, supplemented by searches for published, unpublished and ongoing RCTs in major drug-



approving agencies, clinical trial registries and pharmaceutical company websites up to January 2016. We also contacted the National Institute for Health and Clinical Excellence (NICE, UK), the Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG, Germany) and other relevant organisations and experts in the field for any additional information not already identified. For details, please refer to the full study protocol.<sup>15</sup> No date limits or language restrictions were applied to any of the searches.

At least two persons independently checked reference titles and abstracts thus identified. Full texts of all potentially eligible studies were retrieved, and were further inspected using the same eligibility criteria. Any disagreements were resolved by discussion or in consultation with a third member of the review team. Two or more persons then independently extracted data using the pre-defined data extraction sheet. Two or more independent raters also assessed the quality of each study using the Cochrane risk of bias tool.<sup>16</sup> Each study was then classified as having low risk of bias if none of the domains was rated at high risk of bias and three or less at unclear risk, moderate if one was rated at high risk of bias or none was rated at high risk of bias but four or more at unclear risk, and all other cases were assumed to pertain to high risk of bias.<sup>15</sup>

As we obtained information from various sources including single trial publications, published meta-analyses, regulatory agencies documents and company website information, great efforts were expended to identify and avoid duplication, both before and after data extraction, and to

contact original study authors for supplemental information, if needed.

## ***Analyses***

We examined the proportions of placebo response in the placebo arms of all the identified trials in the following manner. We first tabulated the proportion of placebo response by half-decades since the earliest trial. The log-transformed proportion of responders was meta-analytically synthesized for each half decade, using the random effects model and the method of moments<sup>17</sup> to estimate the between-study component of variance tau-squared with *metan* in Stata Version 14.1 (StataCorp, College Station, Texas, USA). Response was defined as a 50% or greater reduction in the total score from baseline to week 8 on a standardised observer-rating scale for depression. We used Hamilton Rating Scale for Depression (HRSD) or, if HRSD was not used, another standardised and validated observer-rating scale such as Montgomery-Asberg Depression Rating Scale (MADRS). As we were interested in the acute treatment of major depression, assessments had to be done between 4 and 12 weeks, and the ones closest to 8 weeks were prioritised. Longer-term studies were excluded from the statistical analyses if they did not provide data for the 4-12 weeks period. When the number of responders was not reported but baseline mean and endpoint mean and its standard deviation on the depression rating scale were provided, we calculated the number of responders by employing a validated imputation

method.<sup>18</sup> In order to abide by the intention-to-treat principle we used the number randomised as the denominator of our primary analysis and we assumed any participants with an unknown outcome to be non-responders in accordance with a validated method to estimate relative treatment effects.<sup>19</sup>

Next we examined if there is any structural break in this time series, i.e. when a time series abruptly changes at a point in time, by applying the procedure *estat sbsingle* in Stata to our series of meta-analytically pooled response rates for each year from 1978 through 2015. The structural break test helps us to determine when and whether there is a significant change in the data series. We then ran a meta-regression of log of placebo response rate by the year of study completion. We gave preference to this variable over the year of publication because, by definition, the latter is unavailable for unpublished studies. We next ran meta-regression of log of placebo response rates by year of the study completion while controlling for various participant and study design characteristics. Depression severity as measured by different versions of HRSD and MADRS was recoded into HRSD-17 scores by using the conversion table based on the item response theory.<sup>20</sup> Each variable was entered individually in addition to the study year first and when the participant/study characteristics proved to exert statistically significant influence, they were entered together to see whether the relationship between log of placebo response rates and the year of the study was affected. For meta-regression, we used

the procedure *metareg* in Stata with the method of moments estimator for heterogeneity. The percentage of heterogeneity standard deviation that is explained by additional covariates was monitored and reported.

The influential paper by Walsh et al <sup>1</sup> was published in 2002 and included trials up to 2000. We examined whether we could replicate their finding by limiting our sample to trials up to the year 2000, while taking a statistically more appropriate dependent variable and employing meta-analytic methods with proper weighting. We first ran simple meta-regression by year of the study and then examined if additional study characteristics affected the relationship.

### ***Sensitivity analyses***

We conducted the following sensitivity analyses in order to test the robustness of our primary findings.

- (i) excluding studies that imputed the number of responders based on the continuous depression severity scores;<sup>18</sup>
- (ii) using the number of participants analysed instead of the number randomised as denominator when calculating the response rate;
- (iii) using the year of publication instead of the year of study completion;
- (iv) excluding unpublished studies;

- (v) excluding studies at high/moderate risk of bias;
- (vi) using log-odds of placebo response instead of log of placebo response rate as the dependent variable;
- (vii) using pre-post change on the rating scale as the dependent variable.

## RESULTS

### *Selected studies*

The literature search (last update: January 2016) identified 303 placebo-controlled trials. However, in 51 of them either the year of the study or any information about efficacy were missing, so the remaining 252 studies constituted the dataset for our analysis (134 RCTs with published data only, 74 with both published and unpublished data, and 44 studies with unpublished data only) (See Supplementary material for the reference list of included studies and Supplement Figure 1 for PRISMA flowchart). Supplement Table 1 summarises the characteristics of the included studies. Between 1978 and 2015, placebo response rates ranged widely from 0% to 70% but the weighted mean proportion of responders appeared to converge towards 35% to 40%, especially after the early 1990s (Table 1). It must be noted that overall the I-squared remained high (74.1% overall), indicating substantial heterogeneity in placebo response rates among the included trials throughout these years.

### ***Studies after 1991***

The structural break test suggested that the break date was the year 1991 ( $p=0.04$ ), which coincided well with the observed values as reported in Table 1. When pooled through meta-analysis, the summary response rate for this time period was 0.37 (95% CI 0.36 to 0.39) with a heterogeneity standard deviation 0.203. We ran meta-regression to examine the trend after this date and also to explore possible confounders in the relationship. Figure 1 shows meta-regression of log of proportion of placebo response by the year of the study completion after 1991. Since 1991, the relative risk in the proportion of placebo response was 1.00 (95% CI 0.97 to 1.03,  $P=0.99$ ) for every 5-year increase in the study year. When individual participant or study design characteristics were examined, the length of the study and the number of study centers emerged as significant factors: the longer studies had greater response rates, and the multi-center trials had greater response rates than single-center studies (RR 1.03, 95% CI 1.01 to 1.05 for one more week in trial length, and RR 1.32, 95% CI 1.11 to 1.57 for multi-center trials vs single-center trials). When both variables were entered together in the meta-regression, both remained statistically significant predictors (RR 1.03, 95% CI 1.00 to 1.05 for trial duration, and RR 1.31, 95% CI 1.10 to 1.56 for multi-center trials) while the study year remained non-significant (RR 0.99, 95% CI 0.96 to 1.01, for every 5-year increase). This multivariable model explained 2.2%

of the total heterogeneity standard deviation.

### ***Studies before 2000***

We examined if we could replicate Walsh et al's findings by limiting the studies up to the year 2000.<sup>1</sup> The summary response rate from simple meta-analysis was 0.34 (95% CI 0.32 to 0.35) with a heterogeneity standard deviation 0.223. Meta-regression of log of proportion of placebo responders by the study year up to 2000 clearly indicated the increasing placebo response (RR 1.10, 95% CI 1.05 to 1.15,  $P < 0.001$ , for every 5-year increase) (Figure 2, Table 2). When possible confounders were examined individually in meta-regression, the length of the trial, the number of study centers, the dosing schedule and the number of patients randomised emerged as possible confounders (Table 2). In addition to the same tendencies observed for the studies after 1991 for the number of weeks and the number of study centers, studies using fixed dosing schedule showed lower placebo response rates than those using flexible dosing schedule, and the smaller arms tended to show lower response rates. When all these four variables were entered into multivariable meta-regression, the mean response rate for a hypothetical 4-week, single-center trial with fixed dose regimen in the earliest year of publication (1978) was estimated to be 0.19 (95% CI 0.16 to 0.23). Then the response rate increases for a multi-centre trial (RR 1.41, 95% CI 1.22 to 1.62), for a flexible dosing schedule (RR 1.17, 95% CI 1.04 to 1.31)

and for one more week in the length in the trial (RR=1.04, 95% CI 1.01 to 1.06). However, the effect of the study year was no longer statistically significant (RR 1.04, 95% CI 0.99 to 1.09 for every 5-year increase, P=0.18). This multivariable model was able to explain 28.4% of the total heterogeneity and 18.4% of the residual heterogeneity of the meta-regression model with the study year alone. Table 3 summarises the secular changes in these characteristics through the years, and clearly shows that the studies were much shorter and more often single-centered in the 1980s than in 1990s and later, while the flexible dosing regimen became increasingly unpopular in the 2000s and later.

### ***Sensitivity analyses of the primary results***

All sensitivity analyses revealed no effect of year among studies after 1991, providing support for robustness of our main findings (see full results in Supplement Figure 2 and Supplement Table 2).

## **DISCUSSION**

The present study examined the secular changes in the proportion of responders in the placebo arms through the largest systematic review of antidepressant RCTs so far and brought about two insights that shed new light in the current heated debate about diminishing efficacy of



antidepressant drugs. First, the average placebo response rates have stayed constant since 1991 in the range of 35 to 40%. Second, the myth of increasing placebo response was due to the exceptionally low rates in 1980s and before, which were confounded especially by the shorter duration of the trials and the preponderance of single-center studies. These findings are quite contrary to the heretofore almost unanimously held belief in the literature that the antidepressant trials are fraught by increasing placebo response rates (interestingly, the meta-analysis by Rief et al found that only doctor-reported outcomes, such as HAMD scores, showed the rise in placebo response of studies published before 2000, but this was not true for patient-reported outcomes).<sup>1-6,14</sup> Based on our large dataset, however, there is no doubt now that the average placebo response rates have been staying constant since 1990s, as far as antidepressant trials in depression are concerned. These results are robust as they remained consistent in all the sensitivity analyses. It may well be the case, however, that the relationship may be different in other psychiatric disorders.<sup>21,22</sup>

The second insight may do away with much speculation in the literature about the causes of increasing placebo response. Of the variables found to be statistically significant in the literature, when we examined, in our much larger dataset and with appropriate statistical methods, the length of trial, the number of study centres, the number of randomised patients, the baseline severity, the number of arms, and the dosing schedule, only the first two were found to exert

consistent influence independently of each other. Previous findings of the relationship between increasing placebo response rates and study year became non-significant when we controlled for the confounding factors. In other words, the reported increase before 2000 is a methodological artefact because, when the relevant trial design characteristics were held relatively constant after 1990s, the placebo response rates stayed stable. It is also to note here that publication status did not influence this relationship, because excluding unpublished studies did not change the findings (Supplement Table 3). There are several strengths of our study. Firstly, we were able to include 252 placebo-controlled studies of antidepressants between 1978 and 2015, in contrast to 75 in Walsh et al up to 2000<sup>1</sup> or 107 in Undurraga et al up to 2010.<sup>14</sup> The total number of participants on placebo was three to four times larger in the current study than in the previous studies (26 323 vs 6285 or 9925, respectively). We cannot rule out the possibility that our search still failed to identify some unpublished studies, however it is evident that our dataset is more capable of revealing the underlying truth about the placebo response rate than the previous smaller ones. Secondly, we employed statistically more appropriate methods throughout the analyses by taking the log of response rate instead of raw rates, by weighting the studies by their sample sizes, and also by using the objective method to detect the change point in the longitudinal time series.

This study however is not without some limitations. We included trials of an a priori defined list

of first- and second-generation antidepressants but we were not able to include all medications known to act as antidepressants to date. However, this approach was inevitable as there is no uniformly accepted list of effective antidepressants and some compounds may be approved for major depression in some countries but not in others. Moreover, by specifying the drugs to focus on, we think we were able to conduct more systematic and comprehensive search for both published and unpublished trials, which is demonstrated by the size of the current dataset that far exceeds those of any previous study on the topic. To be clinically informative, we included only licensed drugs, therefore we did not include studies of drugs whose development failed. This failure could have happened because of very high response rates. The non-inclusion of placebo response rates from failed development programs could have biased our sample of eligible studies toward a lower rate of placebo response, however we carried out a comprehensive search of more than 250 published and unpublished trials and we do not think that the potential omission of a handful of such studies would materially affect the overall results. The substantial heterogeneity as suggested by high I-squared values suggest that the placebo response rates remain highly variable through the years. To properly address this methodological issue, we therefore used the random effects meta-analytic methods.<sup>23</sup> Only the access to individual patient data would allow researchers to better evaluate moderators and mediators of treatment effect, however this is always not easy and often very time consuming.<sup>24</sup> Moreover, it

should be carried out in a systematic way and across multiple drugs, and not limited to a small sample of available studies about one drug, as previously done in the field of antidepressant trials.<sup>25</sup>

When we replicated Walsh et al.'s findings before 2000, very early studies were on average smaller than those published later. This means that in the meta-regression analysis the study year was correlated with study weights, potentially challenging the assumptions of a joint analysis. When the sample size was examined as a univariable explanatory variable, it was a statistically significant predictor but lost its significance when entered together with other covariates in the multivariable model (Table 2). We also ran a post hoc non-weighted multivariable regression, which indicated that neither the study year nor the sample size was a statistically significant predictor, differently from what previously reported.<sup>26</sup>

In conclusion, even if the average response to placebo has remained constant over the past 25 years, a percentage of 35 to 40% is still a high proportion of patients. Assuming this is the best estimate we have of an average response to placebo, important implications for clinicians and researchers can be drawn. The expectation of improvement, classic conditioning and the contact with a healthcare environment with supportive and therapeutic features contribute to the objective response observed in patients with major depression who are randomised to placebo.<sup>27</sup> These non-pharmacologic aspects, though, are usually not provided to the same

extent in standard clinical practice. Clinicians should create a specific context and level of therapeutic contact, to enhance non-specific effects of treatment and gain greater treatment response.<sup>28</sup> In terms of research, innovation in psychopharmacology is urgently needed not only for drug discovery and development, but also in terms of clinical trials' design. The key question is whether there is still need to have placebo-controlled phase III studies.<sup>29</sup> Clinical research organizations have been running trials in depression for most pharmaceutical companies since the early 1990s and they are financially incentivized to have as many visits as possible to increase the duration and therefore the cost per patient of the trial. Moreover, they tend to use as many centers as possible so that the study is completed sooner. To evaluate the next putative antidepressants, future studies should require the development of more efficient study designs to improve signal detection in drug development studies and an increased antidepressant response in clinical treatment.<sup>30</sup> It is time for academics, pharmaceutical industry and regulators to create new models of drug discovery and drive innovation in the methodology of drug development.<sup>31</sup>

**Contributors**

TAF and AC conceived the study and drafted the manuscript. GS, AC, ACh and TAF designed the methods. LA, YO, NT, YH selected the articles and extracted the data. TAF, ACh and GS analysed the data. TAF, AC and SL interpreted the results. All authors critically revised the manuscript. All authors read and met the ICMJE criteria for authorship and agree with the results and conclusions of this Article.

**Role of the funding source**

This study was supported in part by JSPS KAKENHI (Grant-in-Aid for Scientific Research) Grant Number 26670314 to TAF and by Great Britain Sasakawa Foundation Butterfield Award 4939 to AC and TAF. The funders had no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication. TAF and AC had full access to all the data in the study and had final responsibility for the decision to submit for publication.

**Declaration of interests**

TAF has received lecture fees from Eli Lilly, Janssen, Meiji, MSD, Otsuka, Pfizer and Tanabe-Mitsubishi, and consultancy fees from Sekisui Chemicals and Takeda Science Foundation. He has

received royalties from Igaku-Shoin and Nihon Bunka Kagaku-sha publishers. He has received grant or research support from Mochida and Tanabe-Mitsubishi. He is diplomate of the Academy of Cognitive Therapy. AC was expert witness for Accord Healthcare for a patent issue about quetiapine extended release. SL has received honoraria for lectures from Eli Lilly, Lundbeck (Institute), Pfizer, Janssen, BMS, Johnson and Johnson, Otsuka, Roche, SanofiAventis, ICON, Abbvie, AOP Orphan, Servier; for consulting/advisory boards from Roche, Janssen, Lundbeck, Eli Lilly, Otsuka, TEVA; for the preparation of educational material and publications from Lundbeck Institute and Roche. Eli Lilly has provided medication for a clinical trial led by SL as principal investigator. The other authors declare no competing interests.

### **Acknowledgements**

Andrea Cipriani is supported by the NIHR Oxford Cognitive Health Clinical Research Facility.

## REFERENCES

1. Walsh BT, Seidman SN, Sysko R, Gould M. Placebo response in studies of major depression: variable, substantial, and growing. *JAMA* 2002; **287**(14): 1840-7.
2. Bridge JA, Birmaher B, Iyengar S, Barbe RP, Brent DA. Placebo response in randomized controlled trials of antidepressants for pediatric major depressive disorder. *Am J Psychiatry* 2009; **166**(1): 42-9.
3. Khan A, Bhat A, Kolts R, Thase ME, Brown W. Why has the antidepressant-placebo difference in antidepressant clinical trials diminished over the past three decades? *CNS Neurosci Ther* 2010; **16**(4): 217-26.
4. Khin NA, Chen YF, Yang Y, Yang P, Laughren TP. Exploratory analyses of efficacy data from major depressive disorder trials submitted to the US Food and Drug Administration in support of new drug applications. *J Clin Psychiatry* 2011; **72**(4): 464-72.
5. Fountoulakis KN, Moller HJ. Efficacy of antidepressants: a re-analysis and re-interpretation of the Kirsch data. *Int J Neuropsychopharmacol* 2011; **14**(3): 405-12.
6. Rief W, Nestoriuc Y, Weiss S, Welzel E, Barsky AJ, Hofmann SG. Meta-analysis of the placebo response in antidepressant trials. *J Affect Disord* 2009; **118**(1-3): 1-8.
7. Montgomery SA. The failure of placebo-controlled studies. ECNP Consensus Meeting, September 13, 1997, Vienna. European College of Neuropsychopharmacology. *Eur Neuropsychopharmacol* 1999; **9**(3): 271-6.
8. Rutherford BR, Roose SP. A model of placebo response in antidepressant clinical trials. *Am J Psychiatry* 2013; **170**(7): 723-33.
9. Sinyor M, Levitt AJ, Cheung AH, et al. 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* 2010; **71**(3): 270-9.
10. Papakostas GI, Fava M. Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD. *Eur Neuropsychopharmacol* 2009; **19**(1): 34-40.
11. Khan A, Khan SR, Walens G, Kolts R, Giller EL. Frequency of positive studies among fixed and flexible dose antidepressant clinical trials: an analysis of the food and drug administration summary basis of approval reports. *Neuropsychopharmacology* 2003; **28**(3): 552-7.
12. Rutherford BR, Cooper TM, Persaud A, Brown PJ, Sneed JR, Roose SP. Less is more in antidepressant clinical trials: a meta-analysis of the effect of visit frequency on treatment response and dropout. *J Clin Psychiatry* 2013; **74**(7): 703-15.
13. Landin R, DeBrotta DJ, DeVries TA, Potter WZ, Demitrack MA. The impact of restrictive entry



- criterion during the placebo lead-in period. *Biometrics* 2000; **56**(1): 271-8.
14. Undurraga J, Baldessarini RJ. Randomized, placebo-controlled trials of antidepressants for acute major depression: thirty-year meta-analytic review. *Neuropsychopharmacology* 2012; **37**(4): 851-64.
  15. Furukawa TA, Salanti G, Atkinson L, et al. Comparative efficacy and acceptability of first- and second-generation antidepressants in the acute treatment of major depression: Protocol for a network meta-analysis. *BMJ open* 2016; **6**: e010919.
  16. Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011] Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org); 2011.
  17. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; **7**(3): 177-88.
  18. Furukawa TA, Cipriani A, Barbui C, Brambilla P, Watanabe N. Imputing response rates from means and standard deviations in meta-analyses. *Int Clin Psychopharmacol* 2005; **20**(1): 49-52.
  19. 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.
  20. Carmody TJ, Rush AJ, Bernstein I, et al. The Montgomery Asberg and the Hamilton ratings of depression: a comparison of measures. *Eur Neuropsychopharmacol* 2006; **16**(8): 601-11.
  21. Stein DJ, Baldwin DS, Dolberg OT, Despiegel N, Bandelow B. Which factors predict placebo response in anxiety disorders and major depression? An analysis of placebo-controlled studies of escitalopram. *J Clin Psychiatry* 2006; **67**(11): 1741-6.
  22. Agid O, Siu CO, Potkin SG, et al. Meta-regression analysis of placebo response in antipsychotic trials, 1970-2010. *Am J Psychiatry* 2013; **170**(11): 1335-44.
  23. Nikolakopoulou A1, Mavridis D, Salanti G. Demystifying fixed and random effects meta-analysis. *Evid Based Ment Health* 2014;**17**(2):53-7.
  24. Drazen JM. Sharing Individual Patient Data from Clinical Trials. *N Engl J Med* 2015;**372**:201-202
  25. Mancini M, Wade AG, Perugi G, Lenox-Smith A, Schacht A. Impact of patient selection and study characteristics on signal detection in placebo-controlled trials with antidepressants. *J Psychiatr Res* 2014;**51**:21-9.
  26. Weimer K, Colloca L, Enck P. Placebo effects in psychiatry: mediators and moderators. *Lancet Psychiatry* 2015;**2**(3):246-57.
  27. Fava M. Implications of a Biosignature Study of the Placebo Response in Major Depressive Disorder. *JAMA Psychiatry* 2015;**72**(11):1073-4.

28. Cipriani A, Geddes JR. Placebo for depression: we need to improve the quality of scientific information but also reject too simplistic approaches or ideological nihilism. *BMC Med* 2014;**12**:105.
29. Geddes JR, Cipriani A. Time to abandon placebo control in pivotal phase III trials? *World Psychiatry* 2015;**14**(3):306-7.
30. Rutherford BR, Roose SP: A model of placebo response in antidepressant clinical trials. *Am J Psychiatry* 2013;**170**:723–733.
31. Norman TC, Bountra C, Edwards AM, Yamamoto KR, Friend SH. Leveraging crowdsourcing to facilitate the discovery of new medicines. *Sci Transl Med* 2011;**3**(88):88mr1.

**Table 1. Proportion of placebo responders by half-decades.**

| <b>Years</b>     | <b>No of studies</b> | <b>Range of proportions of responders</b> | <b>Weighted mean proportion of responders (95% CI)</b> | <b>I-square, %</b> | <b>Tau</b> |
|------------------|----------------------|---|--|--------------------|------------|
| <b>1978-1985</b> | 27                   | 0.00 to 0.41                              | 0.30 (0.27 to 0.33)                                    | 12.0               | 0.087      |
| <b>1986-1990</b> | 32                   | 0.08 to 0.50                              | 0.28 (0.25 to 0.31)                                    | 32.1               | 0.171      |
| <b>1991-1995</b> | 34                   | 0.14 to 0.54                              | 0.35 (0.32 to 0.39)                                    | 71.6               | 0.232      |
| <b>1996-2000</b> | 39                   | 0.21 to 0.53                              | 0.38 (0.35 to 0.41)                                    | 69.3               | 0.181      |
| <b>2001-2005</b> | 47                   | 0.18 to 0.58                              | 0.40 (0.37 to 0.42)                                    | 72.8               | 0.176      |
| <b>2006-2010</b> | 46                   | 0.23 to 0.70                              | 0.37 (0.34 to 0.40)                                    | 78.3               | 0.207      |
| <b>2011-2015</b> | 27                   | 0.17 to 0.63                              | 0.36 (0.33 to 0.40)                                    | 84.9               | 0.232      |
| <b>All years</b> | 252                  | 0.00 to 0.70                              | 0.36 (0.35 to 0.37)                                    | 74.1               | 0.211      |

CI: Confidence Interval

Table 2. Factors influencing proportion of placebo responders before 2000

| <i>Covariates in the model</i>                       | <i>No of studies</i> | <i>Study year RR (95% CI)</i><br><i>[for every 5-year increase]</i> | <i>Covariate RR (95% CI)</i><br><i>[interpretation]</i>   | <i>Tau</i> |
|--|----------------------|---|---|------------|
| <b>Model with year as covariate</b>                  |                      |   |   |            |
| <b>Year</b>  | 132                  | 1.10 (1.05 to 1.15)   | -   | 0.196      |
| <b>Models with year and one additional covariate</b> |                      |   |   |            |
| <b>Additional covariate</b>                          |                      |   |   |            |
| <b>Age</b>   | 119                  | 1.10 (1.05 to 1.15)   | 1.00 (0.92 to 1.08)<br>[for every 10-year increase in mean age]                                       | 0.198      |
| <b>Sex</b>   | 57                   | 1.10 (1.02 to 1.18)   | 1.03 (1.00 to 1.07)<br>[for every 10% increase in proportion of women]                                | 0.208      |
| <b>Regions</b>                                       | 130                  | 1.10 (1.05 to 1.15)   | 0.94 (0.77 to 1.14)<br>[North America vs Europe]<br>1.03 (0.81 to 1.31) [Cross-continental vs Europe] | 0.202      |
| <b>Number of study centers</b>                       | 118                  | 1.07 (1.02 to 1.11)   | 1.37 (1.21 to 1.56)<br>[Multi-center vs single-center]  | 0.180      |
| <b>Setting*</b>                                      | 66                   | 1.06 (1.00 to 1.14)   | 0.86 (0.70 to 1.06) [Secondary/Tertiary care vs primary care]   | 0.205      |
| <b>Patient status**</b>                              | 119                  | 1.10 (1.05 to 1.15)   | 1.00 (0.87 to 1.16) [Outpatients vs inpatients]   | 0.202      |
| <b>Baseline eligibility threshold***</b>             | 108                  | 1.09 (1.04 to 1.14)   | 0.96 (0.85 to 1.07)<br>[For every 5 point increase in HAM-D17 threshold score]                        | 0.193      |
| <b>Baseline severity***</b>                          | 111                  | 1.09 (1.04 to 1.14)   | 0.91 (0.83 to 1.00)<br>[For every 5 point increase in baseline HAM-D17]                               | 0.181      |

|  |     |                     |  |       |
|--|-----|---------------------|--|-------|
| <b>Number of arms</b>  | 132 | 1.10 (1.05 to 1.15) | 1.13 (0.78 to 1.63)<br>[one arm more]          | 0.198 |
| <b>Placebo run-in</b>  | 106 | 1.10 (1.04 to 1.15) | 0.89 (0.75 to 1.06) [Present vs absent]        | 0.192 |
| <b>Flexible vs fixed dose regimen</b>                                  | 131 | 1.10 (1.06 to 1.15) | 1.13 (1.01 to 1.27)<br>[Flexible vs Fixed]     | 0.187 |
| <b>Length of trial (in weeks)</b>                                      | 124 | 1.05 (1.00 to 1.11) | 1.05 (1.02 to 1.08)<br>[one week more]         | 0.182 |
| <b>Rescue medication</b>   | 57  | 1.07 (0.99 to 1.15) | 0.90 (0.76 to 1.07)<br>[Present vs absent]     | 0.191 |
| <b>Number of randomised patients</b>                                   | 132 | 1.08 (1.03 to 1.13) | 1.11 (1.02 to 1.22)<br>[100 participants more] | 0.196 |
| <b>Multivariable meta-regression (Using all significant variables)</b> |     |                     |  |       |
| <b>Study year</b>  | 112 | 1.04 (0.99 to 1.09) |  | 0.160 |
| <b>Number of study centers</b>   |     |                     | 1.41 (1.22 to 1.62)                            |       |
| <b>Flexible vs fixed dose regimen</b>                                  |     |                     | 1.17 (1.04 to 1.31)                            |       |
| <b>Length of trial</b>   |     |                     | 1.04 (1.01 to 1.06)                            |       |
| <b>Number of randomised patients</b>                                   |     |                     | 0.97 (0.89 to 1.07)                            |       |

RR: Relative Risk, CI: Confidence Interval.

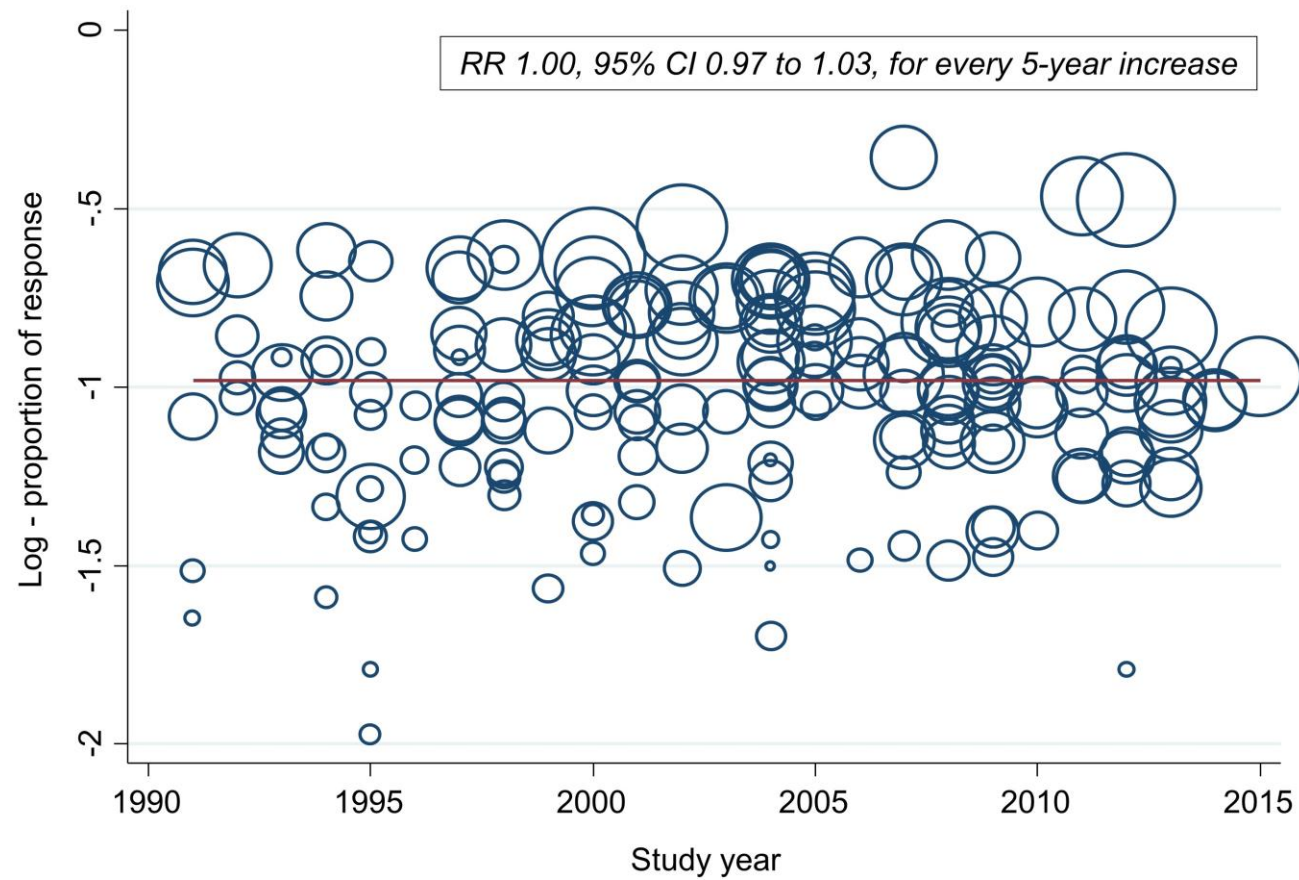
\*: Studies conducted in both primary and secondary/tertiary care were classified as ones in primary care.

\*\*: Studies recruiting both in- and outpatients were classified as ones recruiting inpatients.

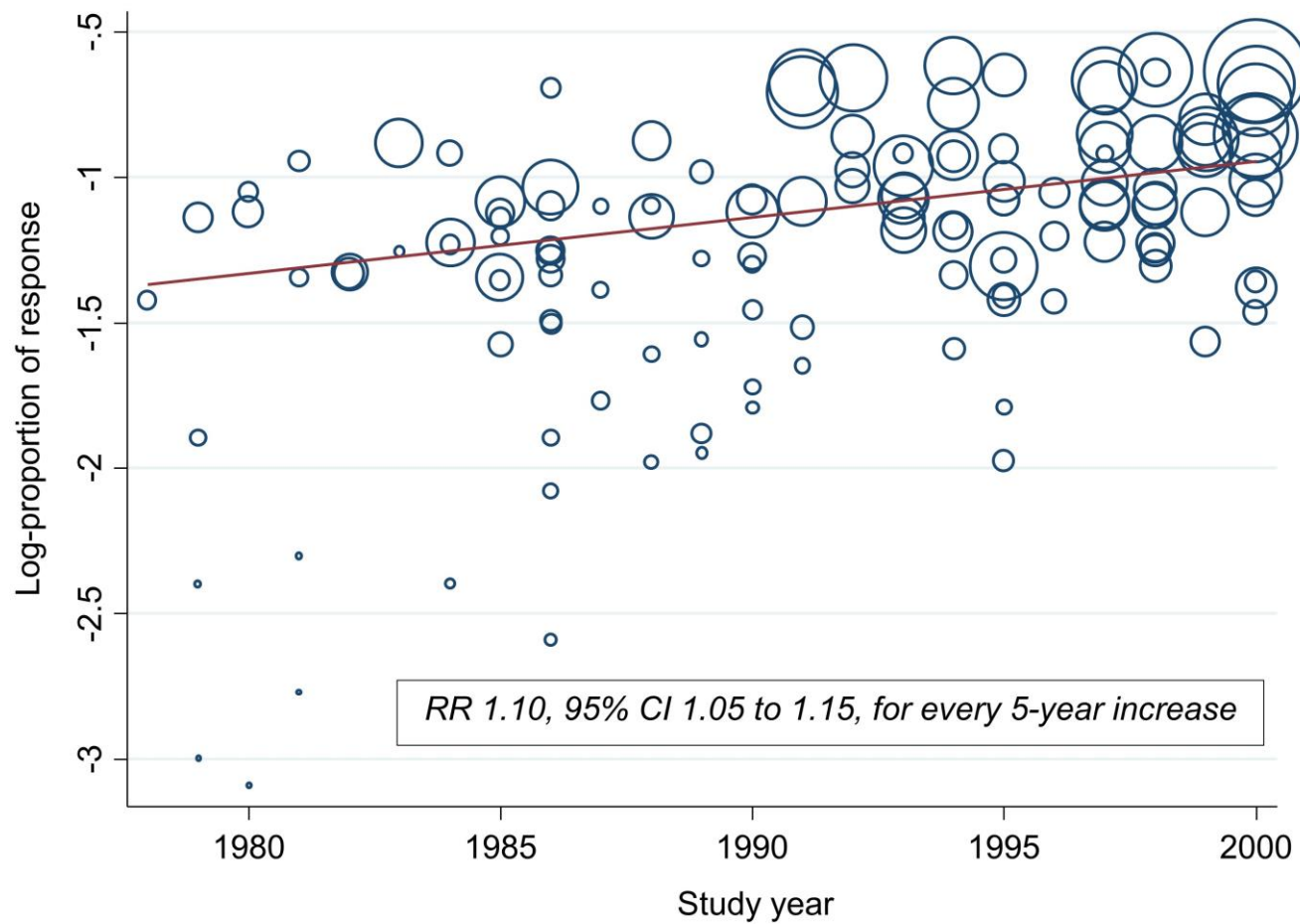
\*\*\*: Converted into Hamilton Rating Scale for Depression-17 using the conversion table<sup>20</sup>

**Table 3. Secular changes in study characteristics of placebo-controlled trials of antidepressants**

| <b>Years</b>     | <b>No of studies</b> | <b>Length of trial, mean weeks (range)</b> | <b>Multi-center trials, %</b> | <b>Flexible dosing, %</b> |
|------------------|----------------------|--|-------------------------------|---------------------------|
| <b>1978-1985</b> | 27                   | 4.6 (4-6)                                  | 61.9                          | 85.2                      |
| <b>1986-1990</b> | 32                   | 5.8 (4-8)                                  | 36.7                          | 75.0                      |
| <b>1991-1995</b> | 34                   | 6.7 (4-12)                                 | 75.9                          | 73.5                      |
| <b>1996-2000</b> | 39                   | 8.0 (4-12)                                 | 89.5                          | 76.3                      |
| <b>2001-2005</b> | 47                   | 7.8 (6-12)                                 | 93.6                          | 42.6                      |
| <b>2006-2010</b> | 46                   | 7.6 (6-12)                                 | 100                           | 40.0                      |
| <b>2011-2015</b> | 27                   | 8.1 (6-12)                                 | 92.6                          | 11.1                      |
| <b>All years</b> | 252                  | 7.1 (4-12)                                 | 81.9                          | 56.8                      |



**Figure 1. Meta-regression of log-proportion of response to placebo by study year since 1991 (193 studies).** RR: Relative Risk; CI: Confidence Interval.



**Figure 2. Meta-regression of log-proportion of response to placebo by study year up to 2000 (132 studies).** RR: Relative Risk; CI: Confidence Interval.