

Has the rising placebo response impacted antidepressant clinical trial outcome? Data from the US Food and Drug Administration 1987-2013

Arif Khan^{1,2}, Kaysee Fahl Mar¹, Jim Faucett¹, Shirin Khan Schilling^{1,3}, Walter A. Brown⁴

¹Northwest Clinical Research Center, Bellevue, WA, USA; ²Department of Psychiatry, Duke University School of Medicine, Durham, NC, USA; ³Department of Psychiatry, University of Connecticut, Hartford, CT, USA; ⁴Department of Psychiatry and Human Behavior, Brown University, Providence, RI, USA

More than fifteen years ago, it was noted that the failure rate of antidepressant clinical trials was high, and such negative outcomes were thought to be related to the increasing magnitude of placebo response. However, there is considerable debate regarding this phenomenon and its relationship to outcomes in more recent antidepressant clinical trials. To investigate this, we accessed the US Food and Drug Administration (FDA) reviews for sixteen antidepressants (85 trials, 115 trial arms, 23,109 patients) approved between 1987 and 2013. We calculated the magnitude of placebo and antidepressant responses, antidepressant-placebo differences, as well as the effect sizes and success rates, and compared these measures over time. Exploratory analysis investigated potential changes in trial design and conduct over time. As expected, the magnitude of placebo response has steadily grown in the past 30 years, increasing since 2000 by 6.4% ($r=0.46$, $p<0.001$). Contrary to expectations, a similar increase has occurred in the magnitude of antidepressant response (6.0%, $r=0.37$, $p<0.001$). Thus, the effect sizes (0.30 vs. 0.29, $p=0.42$) and the magnitude of antidepressant-placebo differences (10.5% vs. 10.3%, $p=0.37$) have remained statistically equivalent. Furthermore, the frequency of positive trial arms has gone up in the past 15 years (from 47.8% to 63.8%), but this difference in frequency has not reached statistical significance. Trial design features that were previously associated with a possible lower magnitude of placebo response were not implemented, and their relationship to the magnitude of placebo response could not be replicated. Of the 34 recent trials, two implemented enhanced interview techniques, but both of them were unsuccessful. The results of this study suggest that the relationship between the magnitude of placebo response and the outcome of antidepressant clinical trials is weak at best. These data further indicate that antidepressant-placebo differences are about the same for all of the sixteen antidepressants approved by the FDA in the past thirty years.

Key words: Antidepressants, clinical trials, placebo response, antidepressant-placebo difference, effect size, success rate, enhanced interview techniques

(*World Psychiatry* 2017;16:181–192)

Fifteen years following the advent of several new antidepressants in the mid-1980s, it became evident that the “success” rate of antidepressant clinical trials was low; less than 50% of trials demonstrated statistical superiority for antidepressants over placebo^{1,2}. Following Walsh et al’s finding³ of a rising placebo response, it was assumed that the clinical trial failure rate was related to this phenomenon⁴.

Investigators have attempted to determine if the increasing placebo response in antidepressant clinical trials observed by Walsh et al³ continues to this day. Meta-analytic reviews of antidepressant clinical trials^{5,6}, or psychotropic trials in general⁷, as well as patient-level data in trials for major depression⁸ have converged in showing that the placebo response has continued to grow over the past 15 years. Furthermore, Khin et al⁹ conducted an internal review for the US Food and Drug Administration (FDA), which seemed to confirm that the magnitude of placebo response was continuing to increase. Although this group of investigators had access to specific data, they did not identify the antidepressant trials that they reviewed.

One discordant voice is a study published by Furukawa et al¹⁰, which contradicts the observation of an increase in placebo response rate in more recent trials. These investigators conducted a review of 252 depression studies, examining the rate of therapeutic response to placebo using various dependent measures. They surmised that the proportion of placebo

responders, defined as patients with 50% or greater reduction in depressive symptoms, had remained the same after 1991. However, no mechanism was offered to explain this shift from a growing placebo response to a steady one¹¹, nor did the authors evaluate the effect of such a phenomenon on the outcome of antidepressant clinical trials.

Concern over the impact of increasing placebo response on antidepressant clinical trials has fueled a line of inquiry looking for variables predicting higher rates of placebo response, based on *post-hoc* analyses^{12,13}. Several hypotheses, such as the idea that more severely depressed patients might be relatively non-responsive to placebo, have been proposed on the basis of associative observations from these analyses^{14,15}. However, prospectively selecting more severely depressed patients for antidepressant clinical trials has neither resulted in a reduction in magnitude of the placebo response nor in enhanced antidepressant-placebo differences¹⁶.

Research has illuminated other possible variables, such as the flexible dosing of the investigational antidepressant, potentially showing a relationship to reduction of placebo response¹⁷. This flexible dosing schedule has been suggested for use in antidepressant clinical trials but, as of now, not fully implemented. Furthermore, retrospective analysis of earlier trials has found that placebo response is higher in trials of longer duration¹⁸ compared to shorter ones, although this phenomenon has not been tested prospectively.

Another hypothesis has been that the magnitude of placebo response and its variability was related to the low reliability among clinicians assessing depressed patients¹⁹⁻²¹. It was then recommended that patient sessions should be audio- or video-taped and audited by a centralized group of specifically trained raters to increase reliability. This type of enhanced interviewing technique has been implemented, although its effects on the outcome of more recent antidepressant trials remain questionable^{22,23}.

What stands out from these studies aiming to elucidate factors possibly mitigating placebo response in antidepressant clinical trials is that such factors are elusive and complex, and that their predictive ability varies across different contexts²⁴. This lack of fruitfulness in pinpointing what may moderate placebo response in antidepressant clinical trials has led to a form of therapeutic nihilism.

In fact, following the observation that antidepressant efficacy in clinical trials appears more robust when severely depressed patients are included, and that antidepressants do not reliably perform better than placebo, criticism has been raised regarding antidepressant's overall therapeutic efficacy and ability to treat the more mildly depressed population²⁵⁻²⁸. However, other investigators do not agree with this view, contending that the magnitude of the antidepressant-placebo response in clinical trials does not reflect the actual therapeutic efficacy of antidepressants in ordinary clinical practice²⁹⁻³².

However, in the midst of this investigative history, it has become obvious that expectations for antidepressant effect have changed as use of psychiatric medications has increased exponentially in the past 30 years³³. For example, currently one in six adults in the US are reported to have taken a psychiatric medication (primarily antidepressants) in the past year³⁴, potentially indicating high regards for antidepressant efficacy.

This observation of a potential increase in expectations for antidepressants has given credence to the theory that placebo response has increased due to the heightened expectations of clinicians and patients. Specifically, studies investigating this theory³⁵⁻³⁷ showed that the higher the risk of receiving placebo in an antidepressant clinical trial, the lower was the magnitude of placebo response. The caveat is that this theory has not been fully tested prospectively.

Given the possibility that the magnitude of placebo response continued to increase in recent antidepressant clinical trials and may have impacted the outcome of these trials, we conducted the present study. We evaluated data from the medical and statistical reviews of sixteen antidepressant programs approved by the FDA from 1987 to 2013, comparing the earlier antidepressant clinical trials to more recent ones.

We decided to conduct this analysis using the FDA clinical trial database³⁸ for several reasons. First, these data are not influenced by publication/investigator/analysis bias, while these selectivity biases are common in the published literature^{39,40}. Second, findings are verified at the source by the FDA staff in order to authenticate them. Third, reviews conducted by the FDA more often provide an analysis that includes the

magnitude of antidepressant response as well as the magnitude of placebo response, and clearly report the statistical analysis used for efficacy approval of the antidepressant. Last, this database is very large, with patient numbers in the tens of thousands, allowing to observe patterns with more confidence.

We hypothesized that the magnitude of placebo response has continued to increase in more recent antidepressant clinical trials, and that such an increase in placebo response may have reduced the frequency of successful trials. Also, we theorized that an increase in placebo response would correspond to a decrease in the antidepressant-placebo differences and observed effect sizes of more recent antidepressant clinical trials. Lastly, we explored if any of the research design features or enhanced interview techniques proposed to help contain placebo response have been implemented, and if so, with what results.

METHODS

Selection of trials

For the purpose of determining if the pattern of increasing placebo response continued in antidepressant clinical trials following Walsh et al's observation³, we formulated groups based on this point in time. We assigned each trial for an investigational antidepressant to the year that the antidepressant was approved, and grouped trials into pre-2000 and post-2000 ones.

We included only acute, parallel-group, double-blind, placebo-controlled trials for investigational antidepressants approved after registering a new drug application (NDA) program with the FDA. Trials were included if they enrolled adult patients with a primary diagnosis of major depressive disorder.

Data from treatment arms evaluating active comparator antidepressants (approved antidepressants not under investigation) were excluded from this analysis, due to the fact that the focus of this examination was to characterize new antidepressants in the process of gaining approval, not performance of established antidepressants.

In addition, we excluded data from treatment arms of investigational antidepressants at dosing levels not approved by the FDA, as shown in product labeling. Therefore, we examined only the clinical trial data from arms with doses expected to guide approved use of the investigational antidepressant.

We excluded depression trials enrolling only geriatric (>65 years old) patients, children (<18 years old) and inpatients, as well as relapse prevention or maintenance studies, as it is not possible to draw comparisons between trials studying unique populations or with confounding differences in experimental design.

Trials included and excluded in this analysis

After review of the FDA database for NDA registrations approved between 1987 and 2013, we identified a total of

sixteen adult depression programs for inclusion in the analysis. The investigational antidepressants (with year of approval) were: fluoxetine hydrochloride (1987), sertraline hydrochloride (1991), paroxetine hydrochloride (1992), venlafaxine hydrochloride (1993), nefazodone hydrochloride (1994), mirtazapine (1996), bupropion hydrochloride SR (1996), venlafaxine hydrochloride ER (1997), citalopram (1998), escitalopram oxalate (2002), duloxetine hydrochloride (2002), desvenlafaxine succinate (2008), trazodone hydrochloride ER (2010), vilazodone hydrochloride (2011), levomilnacipran hydrochloride (2013) and vortioxetine hydrobromide (2013).

These programs comprised a total of 125 efficacy evaluation trials. We excluded 40 trials after applying our selection criteria: six were conducted in a geriatric population, 22 were uncontrolled, four were carried out in inpatients, four had a relapse prevention design, and four used doses not approved by the FDA. Thus, 85 registration trials were included in this analysis.

These 85 trials had 172 treatment arms: 33 were active comparators and 24 utilized a dose of the investigational antidepressant not approved by the FDA. After excluding these 57 arms, 115 active treatment arms of investigational antidepressants at approved doses remained for analysis.

Data analysis

The medical and statistical reviews conducted by the FDA contain the published results of efficacy analysis along with the treatment group raw baseline and change scores on the primary efficacy measure when available. We encountered alternative statistical methods for handling missing data from patient dropout in the reporting and analysis of these efficacy data. These methods included observed cases analysis, analysis of covariance, and last observation carried forward (LOCF). Since data from LOCF analysis were available for all of the trials, we decided to use data (primary efficacy measure scores, p values, and patient numbers) from these LOCF statistical computation tables.

We decided to calculate percent symptom reduction as our measure of response magnitude. We divided the mean change score reported in the FDA reviews by the mean baseline score and multiplied by -100 to get a percent symptom reduction that takes into account variation in baseline and different measurement scales. This measure was calculated for placebo and antidepressant treatment groups separately.

We calculated the average antidepressant-placebo difference, taken by subtracting the placebo percent symptom reduction from the antidepressant percent symptom reduction for each trial arm. In instances where placebo had a greater percent symptom reduction than antidepressant, this measure would be negative.

Success of a treatment arm was defined as it is in the FDA reviews, with a p value threshold of 0.05 for endpoint analysis of the primary efficacy measure.

We calculated effect sizes for individual treatment arm comparisons using Hedges' g formula. This procedure has been used

in previous analyses of antidepressant clinical trials^{39,41}. As noted in Turner et al's paper³⁹, the formula for calculation of Hedges' g requires baseline scores, change scores and confidence intervals, as well as number of patients to generate t scores. NDA packets do not reliably report these data in full, and therefore we followed the statistical workaround method outlined in the supplement to Turner's paper⁴², using the inverse t score function in Microsoft Excel. Precise p values and degrees of freedom are imputed into the function to calculate a t score, which can be transformed to Hedges' g using a specific equation. Hedges' g effect size relies on number of patients and therefore is susceptible to sample size error. We used an appropriate correction to mitigate this risk.

Corrected Hedges' g scores were calculated for each trial arm. We examined effect sizes for trial arms as opposed to means for the trial overall, because the FDA evaluates efficacy for trial arms separately and uses these individual comparisons to support efficacy claims. Since the FDA approval process considers these individual comparisons, we wanted to examine individual treatment arm effects sizes both to retain the variability of signal detection among differing dose levels as well as to replicate the data handling of the FDA approval process. To generate a mean effect size for the two groups of pre-2000 and post-2000, we weighted the corrected effect size by the degrees of freedom to further account for sample size error.

All statistics were performed with IBM Statistical Package for Social Sciences (SPSS). Independent sample t tests were used to compare means from older antidepressant trial arms to the more recent ones, to evaluate if any significant changes had occurred in the distribution of scores from outcome measures. Correlations between year of new drug approval and percent symptom reduction, and between year of new drug approval and mean program effect size, were calculated using Pearson's coefficient.

We calculated frequency of trial design characteristics, including duration (≥ 8 weeks and < 8 weeks), number of trial arms (2 arms or ≥ 3 arms), and dosing schedule (fixed or flexible). We computed percentages of trials using either category of design feature and used chi-square analysis of proportions to explore any evidence of systematic implementation.

Statistical analysis of the results of trials using enhanced interview assessment techniques was not possible because only two (vortioxetine 317 and levomilnacipran MD-02) recent trials out of 34 used such techniques.

RESULTS

Tables 1 and 2 report the program/trial essential characteristics. Prior to 2000, there were nine antidepressant NDA programs, contributing 51 trials and 67 active treatment arms from efficacy tables that met our inclusion/exclusion criteria. The seven programs approved after 2000 supplied 34 trials and 48 active treatment arms for analysis. Four (6.0%) of the treatment arms in pre-2000 trials and 13 (27.1%) of the treatment

Table 1 Characteristics of 51 clinical trials for the approval of nine antidepressants from 1987 to 2000

Protocol number	Dosing schedule	Duration (weeks)	Primary efficacy measure	Placebo			Investigational antidepressant			p value for efficacy calculation	Effect size (Hedges' g)
				No. patients	Baseline/change score on primary efficacy measure	Percent response	No. patients	Baseline/change score on primary efficacy measure	Percent response		
<i>Fluoxetine (1987)</i>											
19	Flexible	4	HAM-D	24	28.2/-5.5	19.5	22	28.6/-12.5	43.7	0.011	0.77
27	Flexible	6	HAM-D	163	28.2/-8.4	29.8	181	27.5/-11.0	40.0	0.012	0.27
25	Flexible	4	HAM-D	24	25.8/-8.8	34.1	18	26.2/-7.2	27.5	0.50	0.21
62-A	Fixed	6	HAM-D	56	4.0/-1.21	30.2	105	3.9/-1.33	34.1	0.46	0.12
				103	3.9/-1.38		103	3.9/-1.38	35.4	0.34	0.16
62-B	Fixed	6	HAM-D	48	24.3/-5.7	23.4	100	3.9/-1.19	30.5	0.50	0.11
				97			97	24.7/-9.8	39.6	0.007	0.48
				97			97	24.1/-9.6	39.8	0.01	0.46
				103			103	24.2/-7.2	29.8	0.34	0.17
<i>Sertraline (1991)</i>											
103	Fixed	6	HAM-D	86	25.3/-7.6	30.0	90	24.8/-10.6	42.7	0.34	0.32
				89			89	24.9/-9.8	39.4	0.102	0.25
				82			82	25.7/-9.9	38.5	0.252	0.23
104	Flexible	8	HAM-D	141	25.4/-8.2	35.0	142	23.3/-11.7	50.2	0.001	0.40
315	Flexible	8	HAM-D	73	22.2/-6.7	29.6	76	23.1/-8.8	38.1	0.46	0.12
<i>Paroxetine (1992)</i>											
01-001	Flexible	6	HAM-D	24	27.4/-10.5	38.3	24	28.0/-13.5	48.2	0.204	0.37
02-001	Flexible	6	HAM-D	53	25.9/-6.8	26.3	51	26.6/-12.3	46.2	0.0042	0.57
02/002	Flexible	6	HAM-D	34	24.9/-5.8	23.3	36	25.0/-10.9	43.6	0.0146	0.60
02/003	Flexible	6	HAM-D	33	28.9/-7.2	24.9	33	28.6/-9.7	33.9	0.3092	0.25
02-004	Flexible	4	HAM-D	38	27.3/-7.2	26.4	36	28.9/-12.2	42.2	0.0018	0.75
03-005	Flexible	6	HAM-D	42	26.8/-4.0	14.9	40	26.1/-10.0	38.3	0.0076	0.60
03-006	Flexible	6	HAM-D	38	28.7/-3.0	10.5	39	29.7/-9.1	30.6	0.0014	0.75
03-001	Flexible	4	HAM-D	38	24.8/-4.7	19.0	40	24.9/-10.8	42.4	0.006	0.63
03-002	Flexible	4	HAM-D	40	25.6/-6.2	24.2	40	24.9/-8.0	32.1	0.0004	0.61
03-003	Flexible	4	HAM-D	42	27.0/-9.2	34.1	39	25.7/-9.3	36.2	0.98	0.01
03-004	Flexible	4	HAM-D	37	27.0/-6.7	24.8	37	27.6/-10.4	37.7	0.04	0.48
09	Fixed	12	HAM-D	51	NR/-8.2	NR	104	NR/-10.6	NR	0.34	0.16
				99			99	NR/-9.0	NR	0.34	0.16
				100			100	NR/-9.4	NR	0.34	0.16

Table 1 Characteristics of 51 clinical trials for the approval of nine antidepressants from 1987 to 2000 (*continued*)

Protocol number	Dosing schedule	Duration (weeks)	Primary efficacy measure	Placebo			Investigational antidepressant			p value for efficacy calculation	Effect size (Hedges' g)
				No. patients	Baseline/change score on primary efficacy measure	Percent response	No. patients	Baseline/change score on primary efficacy measure	Percent response		
Venlafaxine (1995)											
600A-206	Flexible	4	HAM-D	47	28.6/-4.8	16.8	46	28.2/-14.2	50.4	0.006	0.58
600A-301	Flexible	6	HAM-D	78	24.6/-9.5	38.6	64	25.4/-13.9	54.7	0.0004	0.61
600A-302	Flexible	6	HAM-D	75	24.4/-8.9	36.5	65	25.0/-11.9	47.6	0.008	0.45
600A-303	Flexible	6	HAM-D	79	24.6/-9.9	40.2	70	23.6/-10.1	42.8	0.493	0.11
600A-203	Fixed	6	HAM-D	92	25.3/-6.7	26.5	77	26.0/-11.1	42.7	0.004	0.45
600A-313	Fixed	6	HAM-D	75	25.4/-9.5	37.4	72	25.6/-10.9	42.6	0.193	0.21
600A-313	Fixed	6	HAM-D	77	25.6/-11.8	46.1	77	25.6/-11.8	46.1	0.142	0.24
Nefazodone (1994)											
030A2-0007	Fixed	6	HAM-D	47	26.4/-9.8	37.1	47	25.4/-10.7	42.1	0.60	0.11
03A0A-003	Fixed	6	HAM-D	45	25.9/-6.8	26.3	44	25.4/-11.0	43.3	0.03	0.46
03A0A-004A	Fixed	6	HAM-D	77	23.5/-8.5	36.2	76	23.6/-9.0	38.1	0.66	0.71
03A0A-004B	Fixed	6	HAM-D	80	25.0/-9.4	37.6	78	25.4/-12.4	48.8	0.02	0.37
CN104-005	Flexible	8	HAM-D	90	23.5/-8.0	34.0	86	24.4/-12.0	49.2	0.01	0.39
CN104-006	Flexible	8	HAM-D	78	23.8/-8.9	37.4	80	23.5/-10.0	42.6	0.35	0.15
Mirtazapine (1996)											
003-002	Flexible	6	HAM-D	44	24.7/-5.4	21.9	44	24.2/-11.7	48.3	0.0008	0.73
003-003	Flexible	6	HAM-D	45	25.5/-8.8	34.5	45	25.4/-10.4	40.9	0.49	0.14
003-008	Fixed	6	HAM-D	28	25.8/-9.6	37.2	30	26.0/-7.6	29.2	0.293	-0.28
003-020/3220	Flexible	6	HAM-D	39	29.5/-4.8	16.3	40	27.8/-10.3	37.1	0.004	0.66
003-021/3220	Flexible	6	HAM-D	48	24.4/-9.5	38.9	45	24.2/-11.7	48.3	0.223	0.25
003-022/3220	Flexible	6	HAM-D	50	31.2/-9.0	28.8	49	33.0/-16.1	48.8	0.003	0.61
003-024/3220	Flexible	6	HAM-D	48	27.7/-7.7	27.8	50	27.5/-12.1	44.0	0.01	0.53
85027	Flexible	4	HAM-D	61	26.2/-10.9	41.6	64	26.4/-13.4	50.8	0.189	0.23

Table 1 Characteristics of 51 clinical trials for the approval of nine antidepressants from 1987 to 2000 (continued)

Protocol number	Dosing schedule	Duration (weeks)	Primary efficacy measure	Placebo			Investigational antidepressant			p value for efficacy calculation	Effect size (Hedges' g)
				No. patients	Baseline/change score on primary efficacy measure	Percent response	No. patients	Baseline/change score on primary efficacy measure	Percent response		
Bupropion SR (1996)											
203	Fixed	8	HAM-D	117	23.2/-8.1	34.9	113	23.4/-10.2	43.6	0.04	0.27
205	Fixed	8	HAM-D	116	23.4/-8.3	35.5	111	23.6/-9.0	38.1	0.53	0.8
212	Fixed	8	HAM-D	148	23.9/-9.8	41.0	144	24.2/-9.3	38.4	0.30	0.14
Venlafaxine ER (1997)											
208	Flexible	12	HAM-D	91	24.6/-8.7	35.4	85	24.4/-14.9	61.1	0.001	0.50
209	Flexible	8	HAM-D	100	23.6/-6.8	28.8	91	24.5/-11.7	47.8	0.0003	0.53
367	Flexible	8	HAM-D	81	26.6/-13.1	49.2	85	26.5/-15.6	58.9	0.37	0.14
Citalopram (1998)											
85A	Flexible	4	HAM-D	78	33.7/-9.6	28.5	82	33.5/-12.9	38.5	0.0344	0.33
86141	Flexible	6	HAM-D	50	21.0/-4.9	23.3	97	22.2/-6.3	28.4	0.316	0.17
89303	Fixed	6	HAM-D	64	23.7/-10.6	44.7	61	23.0/-13.3	57.8	0.12	0.28
91206	Fixed	6	HAM-D	124	24.6/-9.3	37.8	120	24.4/-12.2	50.0	0.0025	0.39
89306	Fixed	6	MADRS	88	33.1/-16.0	48.3	97	24.5/-12.1	49.4	0.0053	0.37

NR – data not reported or censored in the Food and Drug Administration packet, HAM-D – Hamilton Depression Rating Scale, MADRS – Montgomery-Åsberg Depression Rating Scale
 Bold prints indicate a positive trial arm

Table 2 Characteristics of 34 clinical trials for the approval of seven antidepressants after 2000

Protocol number	Dosing schedule	Duration (weeks)	Primary efficacy measure	Placebo		Investigational antidepressant				p value for efficacy calculation	Effect size (Hedges' g)	
				No. patients	Baseline/change score on primary efficacy measure	Percent response	No. patients	Baseline/change score on primary efficacy measure	Percent response			
<i>Escitalopram (2002)</i>												
MD01	Fixed	8	MADRS	119	29.5/-9.4	31.9	118	28.0/-12.8	45.7	0.0007	0.45	
MD 02	Flexible	8	MADRS	125	28.8/-11.2	38.9	125	28.9/-13.9	48.1	0.0001	0.51	
99001	Fixed	8	MADRS	189	28.7/-13.6	47.4	188	28.7/-12.9	45.0	0.251	0.15	
99003	Flexible	8	MADRS	154	28.7/-12.5	43.6	155	29.2/-16.3	55.8	0.006	0.28	
<i>Duloxetine (2002)</i>												
HMAQa	Flexible	8	HAM-D	57	20.6/-6.5	31.6	56	19.6/-8.5	43.4	0.15	0.27	
HMAQb	Flexible	8	HAM-D	55	20.0/-5.7	28.5	61	19.9/-6.2	31.2	0.95	0.01	
HMATa	Fixed	8	HAM-D	89	17.8/-4.3	24.2	81	17.4/-5.5	31.6	0.138	0.23	
HMATb	Fixed	8	HAM-D	88	17.2/-4.2	24.2	86	18.1/-7.7	42.7	0.003	0.45	
HMBHa	Fixed	9	HAM-D	115	21.1/-5.2	24.5	121	21.5/-9.3	43.0	0.001	0.43	
HMBHb	Fixed	9	HAM-D	136	20.5/-7.2	35.3	123	20.3/-8.9	43.8	0.048	0.25	
<i>Desvenlafaxine (2008)</i>												
332	Fixed	9	HAM-D	150	23.0/-9.6	41.7	150	23.4/-11.5	49.2	0.02	0.27	
333	Fixed	8	HAM-D	161	24.3/-10.8	44.4	147	23.4/-11.0	47.0	0.09	0.20	
223	Fixed	8	HAM-D	78	NR	NR	164	24.3/-13.2	54.3	0.004	0.32	
306	Fixed	8	HAM-D	118	NR/-7.7	NR	158	24.4/-13.7	56.2	0.001	0.37	
308	Fixed	8	HAM-D	124	NR/-9.5	NR	63	NR	NR	0.59	0.09	
304	Flexible	8	HAM-D	114	NR/NR	NR	72	NR	NR	0.52	0.11	
309	Flexible	8	HAM-D	120	NR/-12.5	NR	114	NR/-10.5	NR	0.004	0.38	
317	Flexible	8	HAM-D	125	NR/-9.8	NR	116	NR/-9.6	NR	0.076	0.23	
320	Flexible	8	HAM-D	118	NR/-7.5	NR	113	NR/-10.5	NR	0.002	0.41	
							121	NR/-12.6	NR	0.002	0.40	
							124	NR/-12.1	NR	0.008	0.34	
							120	NR/NR	NR	0.28	0.14	
							117	NR/-13.4	NR	0.381	0.11	
							110	NR/-10.5	NR	0.488	0.09	
							117	NR/-9.1	NR	0.078	0.23	

Table 2 Characteristics of 34 clinical trials for the approval of seven antidepressants after 2000 (continued)

Protocol number	Dosing schedule	Duration (weeks)	Primary efficacy measure	Placebo				Investigational antidepressant				p value for efficacy calculation	Effect size (Hedges' g)
				No. patients	Baseline/change score on primary efficacy measure	Percent response	No. patients	Baseline/change score on primary efficacy measure	Percent response				
										No. patients	Baseline/change score on primary efficacy measure		
<i>Trazodone ER (2010)</i>													
04ACL3-001	Flexible	8	HAM-D	206	22.4/−9.25	41.3	206	23.2/−11.2	48.2	0.0055	0.27		
<i>Vilazodone (2011)</i>													
GNSC-04-DP-02	Flexible	8	MADRS	199	30.7/−9.7	31.6	198	30.8/−12.9	41.9	0.001	0.33		
CLDA-07-DP-02	Fixed	8	MADRS	231	32.0/−10.8	33.8	232	31.9/−13.3	41.7	0.009	0.24		
<i>Levomilnacipran (2013)</i>													
MD-01	Fixed	8	MADRS	175	35.6/−11.6	32.6	176	36.0/−14.8	41.1	0.0186	0.25		
MD-03	Flexible	8	MADRS	214	35.2/−12.2	33.8	177	36.1/−15.6	43.2	0.0038	0.31		
MD-10	Flexible	8	MADRS	185	31.0/−11.3	36.5	176	36.0/−16.5	45.8	0.0005	0.37		
MD-02	Flexible	8	MADRS	182	35.5/−14.2	40.0	215	35.0/−15.3	43.7	0.0051	0.27		
F02695 LP2 02	Flexible	10	MADRS	277	30.5/−14.5	47.5	185	30.8/−14.6	47.4	0.0027	0.31		
<i>Vortioxetine (2013)</i>													
11492A	Fixed	6	MADRS	105	33.9/−14.5	42.8	187	31.2/−14.4	46.2	0.0043	0.30		
305	Fixed	8	HAM-D	139	32.7/−11.3	35.6	175	NR/NR	NR	NR	NR		
13267A	Fixed	8	MADRS	158	31.5/−11.7	37.1	276	30.7/−18.7	60.9	0.0001	0.55		
315US	Fixed	8	MADRS	153	31.5/−12.8	40.6	100	34.0/−20.2	59.4	0.0001	0.55		
316US	Fixed	8	MADRS	155	32.0/−10.8	33.8	139	33.1/−16.2	48.9	0.001	0.40		
11984A	Fixed	8	MADRS	145	NR/−14.8	NR	149	31.8/−17.2	54.1	0.0001	0.45		
317	Fixed	8	MADRS	149	33.4/−12.9	38.6	151	31.2/−18.8	60.3	0.0001	0.45		
							145	31.9/−14.3	44.8	0.224	0.14		
							147	32.0/−15.6	48.8	0.023	0.26		
							154	32.2/−13.0	40.4	0.058	0.19		
							148	32.5/−14.4	44.3	0.002	0.36		
							151	NR/−16.3	NR	0.185	0.15		
							143	34.1/−13.7	40.2	0.597	0.06		
							142	33.6/−13.4	39.9	0.745	0.04		

NR – data not reported or censored in the Food and Drug Administration packet, HAM-D – Hamilton Depression Rating Scale, MADRS – Montgomery-Åsberg Depression Rating Scale
 Bold prints indicate a positive trial arm

Table 3 Evaluation of efficacy outcomes in antidepressant registration trials before and after 2000

	Before 2000	After 2000	p
No. programs	9	7	
No. trials	51	34	
No. active treatment arms	67	48	
Successful treatment arms	47.8% (32/67)	63.8% (30/47)	0.09
% symptom reduction			
Antidepressant	40.6% (± 13.7)	46.6% (± 7.0)	0.005
Placebo	29.8% (± 12.6)	36.2% (± 6.6)	0.005
Mean antidepressant-placebo difference	10.5% (± 9.2)	10.3% (± 5.0)	0.37
Effect size (Hedges' g)	0.30 (± 0.24)	0.29 (± 0.12)	0.42

arms in post-2000 trials had missing baseline or change score data.

Due to missing data, we calculated placebo response magnitude based on 76 out of 85 placebo arms (89.4%), antidepressant response magnitude based on 98 out of 115 treatment arms (85.2%), antidepressant-placebo differences based on 98 out of 115 antidepressant-placebo group comparisons (85.2%), and effect sizes based on 114 out of 115 treatment arms (99.1%).

Prior to 2000, placebo reduced symptoms on average by 29.8% (± 12.6) compared to 36.2% (± 6.6) in programs post-2000, resulting in a significant increase in placebo response by

6.4% ($t = -2.9$, $df = 74$, $p = 0.005$). This represents a 21.5% change over 15 years (Table 3).

Percent symptom reduction as a measure of response magnitude increased by an almost identical 6.0% in the antidepressant treatment arm, from pre-2000 trials at 40.6% (± 13.7) to post-2000 trials at 46.6% (± 7.0) ($t = -2.9$, $df = 96$, $p = 0.005$). This represents a 14.8% change over 15 years (Table 3).

Figure 1 shows placebo and antidepressant response rates over time. Growth rate was nearly parallel in placebo and antidepressant treatments, with both treatment conditions having significant positive relationships (placebo: $r = 0.46$, $p < 0.001$; antidepressants: $r = 0.37$, $p < 0.001$) between time and percent symptom reduction.

The antidepressant-placebo differences have remained equivalent over the years, as a result of matching growth in both treatment condition responses. The mean antidepressant-placebo difference in trials from pre-2000 was 10.5% (± 9.2) as compared to 10.3% (± 5.0) in the trials post-2000 ($p = 0.37$) (Table 3).

Treatment arms for antidepressant clinical trials conducted prior to 2000 were successful in 47.8% of cases (32 out of 67 treatment arms), compared with a treatment arm success rate of 63.8% (30 out of 47) in antidepressant trials post-2000. Chi-square analysis of proportions determined that this difference was not statistically significant ($p = 0.09$).

Effect sizes based on number of patients and p values from individual treatment arm LOCF analysis revealed no significant change over the 31 years of antidepressant program data. The average weighted effect size across trial arms conducted

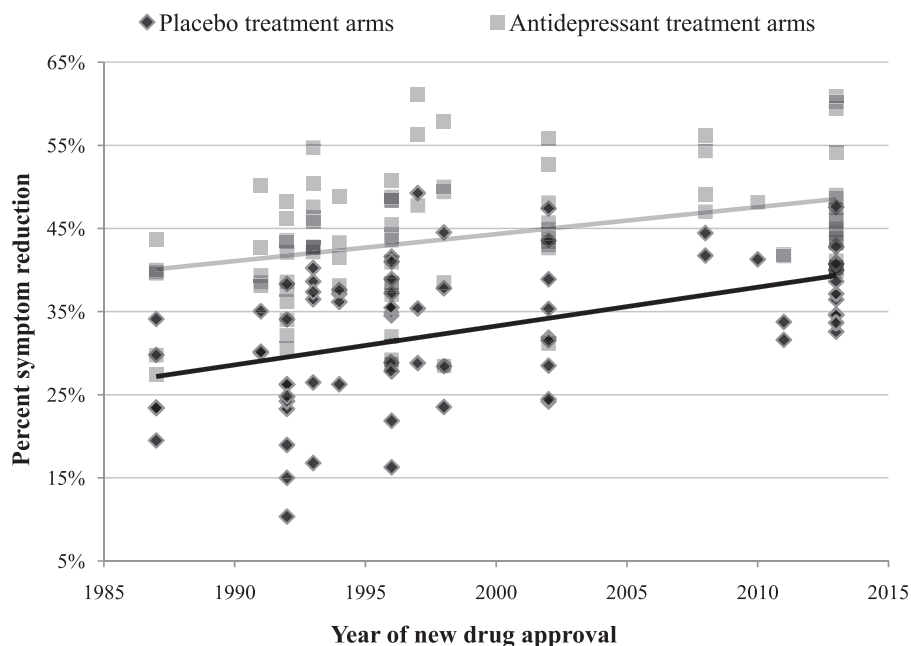


Figure 1 Percent symptom reduction in 74 placebo and 92 antidepressant treatment arms from 85 clinical trials for 16 antidepressant approval programs plotted with time. The correlation between year of new drug approval and percent symptom reduction was significant in both the placebo ($r = 0.46$, $p < 0.001$) and the antidepressant group ($r = 0.37$, $p < 0.001$).

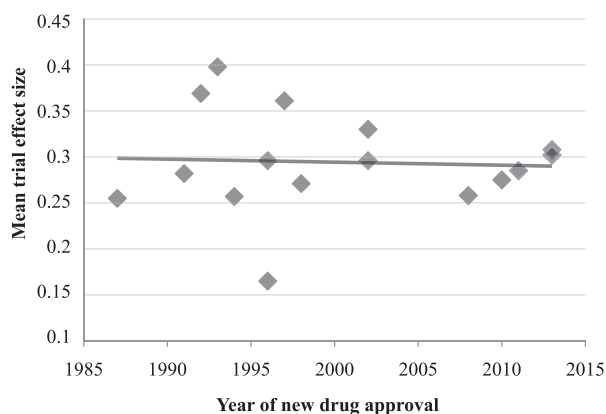


Figure 2 Mean effect size (Hedges' *g*) of antidepressant clinical trials based on year of approval. There was no significant relationship between year of new drug approval and mean program effect size ($r = -0.06$, $p = 0.85$).

before 2000 was 0.30 (± 0.24), while for trials after 2000 it was 0.29 (± 0.12) ($p = 0.42$).

Figure 2 shows this trend of stability in effect size throughout the years simplified by averaging trial effect sizes to generate overall values for each antidepressant program. Program effect sizes were not correlated with any kind of change over time ($p = 0.85$).

The trial design suggestions¹², including enhanced rater interview techniques¹⁹⁻²¹, put forth by investigators based on *post-hoc* analyses of placebo response were not implemented in recent clinical trials. Specifically, the trends examined were opposite in direction to the modifications in trial design previously suggested: trials were of longer duration, had a greater number of treatment arms, and rarely used flexible dosing schedules; all elements previously corresponding to higher placebo response. There was no observed association between trial design features and trial outcomes in post-2000 trials (see Table 2 for trial design characteristics).

Regarding enhanced interview techniques¹⁹⁻²¹, two out of 34 recent antidepressant clinical trials submitted for review by the FDA used such techniques. Neither of these (trial 317 for vortioxetine⁴³ and trial MD-02 for levomilnacipran⁴⁴) was successful.

DISCUSSION

Given the present state of uncertainty in the research surrounding placebo response in antidepressant clinical trials and the importance of this phenomenon, this study aimed to evaluate if placebo response as measured by symptom reduction has continued to rise over the past 15 years compared to the earlier 15 years. The study also attempted to determine if decreases in success rate and measures of antidepressant-placebo differences accompanied the growth in symptom reduction with placebo.

The study showed that the pattern of increase in placebo response noted in 2001 by Walsh et al³ has continued. The magnitude of symptom reduction with placebo has steadily increased from 29.8% to 36.2% ($p = 0.005$). These results converge with the findings by Khin et al⁹ and other investigators⁵⁻⁸ that placebo symptom reduction has continued to increase in more recent antidepressant clinical trials.

The increase in placebo response observed in recent antidepressant clinical trials is in contrast with a recent study by Furukawa et al¹⁰, reporting a stability in placebo response rate after 1991. We attribute this discrepancy to differences in study design. That study included data from published sources, which have been shown to contain selection bias^{39,40} and frequently use different statistical analyses from those performed by the FDA reviewers. Therefore, use of published sources may have resulted in different datasets. Additionally, our study used percent symptom reduction as a measure of placebo response and this value is on a continuum, allowing for analysis of more subtle changes than a binary measure such as number of patients meeting a therapeutic response threshold, as used by Furukawa et al¹⁰ and many others.

Contrary to expectations, given our finding of a continued increase in placebo response over time, the success rate of antidepressant clinical trials has gone up over the past 15 years (from 47.8% to 63.8%, $p = 0.09$). This has occurred as the magnitude of the antidepressant response has also gone up considerably (from 40.6% to 46.6%, $p = 0.005$).

In essence, both the magnitude of placebo response and antidepressant response have steadily increased over the past thirty years among these sixteen new antidepressant programs. The success rate of antidepressant trials has remained about the same, showing a modest increase in recent years. This is confirmed by the finding that treatment arm effect sizes have remained about the same, with a distribution around 0.30, and antidepressant-placebo differences continue to show a 10% antidepressant advantage regardless of placebo response. In other words, the newer antidepressants appear about as efficacious as the older ones.

Potential remedies that have been suggested in order to mitigate placebo response, such as changes in study designs (use of flexible dosing, shorter duration of trials, and fewer number of treatment arms¹²) seem not to have been systematically implemented or to have had effect on the outcomes of more recent antidepressant trials. Our exploration also suggests that these trial design and conduct factors may not be causally related to the magnitude of placebo response (see Table 2), so that the prospective implementation of these suggestions may not have the effect expected based on theory or observed from retrospective analysis. In particular, the two antidepressant clinical trials that prospectively used enhanced interview techniques failed to show superiority over placebo in NDA programs for vortioxetine and levomilnacipran.

In this context, it is important to note that the current results do not support earlier studies regarding the impact of

placebo response on trial outcomes, which found that the magnitude of placebo response was inversely associated with the frequency of positive outcomes in trials conducted between 1987 and 1999⁴. This relationship holds true for those earlier trials, but has dissolved in the more recent post-2000 trials.

What these current data show is that, in spite of the continuing growth of placebo response, antidepressants appear to maintain an advantage of about 10% (effect size of 0.30, a modest one), suggesting that acting to mitigate placebo response may not be a critical component of the success and outcomes of efficacy analysis in antidepressant clinical trials.

Potential mechanisms explaining the growth in placebo response and relationship to trial outcomes were not fully explored in this study. However, we noticed that there has been a substantial increase in the sample size in both placebo and antidepressant treatment arms in recent years. As described by Liu et al⁴⁴, increased sample size has been associated with clinical trial outcomes of investigational hypertension medications, and the relative mechanism calls for further exploration.

A drawback to our study is that it was an observational *post-hoc* analysis rather than prospective in design. More important, FDA medical and statistical reports do not include subject-level data. This summarization of data in FDA reviews of new investigational antidepressants does not allow a more detailed analysis. However, the sponsoring pharmaceutical companies or the FDA may undertake such an analysis to provide better insight into the relationship between placebo response and antidepressant clinical trial outcomes.

In conclusion, the results of this study suggest that the relationship between the magnitude of placebo response and the success of antidepressant clinical trials is weak at best. These data indicate that the antidepressant-placebo differences are about the same for all of the sixteen antidepressants approved by the FDA in the past thirty years. This finding has implications for guiding future clinical trials and warrants exploratory analysis of other potential factors that may influence the outcome of antidepressant trials.

REFERENCES

- Khan A, Warner HA, Brown WA. Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials. *Arch Gen Psychiatry* 2000;57:311-7.
- Khan A, Khan S, Brown WA. Are placebo controls necessary to test new antidepressants and anxiolytics? *Int J Neuropsychopharmacol* 2002;5:193-7.
- Walsh BT, Seidman SN, Sysko R et al. Placebo response in studies of major depression: variable, substantial, and growing. *JAMA* 2001;287:1840-7.
- Khan A, Detke M, Khan S et al. Placebo response and antidepressant clinical trial outcome. *J Nerv Ment Dis* 2003;191:211-8.
- Rief W, Nestoriuc Y, Weiss S et al. Meta-analysis of the placebo response in antidepressant trials. *J Affect Disord* 2009;118:1-8.
- Undurraga J, Baldessarini RJ. Randomized, placebo-controlled trials of antidepressants for acute major depression: thirty-year meta-analytic review. *Neuropsychopharmacology* 2012;37:851-64.
- Weimer K, Colloca L, Enck P. Placebo effects in psychiatry: mediators and moderators. *Lancet Psychiatry* 2015;2:246-57.
- Mancini M, Wade AG, Perugi G et al. Impact of patient selection and study characteristics on signal detection in placebo-controlled trials with antidepressants. *J Psychiatr Res* 2014;51:21-9.
- Khin NA, Chen Y, Yang Y et al. 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;72:464-72.
- Furukawa TA, Cipriani A, Atkinson LZ et al. Placebo response rates in antidepressant trials: a systematic review of published and unpublished double-blind randomized controlled studies. *Lancet Psychiatry* 2016;3:1059-66.
- Enck P. Placebo response in depression: is it rising? *Lancet Psychiatry* 2016;3:1005-6.
- Khan A, Kolts RL, Thase ME et al. Research design features and patient characteristics associated with the outcome of antidepressant clinical trials. *Am J Psychiatry* 2004;161:2045-9.
- Papakostas GI, Ostergaard SD, Iovieno N. The nature of placebo response in clinical studies of major depressive disorder. *J Clin Psychiatry* 2015;76:456-66.
- Khan A, Leventhal RM, Khan SR et al. Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. *J Clin Psychopharmacol* 2002;22:40-5.
- Khan A, Brodhead AE, Kolts RL et al. Severity of depressive symptoms and response to antidepressants and placebo in antidepressant trials. *J Psychiatr Res* 2005;29:145-50.
- Khan A, Schwartz K, Kolts RL et al. Relationship between depression severity entry criteria and antidepressant clinical trial outcomes. *Biol Psychiatry* 2007;62:65-71.
- Khan A, Khan SR, Walens G et al. 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;28:552-7.
- Khan A, Khan SR, Leventhal RM et al. Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials: a replication analysis of the Food and Drug Administration database. *Int J Neuropsychopharmacol* 2001;4:113-8.
- Demitrack MA, Faries D, Herrera JM et al. The problem of measurement error in multisite clinical trials. *Psychopharmacol Bull* 1998;34:19-24.
- Kobak K, Thase ME. Why do clinical trials fail? The problem of measurement error in clinical trials: time to test new paradigms? *J Clin Psychopharmacol* 2007;27:1-5.
- Kobak KA, Feiger AD, Lipsitz JD. Interview quality and signal detection in clinical trials. *Am J Psychiatry* 2005;162:628.
- Khan A, Fauceit J, Brown WA. Magnitude of placebo response and response variance in antidepressant clinical trials using structured, taped, and appraised rater interviews compared to traditional rating interviews. *J Psychiatr Res* 2014;51:88-92.
- Khan A, Fauceit J, Brown WA. Magnitude of change with antidepressants and placebo in antidepressant clinical trials using structured, taped and appraised rater interview (SIGMA-RAPS) compared to trials using traditional semi-structured interviews. *Psychopharmacology* 2014;231:4301-7.
- Stein DJ, Baldwin DS, Dolberg OT et al. Which factors predict placebo response in anxiety disorders and major depression? An analysis of placebo-controlled studies of escitalopram. *J Clin Psychiatry* 2006;67:1741-6.
- Kirsch K, Deacon BJ, Huedo-Medina TB et al. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med* 2008;5:e45.
- Fournier JC, DeRubeis RJ, Hollon SD. Antidepressant drug effects depression severity: a patient-level meta-analysis. *JAMA* 2010;303:47-53.
- Naudet F, Millet B, Charlier P et al. Which placebo to cure depression? A thought-provoking network meta-analysis. *BMC Med* 2013;11:230.
- Moncrieff J. Antidepressants: misnamed and misrepresented. *World Psychiatry* 2015;14:302-3.
- Fountoulakis KN, Möller H. Efficacy of antidepressants: a re-analysis and re-interpretation of the Kirsch data. *Int J Neuropsychopharmacol* 2011;14:405-12.
- 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.
- Montgomery SA. Antidepressant or antidepressant plus placebo effect? *World Psychiatry* 2015;14:303-4.
- Kasper S, Dold M. Factors contributing to the increasing placebo response in antidepressant trials. *World Psychiatry* 2015;14:304-6.
- Khan A, Brown WA. Antidepressants vs. placebo in major depression: an overview. *World Psychiatry* 2015;14:294-300.

34. Moore TJ, Mattison DR. Adult utilization of psychiatric drugs and differences by sex, age, and race. *JAMA Intern Med* 2017;177:274-5.
35. 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-analysis. *J Clin Psychiatry* 2010;71:270-9.
36. 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;19:34-40.
37. Rutherford BR, Roose SP. A model of placebo response in antidepressant clinical trials. *Am J Psychiatry* 2013;170:723-33.
38. Food and Drug Administration. www.accessdata.fda.gov.
39. Turner EH, Matthews AM, Linardatos E et al. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008;358:252-60.
40. Lee K, Bacchetti P, Sim I. Publication of clinical trials supporting successful new drug applications: a literature analysis. *PLoS Med* 2008;5:1348-56.
41. Gibertini M, Nations K, Whitaker J. Obtained effect size as function of sample size in approved antidepressants: a real-world illustration in support of better trial design. *Int Clin Psychopharmacol* 2012;27:100-6.
42. Turner EH, Matthews AM, Linardatos E et al. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008;358:252-60.
43. Mahableshwarkar A, Jacobsen P, Serenko M et al. A randomized, double-blind, placebo-controlled study of the efficacy and safety of 2 doses of vortioxetine in adults with major depressive disorder. *J Clin Psychiatry* 2015;76:583-91.
44. Liu KS, Snively DB, Ball WA et al. Is bigger better for depression trials? *J Psychiatr Res* 2008;42:622-30.

DOI:10.1002/wps.20421