Determinants of antidepressant response: implications for practice and future clinical trials

Jonathan Rabinowitz, PhD[a], Nomi Werbeloff, PhD[b], Francine S. Mandel, PhD[c], Lauren Marangell, MD[d], François Menard, MD[e], Shitij Kapur, MD[f]

(a) Bar Ilan University, Ramat Gan, Israel; (b) University College London, London (c) Pfizer (at time work conducted); (d) Eli Lilly (at time work conducted); (e) Lundbeck SAS (at time work conducted); (f) University of Melbourne, Australia

Corresponding author: Jonathan Rabinowitz, PhD, Bar Ilan University, Ramat Gan, Israel; Mailing address: 22 Akiva Street, Suite 5, Raanana, Israel, jonathan.rabinowitz@biu.ac.il; Phone: +972-54-464-3889

Abstract

Background: Response to antidepressants in major depressive disorder is variable and determinants are not well understood or used to design clinical trials. We aimed to understand these determinants. *Methods*: Supported by Innovative Medicines Initiative, as part of a large public-private collaboration (NEWMEDS), we assembled the largest dataset of individual patient level information from industry sponsored randomized placebo-controlled trials of antidepressant drugs in adults with MDD. We examined patient and trial-design-related determinants of outcome as measured by change on Hamilton Depression Scale or Montgomery–Asberg Depression Rating Scale in 34 placebo-controlled trials (drug, n=8260; placebo, n=3957). *Results*: While it is conventional for trials to be 6-8 weeks long, drug-placebo differences were nearly the same at week 4 as at week 6 and with lower dropout rates. At the multivariate level, having any of these attributes was significantly associated with greater drug vs. placebo differences on symptom improvement: female, increasing proportion of patients on placebo, centers located outside of North America, centers with low placebo response (regardless of active treatment response) and using randomized withdrawal designs. *Limitations*- Data on compounds that failed were not available to us. Findings may not be relevant for new mechanisms of action. *Conclusions*- Proof of concept trials can be shorter and efficiency improved by selecting enriched populations based on clinical and demographic variables, ensuring adequate balance of placebo patients, and carefully selecting and monitoring centers. In addition to improving drug discovery, patient exposure to placebo and experimental treatments can be reduced.

Introduction

Antidepressants were first discovered in the 1950's and in the late 1980's serotonin reuptake inhibitors (SSRIs) were introduced following a large number of double-blind randomized placebo-controlled trials with different compounds. Most of these SSRI trials were six to eight weeks in duration without stratification. They all included adult patients with major depressive disorder, regardless of symptom profile – despite evidence that any or all of the following factors may affect clinical response; age, sex (Kornstein and McEnany, 2000); (Khan et al., 2005), geographic region (Khin et al., 2011) (See studies in Supplementary table). In addition, the literature suggests that trials could be shorter (Rutherford et al., 2009; Tedeschini et al., 2011) and that removing centers with unrealistically high or low placebo response (blinded to active treatment response) (Mallinckrodt and Prucka, 2010; Merlo-Pich et al., 2010) could heighten placebo-active treatment differences. However, these measures to improve trials have not been adequately tested to confidently include them in clinical trials. To complicate matters further, nearly half the patients dropped out of these trials (Rutherford et al., 2013) and there are international differences in study results according to European Medicines Agency (2009) raising methodological questions. These findings, along with a fair number of negative or failed trials (where an established drug fails to separate from placebo) (Khin et al., 2012), the moderate superiority over placebo (Leucht et al., 2009) and the rising cost and difficulties of completing these trials, have led to questions whether the current approach to trials is most effective – and whether more focused and

shorter trials might yield informative results especially in early phases of clinical development.

To address the above mentioned impediments to drug development, the National Advisory Mental Health Council (2010) has recommended sharing of data to improve efficiency and decrease cost of therapeutic development. This could enable identifying moderators and mediators of treatment effects, and facilitate establishing a biologically-based discovery process. In concert with this, as part of the European Union funded Innovative Medicines Initiative, an academic and industry collaboration, we merged individual patient data from 34 randomized controlled trials (RCT) from four pharmaceutical companies. We explored determinants of antidepressant response in major depressive disorder, optimal trial duration, and whether these findings could be used to design more efficient trials in general, and specifically proof of concept trials.

We examined which key demographic and clinical variables, as well as study design features, influenced response, and if so, in what way. Next, as treatment response may be reached earlier than six weeks, we tested if study conclusions could have been reached earlier. Finally, based on previous literature on the inflation of baseline scores stemming from enrollment pressures (e.g., DeBrota et al., 1999; Kobak et al., 2010), we examined whether patients who just met symptom inclusion criteria were overrepresented and whether this appeared to affect study results. We speculated that if there was an overrepresentation of patients just meeting inclusion criteria, this may suggest that scores may have been inflated for purposes of including them in the study (the so-called

"baseline inflation"). We then examined whether their exclusion might have resulted in different conclusions.

Method

The NEWMEDS repository includes anonymized patient data from controlled studies to treat depression from the 39 randomized placebo-controlled trials (n=12,217) (1983-2007) of citalopram, duloxetine, escitalopram, quetiapine and sertraline. This included all the acute placebo-controlled trials of major depressive disorder in non-enriched (e.g., no major psychiatric comorbidities) adult populations, sponsored or owned by Pfizer, Eli Lilly, AstraZeneca and Lundbeck. We examined patient and trial-design-related determinants of outcome as measured by change on the Hamilton Depression Scale or the Montgomery–Asberg Depression Rating Scale in 34 placebo-controlled trials (drug, n=8260; placebo, n=3957). Eight, out of 22 active-placebo studies, were negative studies, and 5/17 studies with active comparators were failed studies (no difference on study drug and active comparator vs. placebo). Five of 39 studies were relapse prevention studies with open label randomized withdrawal designs prior to randomization of responders.

Results of the individual studies (listed in Supplementary Table 1) have been publicized. These data have not been previously pooled into a single dataset. All drugs were grouped and compared to placebo. Each study had been approved by the relevant IRB when and where it was conducted. All studies included informed written consent of study participants. The first and second authors of this paper had full access to all the

data in the studies, conducted all of the statistical analyses and take responsibility for the integrity of the data and the accuracy of the data analysis. There was no commercial funding for this work.

Measures

Studies used the MADRS (19 studies) or HAM-D (34 studies) and 14 studies used both. For combined analysis, we estimated the HAM-D based on the MADRS using equipercentile scaling (relative rank order within each measure). For randomized withdrawal designs, double-blind period baseline was used for change from baseline calculations.

Completeness of data

Complete data was available for all 12,217 subjects on sex, trial identifier, year of study and study arm. Data was missing on age for 15 subjects, on region for 434 subjects and site identifier for 6329 subjects. For purposes of analysis the 434 subjects with missing region were included in the "other" region group and age for the missing 15 subjects was replaced with mean age. Site was only included in one of set of analyses.

Analysis plan

Differential effects of key variables available at baseline on drug vs. placebo response were examined primarily based on the literature using a pre-specified analytic plan. The individual participant data from all studies were modelled simultaneously while accounting for the clustering of participants within studies as per the one-step approach to individual participant data meta-analysis, as described by Riley, Lambert &

Abo-Zaid (Riley et al., 2010). Specifically, we conducted a multi-level Mixed Model Repeated Measures (MMRM) analyses with subjects nested within studies, controlling for baseline using scaled identity matrix. Patient level fixed effects studied were age (quartile), sex and treatment (drug vs. placebo). Study level-fixed effects were: investigated drug, region (Western Europe, Eastern Europe, North America), proportion of patients on placebo (25% or less, 26% to 35%, greater than 35%), design (standard vs. withdrawal), outcome measure (HAM-D or MADRS) and year of study. Adjusted marginal means, F test, degrees of freedom, p value and Cohen's d effect size score are reported. Effects of study year were examined by testing for linear effects in placebo drug difference using the MMRM estimated marginal means. Because site was not available for 52% of the subjects, site was not included as a level in the main analysis but studied in a second round of analysis directed at studying the effects of sites with vs. without unrealistically high placebo response.

A separate round of analysis was done of baseline inflation as only some studies had symptom level inclusion criteria. Given difficulties in recruiting, patients who are just below the eligibility threshold may have had their scores unintentionally inflated so that they may be included. These patients would be expected to show a more pronounced improvement early on in treatment (with both drug and placebo), and by increasing overall response in the placebo group may mitigate against finding a true difference. Specifically, twenty-eight studies had lower symptom level inclusion criteria (not including randomized withdrawal designs). These studies have screening data on 8990 patients. Of these, 395 patients (4.4%) had screening scores below the bottom inclusion

criteria. These patients were removed from further analyses. To examine baseline inflation, subjects were grouped based on 5-point grouping of their baseline score from the bottom symptoms inclusion criteria. A potential baseline inflation was defined as patients with screening scores within 5 points of the bottom inclusion criteria (4,175 patients (48.6%) met this criteria and were the largest group (the next adjacent 5 point groups: 35.6%, 13.1%, 2.1% and 0.3%).

In addition, we tested to see whether shorter trials might be feasible. We examined percent of six-week difference between drug and placebo already discernible at each previous week. For example, if week 6 total difference between drug and placebo was five points and the week 5 difference was four points, then 80% (4/5) of this difference was discernible at week 5. Since in most trials a difference is considered to be statistically significant at a p-value of <0.05, we examined if a drug-placebo difference which met this criteria at week six would also have met this criteria had the trial been stopped earlier (e.g., at three, four, or five weeks). All analyses were conducted using SPSS Version 23 (IBM, Chicago, USA).

Results

Supplementary Table 1 includes details on study arms, regions and type of dosing for all 39 randomized placebo-controlled trials from the NewMeds repository. These include 4 citalopram trials, 13 duloxetine trials, 4 escitalopram trials, 4 quetiapine trials and 14 sertraline trials. Of the 12,217 included in the repository, 64.9% were females. Mean age at study entry was 45.2 ± 15.1 years.

Table 1 presents the results of multilevel analysis. Of the *patient level effects* sex was significantly associated with placebo vs. active treatment difference in response. Females had less placebo response than males and more active treatment response. Neither age nor symptom score at baseline being near bottom inclusion criteria (i.e., possible baseline inflation) were significantly associated with placebo vs. active treatment difference in response.

At the *site level* sites with unrealistically high placebo response (40% or more at week four) had markedly less placebo vs. active treatment difference than sites that did not have unrealistically high placebo response. Table 2 shows a breakdown of the 19 studies that included information on study sites, in 3 there were no centers with inflated placebo change scores. In all but 1 there was improvement after removing the centers with inflated change scores (see Table 2).

At the *study-level* region, proportion of patients on placebo, and randomized withdrawal design were significantly associated with increased placebo-active separation, whereas fixed vs. flexible dosing and study measure (MADRS vs. HAM-D) was not. Specifically, for each successive increase in proportion of subjects on placebo there was greater placebo vs. active separation. While there were differences in placebo-active separation between years, the differences were not linear, and in fact there was a linear increase in response in both the treatment and placebo arms. There was the least difference between placebo vs. active treatment in North America as compared to Europe.

Drug-placebo differences and trial duration. The percent of the total drug-placebo difference at 6 weeks that was discernible at earlier weeks was as follows: week 5: 87%; week 4: 79%; week 3: 60%; week 2: 45%; week 1: 13%. Active treatment was significantly superior to placebo at week 6 in 15 trials. At week 4 all these trials still show significant differences, as do two additional trials that were not significant at week 6. Additionally, completion rates were higher at 4 weeks than at 6 weeks (85.3% vs. 75.8%, respectively), and this was evident in all trials.

Discussion

Based on this unprecedented private-public collaboration that enabled merging data from the majority of placebo-controlled studies of SSRIs conducted by four pharmaceutical companies over the last two decades, we were able to identify response determinants that could help improve efficiency of future drug discovery trials in this area. We found that such trials can be shorter and placebo vs. active treatment differences increased by increasing proportion of patients on placebo, excluding centers with high placebo response, using randomized withdrawal designs and also possibly by increasing proportion of females, although this may not be practical as they constituted 65% of subjects. With regard to baseline inflation, while patients scoring within 5 points of the bottom inclusion criteria were the largest group, baseline inflation appears to be minor in these studies. However, where it exists it is adversely affects signal detection as it represents a type of unreliability.

While males and females showed almost identical treatment response, the placebo response was notably lower among women. These findings are in contrast to previous

findings suggesting that women respond better to SSRIs (Kornstein and McEnany, 2000); (Khan et al., 2005), and that there are no sex differences in the response to placebo (Casper et al., 2001; Kornstein and McEnany, 2000; Quitkin et al., 2002).

Better outcomes in some continents is in keeping with the European Medicines Evaluation Agency (EMA) suggestion that that geographic differences in outcomes may be related to intrinsic (genetic, physiological, and pathological conditions) and extrinsic (environmental, e.g., climate, culture, medical practice) factors (2009). Relating to depression trials, the differences may be related to the fact that patients in the United States may have participated in more trials and, thus, had exposure to more medications, thus lowering response compared to patients in Europe. These findings differ from those of Khin et al. (2011) who reported that although the observed placebo and drug responses at non-US sites tended to be larger than at US sites, the treatment effect was similar. Future studies should include an inventory of patients' experience in previous clinical trials and detailed medication history.

Similar to others (Borges et al., 2014), our results support the use of randomized withdrawal designs in increasing signal detection. Our results also suggest that MADRS and HAMD yield similar results. This is probably because these scales fundamentally capture similar domains of MDD. Additionally, excluding all patients from sites that had 40% or greater placebo response resulted in improvement in the efficacy of all studies examined. These results appear to differ from those of Gelwicks et al (2002) who did not find consistency in site performance on drug-placebo separation in *consecutive* studies of duloxetine. Similar to our findings are those of Targum et al (2014) who removed sites

with unrealistic placebo response using the band-pass approach (Merlo-Pich et al., 2010), post-hoc, *with-in* a study and found that it improved results. In keeping with the suggestion of Mallinckrodt and Prucka (2010) that the approach be tested prospectively, removing sites with unrealistic placebo response could be incorporated into the study design in two ways. The first is by having a pre-specified sensitivity analysis in the statistical analysis plan of re-analyzing data after removing sites, or possibly even raters, with unrealistic placebo response. Another way that this can be used is by planning a study with an interim analysis in which subjects are not recruited until the interim analysis is completed and the placebo data from the interim analysis is used to identify sites, or possibly raters, with high placebo response. Those centers or raters would be excluded from recruiting additional subjects for the rest of the study. This could result in improving signal detection.

The apparent overrepresentation of patients just meeting symptom eligibility criteria supports concerns that investigators may inflate scores to allow including additional patients. This finding previously reported in studies of anxiety (Williams et al., 2015) and depression (DeBrota et al., 1999). In those studies, baseline scores obtained in real-time through two different rating methodologies were compared, concluding that clinician-rated severity may be inflated so that patients reach the inclusion threshold. Taken together, these results suggest that in cases where there is evidence of baseline inflation, it would be prudent to include in the statistical analysis plan a secondary analysis after removing persons just meeting eligibility criteria.

Our results show that trials can both be shorter and have fewer patients. Based on the effect size for antidepressants, which we found to be in these studies d=.35, a conventional 6 week trial based on this effect size requires approximately 290 patients per arm (a total of 3480 patient weeks of exposure to both placebo and drug). Using the information identified herein, trial duration could be reduced to four weeks for 173 patients per arm (a total of 1384 weeks of patient exposure) as effect sizes could be increased by selecting patient groups showing increased treatment response. In addition to having fewer weeks of exposure, shorter trials have the advantage of higher completion rates, as shown by our data. In addition, trial recruitment for shorter trials will probably be easier, exposure to placebo more acceptable and more ethically justifiable, and retention rates should be even higher than shown here, as patients may be willing to stay in a shorter study with the end in sight. Shorter trials also cost less money and lower dropout rates result in less imputation of missing data.

We found that in the years of our studies 1983-2007 both treatment effect and placebo response increased over time, with a similar trend observed for the association between study year and drug-placebo difference. This is similar to Khan et al (2017) based on Food and Drug Administration (FDA) reviews for sixteen antidepressants (85 trials, 115 trial arms, 23,109 patients) approved between 1987 and 2013, who found that the magnitude of placebo response and active treatment response steadily rose in the past 30 years. This is in contrast to the findings of Khin et al. (2011) who examined 81 studies conducted in a similar time period and found that the placebo response showed a modest

increase over the observation period but the treatment effect clearly diminished, resulting in decreasing drug-placebo separation over time.

Our finding that treatment response increased with the proportion of subjects on placebo is consistent with a meta-analysis of depression studies (Papakostas and Fava, 2009), an analysis of a patient registry of antipsychotic trials (Mallinckrodt et al., 2011), a review of trials across psychiatry (Weimer et al., 2015) and has been reported in other areas of medicine as well (Enck et al., 2011). This finding has been attributed to expectancy; if the proportion of subjects on placebo is low then the expectation of both subject and investigator is that a given subject is on active treatment (Enck et al., 2011).

This study has several important limitations. Data is representative of clinical trials of medications that are proven superior to placebo on antidepressant effect; however, data on compounds that failed in the last two decades were not available to us. This leads to underestimation of the placebo response, and an overestimation of the drug-placebo difference. In addition, we were not able to study other important variables such as duration of the current episode, number of previous failed treatments, comorbidities, outpatient or inpatient status as these were not available in the data sets.

Our analysis was conducted on data from placebo controlled trials of citalopram, duloxetine, escitalopram, quetiapine and sertraline. The compounds on which our conclusions are based, with the exception of quetiapine, like all compounds currently available for clinical use, share the serotonin reuptake mechanism as their common mechanism. While this provides the only data-driven estimate for future drugs, it is conceivable that newer drugs working on different mechanisms may show a different

profile or timeframe of response. The results of this work may not be generalizable to compounds not included in this work and we were not able to test the results by compound, however the results replicated in most studies. All medications were given orally so that data cannot be generalized to other formulations. Future work should attempt to replicate these findings using data from compounds with different mechanisms of action.

The focus of our work is on efficacy as measured by the HAM-D or MADRS, and our objective was to test the possibility of conducting shorter proof of principle trials. Time to discontinuation is an important pragmatic outcome measure, which reflects both safety and tolerability, but, is more relevant in longer trials and thus was not a measure of interest for this paper. Our data suggest that including more women may be good for statistical power. While our analysis suggests enrollment criteria that maximize drug-placebo differences, using selective criteria (e.g., age groups or symptom severity) may decrease generalizability of results to routine clinical practice and make recruitment more difficult as it is likely to slow enrollment and delay trial conduct. However, in the early stages of drug development, where finding evidence of efficacy is more critical than generalizability, our data suggest a way forward.

In summary, sex, proportion of patients on placebo, centers or raters with high placebo response, randomized withdrawal designs and geographical location all significantly influence outcome of depression trials with effect sizes that are clinically relevant. Proof of concept trials can be shorter and efficiency improved by selecting enriched populations based on clinical and demographic variables, ensuring adequate

balance of placebo patients, carefully selecting and monitoring raters and centers and using randomized withdrawal designs. In addition to improving drug discovery, patient exposure to placebo and experimental treatments can be reduced.

Supplementary Table 1 List of studies included

Table 1. Baseline to endpoint (MMRM) change on the HAMD for placebo and drug difference by key variables (LS Mean, 95% CI) (a)

	Predicted mean (se) decline last observation		Multivariate test
	Placebo n=3957	Drug n=8260	Interaction Drug vs. Placebo Difference
Sex			df=1, 55810.614, f=20.853, p<.001, Cohen's d=.09
Female (n=7933)	3.65 (.157) n=2595	4.46 (.121) n=5338	
Male (n=4284)	3.72 (.166) n=1362	3.85 (.128) n=2922	
Age quartile			df=3, 51497.108, f=0.781, p<.50
Q 1 <33	3.22 (.21) n=952	2.77 (.17) n=2079	
Q 2 33-44	3.55 (.20) n=1023	3.06 (.17) n=2225	
Q 3 45-53	3.04 (.17) n=894	3.04 (.17) n=1899	
Q 4 >53	3.99 (.20) n=1082	3.54 (0.17) n=2067	
Region			df=3, 22989.683, f=6.53, p<.001, d=.07
Western Europe	2.61 (.16) n=905	4.02 (.164) n=1661	
Eastern Europe	3.28 (.23) n=390	2.77 (0.194) n=1114	
North America	0.89 (.161) n=2369	1.80 (0.127) n=4921	
Other & missing	7.03 (.33) n=293	6.75 (.251) n=564	
Proportion subjects on placebo			df=2, 46186.057, f=5.21, p=.003, d=.06
Up to 25%	2.97 (.269) n=841	3.21 (.209) n=3065	
26% to 34%	3.92 (.182) n=1256	4.16 (0.133) n=2814	
35% to 50%	4.16 (.133) n=1860	5.10 (.121) n=2381	
Fixed vs. flexible dose			df=1, 55003.320, f=0.29, p<.86
Fixed	2.94 (0.176) n=2405	3.36 (.13) n=5733	
Flexible	4.43 (.176) n=1552	4.96 (.148) n=2527	
Study year (linear test (c)	r=.48	r=.49	r=.007, p<.80
Randomized withdrawal			df=1, 21607.274, f=9.50, p=.002, d=.06
No	8.63 (.153) n=3328	9.72 (.13) n=7510	
Yes	-1.26 (.280) n=629	-1.41 (.201) n=750	
Measure			df=1, 52010.315, f=0.52, p=.47
HAMD	3.33 (.21) n=3335	3.69 (0.178) n=7278	
MADRS (only)	4.03 (.17) n=622	4.62 (.128) n=982	
Sites with high placebo response (d)			df=1, 8828.986, f=60.72, p<.001, d=.34
Yes	11.97 (.407) n=109	9.27 (.292) n=232	
No	3.14 (.346) n=636	4.19 (.254) n=1561	
Baseline inflation (e)			df=1, 22110.355, f=1.24, p=.26, d=.02
Yes	8.00 (.412) n=107	10.12 (.149) n=288	
No	7.99 (.286) n=2683	9.96.097) n=5912	

(a) Mixed Model Repeated Measures: subjects nested within studies, controlling for baseline. Patient level fixed effects: age, sex, drug vs. placebo; Study level-fixed effects: investigated drug, region, proportion of patients on placebo, design (standard vs. withdrawal), study measure year of study. (b) Spearman rank order correlation study year and adjusted mean from MMRM. (c) Excluding studies without site variable; (d) Excluding studies with no baseline symptom inclusion level.

Table 2. Inflated placebo response by site

Study	All sites included			number of centers removed	Sites with high placebo response (>=40%) removed			Difference in d
	n active	n placebo	Cohen's d=		n active	n placebo	Cohen's d=	
1	94	26	-0.325	3/10	72	20	0.075	0.4
2	254	86	0.224	2/8	238	79	0.273	0.049
3	299	150	0.473	1/8	240	121	0.5	0.027
4	211	104	0.247	1/5	167	82	0.272	0.025
5	225	110	0.182	1/5	158	77	0.302	0.12
6	159	40	-0.02	2/4	82	22	0.221	0.241
7	88	42	-0.058	2/6	66	31	0.101	0.159
8	137	34	-0.057	4/13	99	25	0.091	0.148
9	146	73	0.208	2/13	130	62	0.311	0.103
10	89	91	0.435	1/3	55	57	0.387	-0.048
11	134	66	0.159	12/18	31	12	0.612	0.453
12	185	89	0.163	4/18	143	71	0.319	0.156
13	154	147	0.256	4/38	145	142	0.321	0.065
14	257	117	0.04	5/58	235	107	0.167	0.127
15	268	142	-0.218	12/60	230	118	0.004	0.222
16	468	51	0.627	4/14	350	32	1.086	0.459

Note: Negative ES means that placebo did better than active treatment.

Acknowledgment: The research leading to these results has received support from the Innovative Medicine Initiative Joint Undertaking under grant agreement n° 115008 of which resources are composed of European Federation of Pharmaceutical Industries and Associations (EFPIA) in-kind contribution and financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013). Funding source was not involved in the collection, analysis, and interpretation of data; in the writing of the report; and nor in the decision to submit the paper for publication.

References

2009. EMEA reflection paper on the extrapolation of results from clinical studies conducted outside Europe to the EU-population, in: EMEA/CHMP/EWP/692702 (Ed.). European Medicines Agency, London.

Borges, S., Chen, Y.F., Laughren, T.P., Temple, R., Patel, H.D., David, P.A., Mathis, M., Unger, E., Yang, P., Khin, N.A., 2014. Review of maintenance trials for major depressive disorder: a 25-year perspective from the US Food and Drug Administration. J Clin Psychiatry 75, 205-214.

Casper, R.C., Tollefson, G.D., Nilsson, M.E., 2001. No gender differences in placebo responses of patients with major depressive disorder. Biol Psychiatry 49, 158-160.

DeBrota, D.J., Demitrack, M.A., Landin, R., Kobak, K.A., Greist, J.H., Potter, W.Z., 1999. A comparison between interactive voice response system–administered HAM-D and clinician-administered HAM-D in patients with major depressive episode, 39th Annual Meeting, New Clinical Drug Evaluation Unit, Boca Raton, FL.

Enck, P., Klosterhalfen, S., Weimer, K., Horing, B., Zipfel, S., 2011. The placebo response in clinical trials: more questions than answers. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 366, 1889-1895.

Gelwicks, S., Debrota, D., Engelhardt, N., Potter, W., 2002. Analysis of site performance across multiple clinical trials, 42nd annual meeting of the National Clinical Drug Evaluation Unit (NCDEU).

Khan, A., Brodhead, A.E., Schwartz, K.A., Kolts, R.L., Brown, W.A., 2005. Sex differences in antidepressant response in recent antidepressant clinical trials. J Clin Psychopharmacol 25, 318-324.

Khan, A., Fahl Mar, K., Faucett, J., Khan Schilling, S., Brown, W.A., 2017. Has the rising placebo response impacted antidepressant clinical trial outcome? Data from the US Food and Drug Administration 1987-2013. World Psychiatry 16, 181-192.

Khin, N.A., Chen, Y.F., Yang, Y., Yang, P., Laughren, T.P., 2011. 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 72, 464-472.

Khin, N.A., Chen, Y.F., Yang, Y., Yang, P., Laughren, T.P., 2012. Exploratory analyses of efficacy data from schizophrenia trials in support of new drug applications submitted to the US Food and Drug Administration. The Journal of clinical psychiatry 73, 856-864.

Kornstein, S.G., McEnany, G., 2000. Enhancing pharmacologic effects in the treatment of depression in women. J Clin Psychiatry 61 Suppl 11, 18-27.

Leucht, S., Arbter, D., Engel, R.R., Kissling, W., Davis, J.M., 2009. How effective are second-generation antipsychotic drugs? A meta-analysis of placebo-controlled trials. Mol Psychiatry 14, 429-447.

Mallinckrodt, C.H., Prucka, W.R., 2010. The enrichment window approach as a means of dealing with placebo response in antidepressant clinical trials. Clin Pharmacol Ther 88, 592-594.

Mallinckrodt, C.H., Tamura, R.N., Tanaka, Y., 2011. Recent developments in improving signal detection and reducing placebo response in psychiatric clinical trials. J Psychiatr Res 45, 1202-1207.

Merlo-Pich, E., Alexander, R.C., Fava, M., Gomeni, R., 2010. A new population-enrichment strategy to improve efficiency of placebo-controlled clinical trials of antidepressant drugs. Clin Pharmacol Ther 88, 634-642.

National_Advisory_Mental_Health_Council's_Workgroup, 2010. From Discovery to Cure: Accelerating the Development of New and Personalized Interventions for Mental Illnesses, in: National_Advisory_Mental_Health_Council (Ed.). National Institutes of Health Washington, DC.

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

Quitkin, F.M., Stewart, J.W., McGrath, P.J., Taylor, B.P., Tisminetzky, M.S., Petkova, E., Chen, Y., Ma, G., Klein, D.F., 2002. Are there differences between women's and men's antidepressant responses? Am J Psychiatry 159, 1848-1854.

Riley, R.D., Lambert, P.C., Abo-Zaid, G., 2010. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ 340, c221.

Rutherford, B.R., Cooper, T.M., Persaud, A., Brown, P.J., Sneed, J.R., Roose, S.P., 2013. Less is more in antidepressant clinical trials: a meta-analysis of the effect of visit frequency on treatment response and dropout. J Clin Psychiatry 74, 703-715.

Rutherford, B.R., Sneed, J.R., Roose, S.P., 2009. Does study design influence outcome?. The effects of placebo control and treatment duration in antidepressant trials. Psychother Psychosom 78, 172-181.

Targum, S.D., Burch, D.J., Asgharnejad, M., Petersen, T., Gomeni, R., Fava, M., 2014. Use of band-pass filter analysis to evaluate outcomes in an antidepressant trial for treatment resistant patients. Eur Neuropsychopharmacol 24, 1188-1195.

Tedeschini, E., Fava, M., Papakostas, G.I., 2011. Placebo-controlled, antidepressant clinical trials cannot be shortened to less than 4 weeks' duration: a pooled analysis of randomized clinical trials employing a diagnostic odds ratio-based approach. J. Clin. Psychiatry. 72, 98-113.

Weimer, K., Colloca, L., Enck, P., 2015. Placebo effects in psychiatry: mediators and moderators. Lancet Psychiatry 2, 246-257.

Williams, J.B., Kobak, K.A., Giller, E., Reasner, D.S., Curry, L., Detke, M.J., 2015. Comparison of Site-Based Versus Central Ratings in a Study of Generalized Anxiety Disorder. J Clin Psychopharmacol 35, 654-660.