

ORIGINAL ARTICLE

Determining maximal achievable effect sizes of antidepressant therapies in placebo-controlled trials

Fredrik Hieronymus^{1,2,3}  | Alexander Lisinski³ | Magnus Hieronymus⁴ | Jakob Näslund³ | Elias Eriksson³  | Søren Dinesen Østergaard^{1,2}

¹Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

²Department of Affective Disorders, Aarhus University Hospital, Aarhus, Denmark

³Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, Sweden

⁴Swedish Meteorological and Hydrological Institute, Norrköping, Sweden

Correspondence

Fredrik Hieronymus, Department of Clinical Medicine, Aarhus University, Palle Juul-Jensens Boulevard 99, Aarhus N 8200, Denmark.
Email: fredrik.hieronymus@neuro.gu.se

Funding Information

FH is supported by The Lundbeck Foundation (grant number: R303-2018-3429). SDØ is supported by the Novo Nordisk Foundation (grant number: NNF20SA0062874), the Lundbeck Foundation (grant numbers: R358-2020-2341 and R344-2020-1073), The Danish Cancer Society (R283-A16461), the Central Denmark Region Fund for Strengthening of Health Science (grant number: 1-36-72-4-20), The Danish Agency for Digitisation Investment Fund for New Technologies (grant number 2020-6720) and Independent Research Fund Denmark (grant number: 7016-00048B). These funding bodies were not involved in the study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

Abstract

Objective: Antidepressants outperform placebo with an effect size of around 0.30. It has been suggested that effect sizes as high as 0.875 are necessary for a minimal clinically important difference. Whether such effect sizes are achievable in placebo-controlled trials is unknown. Therefore, we aimed to assess what effect sizes are theoretically achievable in placebo-controlled trials of antidepressants.

Methods: Patient-level analyses comparing Hamilton Depression Rating Scale (HDRS-17) outcomes for simulated antidepressant therapies to placebo-treated participants ($n = 2201$) from clinical trials of selective serotonin reuptake inhibitors.

Results: An optimally effective antidepressant, where all treated participants achieve HDRS-17 scores comparable to those displayed by healthy volunteers (remission-type model), had a maximum effect size of 1.75, with a mean difference of 11.6 points on the HDRS-17. In simulations where patients received an additional 50% symptom reduction over that obtained with placebo (improvement-type model), the maximum effect size was 1.08 with a mean HDRS-17 difference of 7.2. When adjusting for normal rates of treatment discontinuation, maximum effect sizes were 1.10 (remission-type model) and 0.76 (improvement-type model) with HDRS-17 mean differences of 8.8 and 5.6, respectively.

Conclusions: Three methodological issues (i) a large and variable placebo response, (ii) a high rate of dropout and (iii) HDRS-17-ratings significantly larger than zero in healthy volunteers, reduce the degree of treatment-placebo separation achievable in depression trials. Assuming that those who discontinue treatment have only partial response, even a highly effective antidepressant would have difficulties surpassing such effect size cut-offs as have been suggested to signify a minimal clinically important difference.

KEYWORDS

antidepressive agents, depression, minimal clinically important difference, placebo effect, serotonin uptake inhibitors

1 | INTRODUCTION

The magnitude of drug-placebo differences for antidepressant therapies has been extensively debated.¹ Some authors have argued that the statistically significant effects observed in clinical trials—an effect size of about 0.30 or an HDRS-17 difference of 2 points—are far below commonly championed cut-offs for clinical relevance, and that antidepressant therapies should therefore not be used.²⁻⁵ In this vein, previous treatment guidelines from the National Institute of Health and Care Excellence (NICE) refrained from recommending a number of established treatments for major depression since their effect size (ES) as compared to placebo was below 0.50.⁶ Others have argued for even higher cut-offs, an analysis comparing scores on the Clinical Global Impression (CGI) and the 17-item Hamilton Depression Rating Scale (HDRS-17)⁷ suggested a CGI rating of minimally improved to correspond to an improvement of 7 points on the HDRS-17, or an ES of 0.875, hence promoting the perception that antidepressant therapies displaying treatment differences smaller than these lack clinical relevance.²⁻⁵ While the cut-offs described above originate from within-patient observations (ie the difference between the baseline and endpoint rating), they have been interpreted as implying that HDRS-17 differences between treatments smaller than these are clinically imperceptible.²⁻⁵ There are several problems with applying a metric derived from within-person change scores to between-treatment differences,¹ one issue that has received scarce attention is whether treatment-placebo differences of such magnitudes are at all achievable in antidepressant trials. Given (i) that placebo-treated participants often display mean endpoint scores of less than 15 HDRS-17 points, with standard deviations (SDs) around 7.5,⁷ (ii) that it is impossible to score less than zero, (iii) that healthy volunteers on average display HDRS-17 scores around 3 (SD 3.2)⁸ and (iv) that dropouts—who tend to display much higher average HDRS-17 scores than healthy volunteers or remitted depressed individuals—are often in the 30–40% range, it is unclear if between-treatment differences of 7 HDRS-17 points (or an ES of 0.875) are realistic targets.

The purpose of the present study was to determine the maximal effect size magnitude possible to obtain in placebo-controlled antidepressant trials. The analyses were based on HDRS-17 ratings from patients with major depression who have been treated with placebo in randomized controlled trials of selective serotonin reuptake inhibitors (SSRIs). Using the placebo population as baseline, we created a duplicate population for which we simulated treatment with effective antidepressant therapy. Effect sizes achieved under different assumptions of efficacy were then calculated by comparing simulated antidepressant treatment to real-world placebo outcomes. We also assessed how treatment non-completion and time to maximum effect impacted achievable effect sizes.

Significant outcomes

- In our simulations, a theoretically optimal slow-acting antidepressant, where everyone treated reach depression levels on par with healthy volunteers (remission-type model), would display an effect size of 1.75. An antidepressant where instead of everyone treated improves an additional 50% over what they would have improved on placebo (improvement-type model) would have an effect size of 1.08 at endpoint.
- When adjusting for normal rates of dropout and partial efficacy in treatment non-completers, maximum achievable effect sizes were 1.10 for a remission-type model and 0.76 for an improvement-type model, at endpoint.
- Even under liberal assumptions, for example that treatment non-completers have partial efficacy and that active treatment can never lead to a worse outcome than placebo treatment, even a ‘perfect’ antidepressant (with a 100% underlying rate of remission or improvement) would have difficulties surpassing some of the cut-offs suggested to signify a minimal clinically relevant effect in depression (eg an effect size of 0.875).

Limitations

- Simulated effect sizes are based on the short-term placebo response to selective serotonin reuptake inhibitors and do not necessarily translate to rapid-acting antidepressant treatments or to other treatment modalities.
- The results are based on outcomes from the Hamilton Depression Rating Scale and need not generalize to other depression outcome measures or to measures of other relevant constructs (eg quality of life or functioning).
- Future treatments with higher efficacy may have better retention rates even if their side effect burden is similar. Such treatments would demonstrate higher efficacy in simulations adjusted for dropout as per our methodology.

1.1 | Aims of the study

To provide an upper limit on the maximum achievable effect size for a theoretically perfect antidepressant and to assess what effect sizes can be expected in placebo-controlled trials of slow-acting antidepressants relative to their true underlying efficacy.

2 | METHODS

This is a simulation study based on outcome data from actual placebo-treated subjects who have participated in acute-phase trials of antidepressants for major depression. Using the placebo data as the control group, we investigate maximum achievable effect sizes for simulated antidepressant treatments with varying rates of efficacy.

2.1 | Data acquisition

The data set consisted of all placebo-treated patients who participated in acute-phase placebo-controlled trials in adult major depression during the development programs for citalopram, paroxetine and sertraline. These data are available to us for offline use due to data sharing agreements with the respective manufacturer. Data can be requested from these manufacturers and/or from clinical data sharing websites like ClinicalStudyDataRequest.com and Vivli.org. Participants receiving active treatment were included only to inform rates of treatment non-completion. Trials had to utilize the HDRS-17 scale and have a scheduled evaluation at week 6 to be eligible for inclusion. Additional details on the study population have been presented previously.⁹

2.2 | Analyses

The last available observation up until the week 6 evaluation for all placebo-treated participants was included. This population served both as the control group (for the effect of placebo) and as baseline for simulated antidepressant treatment. For each simulation, the placebo population was duplicated and the effects of simulated antidepressant treatment modelled onto the duplicated population.

We first assessed an improvement-type effect of antidepressant treatment where patients undergoing simulated treatment receive on average a 50% improvement over the endpoint score of their placebo counterparts. Specifically, endpoint HDRS-17 scores for patients undergoing simulated treatment were the endpoint scores of the corresponding patients in the placebo distribution multiplied with an improvement factor. These were modelled using a normal distribution with a mean of 0.5 and a standard deviation of 0.1. To preclude changes too small or too large to fit with the clinical perception of response, improvement factors were bounded at 0.2 and 0.8. We considered models where 0%, 20%, 40%, 60%, 80% and 100% of participants were responsive to treatment, defined as mentioned above. In the first run, all patients were modelled as achieving this improvement regardless of time under treatment as long as the patient had at least one post-baseline HDRS-17 evaluation (since otherwise

improvement is impossible to detect). Patients who only had a baseline observation retained the score of their placebo duplicate. In a second run, we accounted for dropouts by modelling improvement as linearly related to time under treatment. A participant that was randomized to obtain a 60% decrease in endpoint score (an improvement factor of 0.4) had they completed 6 weeks of treatment would thus have their score decreased by 10% if their last visit was at week 1, by 20% if their last visit was at week 2, etc. Dropout rates for simulated treatment were taken from the active treatment arms of the included trials.

In the second simulation, we assumed a remission-type effect of the simulated antidepressant treatment. Patients undergoing simulated treatment were thus given a new endpoint score drawn from a distribution approximating the HDRS-17 scores that can be expected in healthy volunteers. The score distribution for remitted patients was generated using an exponential distribution with a rate parameter, 0.3125, chosen to correspond to a mean and SD of 3.2 in healthy volunteers as reported by Zimmermann and co-workers.⁸ Since some patients receiving placebo displayed low HDRS-17 scores at endpoint, we introduced an additional criterion according to which simulated remitters whose new endpoint score was higher than the endpoint score of their placebo duplicate retained the placebo endpoint score (ie so that simulated treatment can never increase endpoint scores and thereby make it more difficult to achieve treatment-placebo separation). Using the same strategy as for the simulation of improvement-type mechanisms, we first simulated runs where all patients with at least one post-baseline evaluation remitted, defined as mentioned above, and then considered models in which chance of remission was linearly related to time under treatment. For example, in a simulation where 60% of participants undergoing simulated treatment obtained remission, a subject dropping out after week 1 would have a 10% chance of remission, a patient dropping out after week 2 would have a 20% chance of remission, etc.

2.3 | Sensitivity analyses

In order to explore the influence of time under treatment and dropout rates, we ran simulations using placebo HDRS-17 scores obtained from the week one evaluation and from the week 12 evaluation (in trials with such duration), respectively. For week one data, only dropouts prior to any post-baseline visit are relevant and there were thus no analyses assessing the impact of dropouts. For week 12 data, we first ran these using dropout figures from the full population, with maximum effect being achieved if the patient received treatment for ≥ 6 weeks. This was done to isolate the effect of lower endpoint scores in the placebo group in longer duration

trials so that the impact of this factor could be assessed. To assess the impact of an antidepressant effect that was slower to develop than in the primary analyses, we ran the same analyses using dropout figures from 12-week trials, with the treatment effect being linearly related to time and a full effect only being achieved if the patient remained in the study for ≥ 12 weeks.

All analyses were conducted in R version 3.6.3.

3 | RESULTS

Characteristics of the included studies are displayed in Table 1. Observations were available for 7170 participants, 2201 of which had been allocated to placebo. The overall dropout rate for SSRI-treated participants was 31.9%, with the last available evaluation being the baseline evaluation for 4.6% (230) of cases, the week 1 evaluation for 8.1% (400), the week 2 evaluation for 5.7% (285), the week 3 evaluation for 4.9% (243), the week 4 evaluation for 8.4% (416) and the week 5 evaluation for 0.2% (9). The overall dropout rate was not significantly different ($p = .483$) from that of placebo-treated participants (32.7%).

3.1 | Improvement- and remission-type mechanisms

The mean placebo endpoint HDRS-17 rating for trials with a week 6 evaluation was 15.5 with a standard deviation of 7.73. Figure 1A-F detail the results of simulating an improvement-type antidepressant effect (ie a 50% additional reduction over that obtained with placebo) with full effect of treatment being achieved as long as the individual has at least one post-baseline visit. As shown in Figure 1E-F, a rate of improvement above 80% would be needed to surpass an ES of 0.875 and a rate close to 100% would be needed to reach a mean difference of 7 HDRS-17 points.

Figure 2A-F show the same improvement-type analysis but with full effect of treatment being achieved only if an individual stays in the study for six weeks. Under these assumptions, models where 100% of study completers experienced a 50% larger improvement than that obtained with placebo resulted in an ES of 0.76 and a mean difference of 5.55 (Figure 2F).

Figures 3 and 4 detail the results of the corresponding simulations using a remission-type antidepressant effect. In the case of no dropouts, 100% of subjects remitting resulted in an ES of 1.75 and a mean difference of 11.6 (Figure 3F), while a 60%

Study	Placebo (n)	Female (n)	Age (mean)	Trial duration	Trial conducted
GSK/002	171	93	42.5	6 weeks	1985–1987
GSK/003	244	122	40.4	6 weeks	1985–1986
GSK/009	53	29	40.6	6 weeks	1985–1986
GSK/115	117	85	41.5	12 weeks	1991
GSK/128	140	100	42.8	12 weeks	1991
GSK/251	129	84	41.3	8 weeks	1992–1993
GSK/276	21	12	43.4	6 weeks	1982–1984
GSK/279	10	7	51.6	6 weeks	1983–1986
GSK/448	103	67	38.4	12 weeks	1996–1997
GSK/449	110	66	40.7	12 weeks	1996–1997
GSK/487	109	69	69.4	12 weeks	1996–1997
GSK/810	148	91	38.6	8 weeks	2001–2002
LB/89303	66	45	43.7	6 weeks	1989–1990
LB/91206	129	71	37.8	6 weeks	1992–1993
PZ/103	91	40	37.3	6 weeks	1984–1985
PZ/104	150	78	40.1	8 weeks	1984–1986
PZ/109	104	69	41.0	8 weeks	1987–1989
PZ/111	110	72	41.4	8 weeks	1989–1990
PZ/238	40	28	37.2	10 weeks	1987–1989
PZ/240	35	18	68.6	10 weeks	1987–1990
PZ/247	42	22	68.6	10 weeks	1987–1990
PZ/315	79	61	43.2	8 weeks	1984–1986

TABLE 1 Study characteristics

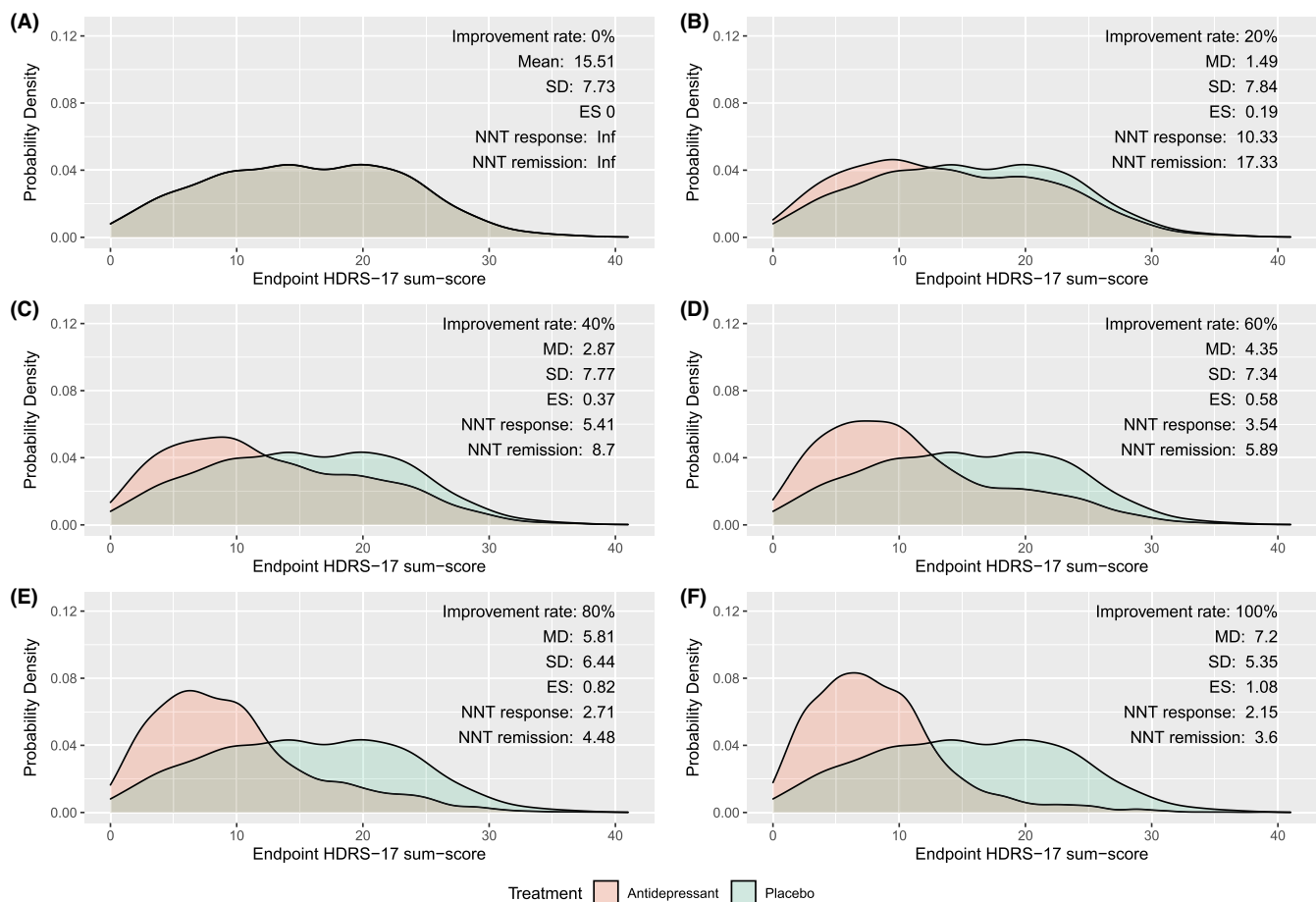


FIGURE 1 Improvement-type mechanism, full efficacy in treatment non-completers. Distribution of HDRS-17 scores at endpoint after six weeks of treatment in patients having received simulated antidepressant treatment (pink) or placebo-treatment (light blue), respectively, with a response rate of 0% (A), 20% (B), 40%, (C), 60% (D), 80% (E) and 100% (F). Simulated treatment effects are fully realized in all participants with at least one post-baseline evaluation. 2201 placebo endpoint scores were included in this analysis

remission rate yielded an ES of 0.85 and a mean difference of 6.96 (Figure 3D), that is an outcome close to what has been suggested as cut-off for a clinically relevant effect. If a full effect of treatment was only achieved in participants who did not dropout prior to week 6, a remission rate of 80% would be required to obtain an ES of 0.85 and a mean difference of 7.04 (Figure 4E).

3.2 | Sensitivity analyses

The mean HDRS-17 score at week one for placebo-treated participants was 19.8 with an SD of 5.53. Figures S1 and S2 show the results of simulating improvement- and remission-type effects fully realized at this timepoint. Due to the lower SD, an ES of 0.93 is achieved already at 60% rate of improvement (Figure S1D), whereas a mean difference of >7 is achieved first at an 80% rate of improvement. For remission-type mechanisms, an ES of 0.81 is reached at 40% remission (Figure S2C) with a mean difference of 6.26.

At week 12, the mean endpoint score in the placebo-treated group was 13.9 with an SD of 7.99. Figures S3 and

S4 show simulations of improvement- and remission-type mechanisms with full efficacy if the patient had an evaluation at or after week 6. Under these conditions, the improvement-type mechanism achieved a maximum ES of 0.69 with a mean difference of 5.06 (Figure S3F) and the remission-type mechanism maxed out at an ES of 0.98 with a mean difference of 7.65 (Figure S4F).

Figures S5 and S6 detail the results of simulations with an antidepressant effect that is slower to develop (reaching full strength after 12 weeks of treatment). Maximal ES for a improvement-type mechanism under these conditions was 0.59 and the maximum mean difference was 4.48 (Figure S5F). The corresponding figures for remission-type models yielded a maximum ES of 0.83 with a mean difference of 6.76 (Figure S6F).

4 | DISCUSSION

The primary finding of this study is that previously suggested criteria for what should be regarded as lower limit

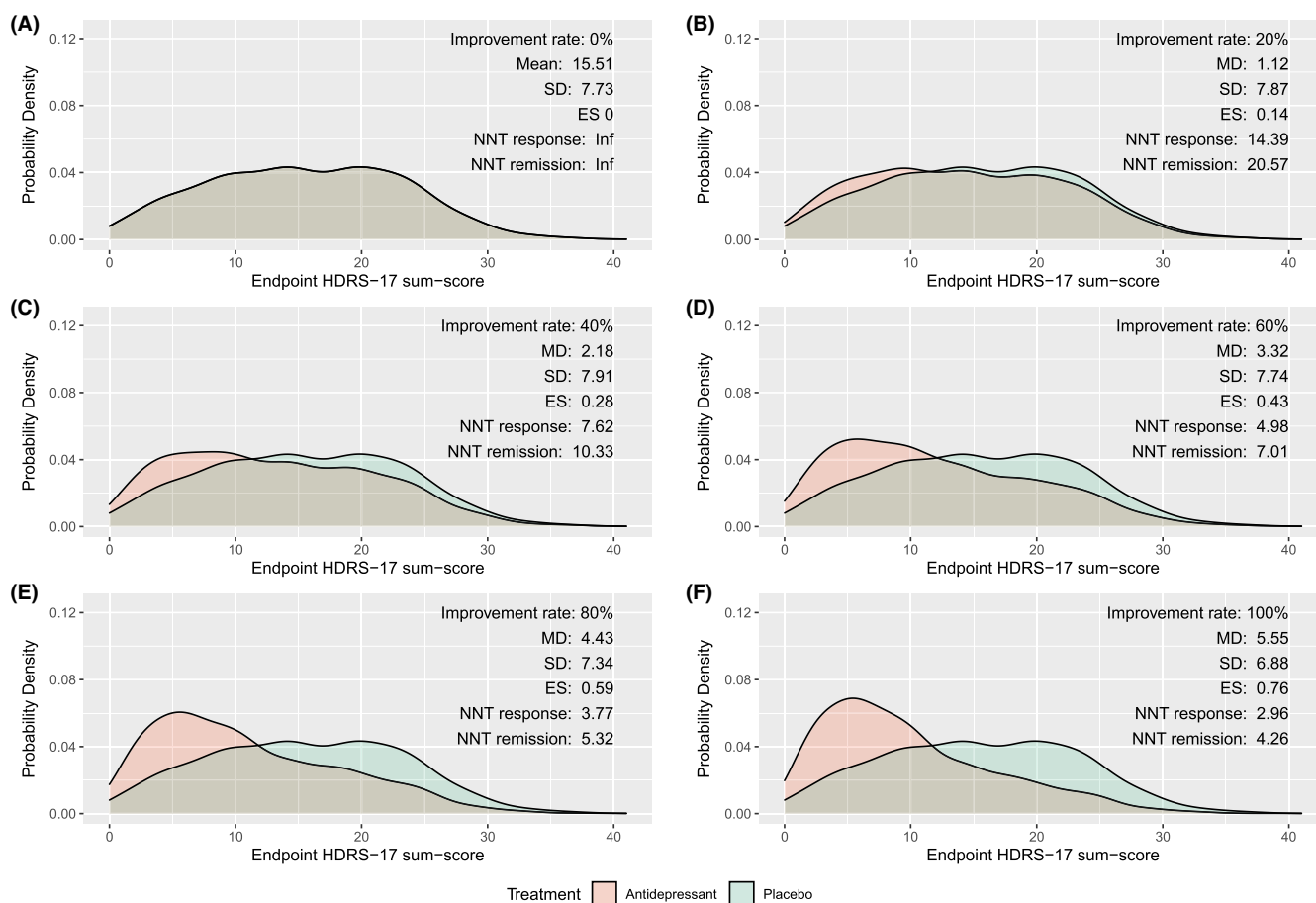


FIGURE 2 Improvement-type mechanism, partial efficacy in treatment non-completers. Distribution of HDRS-17 scores at endpoint after 6 weeks of treatment in patients having received simulated antidepressant treatment (pink) or placebo-treatment (light blue), respectively, with a response rate of 0% (A), 20% (B), 40%, (C), 60% (D), 80% (E) and 100% (F). Simulated treatment effects are linearly related to time and fully realized after 6 weeks of treatment. 2201 placebo endpoint scores were included in this analysis

for a difference between an antidepressant and placebo to be regarded clinically relevant, that is an ES of 0.875 or a mean difference of 7 HDRS points, are likely above what is possible to achieve with a slow-acting antidepressant therapy given the current design of antidepressant trials. Even under the unrealistic assumption of full efficacy in study non-completers, simulations thus show that 80% of subjects on active treatment must display an additional 50% improvement than if treated with placebo, or that 60% must display remission, for these cut-offs to be surpassed more often than not. Given the dropout rate normally seen in antidepressant trials, and assuming the effect of active treatment to be only partial in non-completers, the maximum achievable ES is 0.76 for improvement-type simulations and 1.10 for remission-type models.

The assumption of full efficacy also in treatment non-completers, which was used in some simulations, is obviously unrealistic for slow-acting antidepressant therapies. Similarly, simulating treatment effects in dropouts as being linearly related to time is likely overly optimistic as there is very little

separation on the HDRS-17 sum-score during the first two weeks of treatment with, for example an SSRI.¹⁰ While likely more realistic to model treatment effects as being, for example, quadratically related to time under treatment, these conservative design choices were made in order not to overstate the difficulties in achieving large ESs. For the same reason, we did not account for other factors which may make treatment-placebo separation difficult to achieve, including (i) that trial participants in remission may display substantially higher average HDRS-17 scores than healthy volunteers,^{8,11,12} (ii) that HDRS-17 captures symptoms which may be elicited as treatment-emergent side effects^{7,13} and (iii) that no participant responded worse to simulated antidepressant therapy than to placebo, which cannot be assumed to hold true in reality. Also in other regards, we applied a conservative approach; we hence did not consider placebo samples other than the intention to treat population and did not conduct analyses stratified by initial severity, since the hereby obtained subpopulations (ie completers and/or mildly depressed subjects) would have displayed lower endpoint scores than

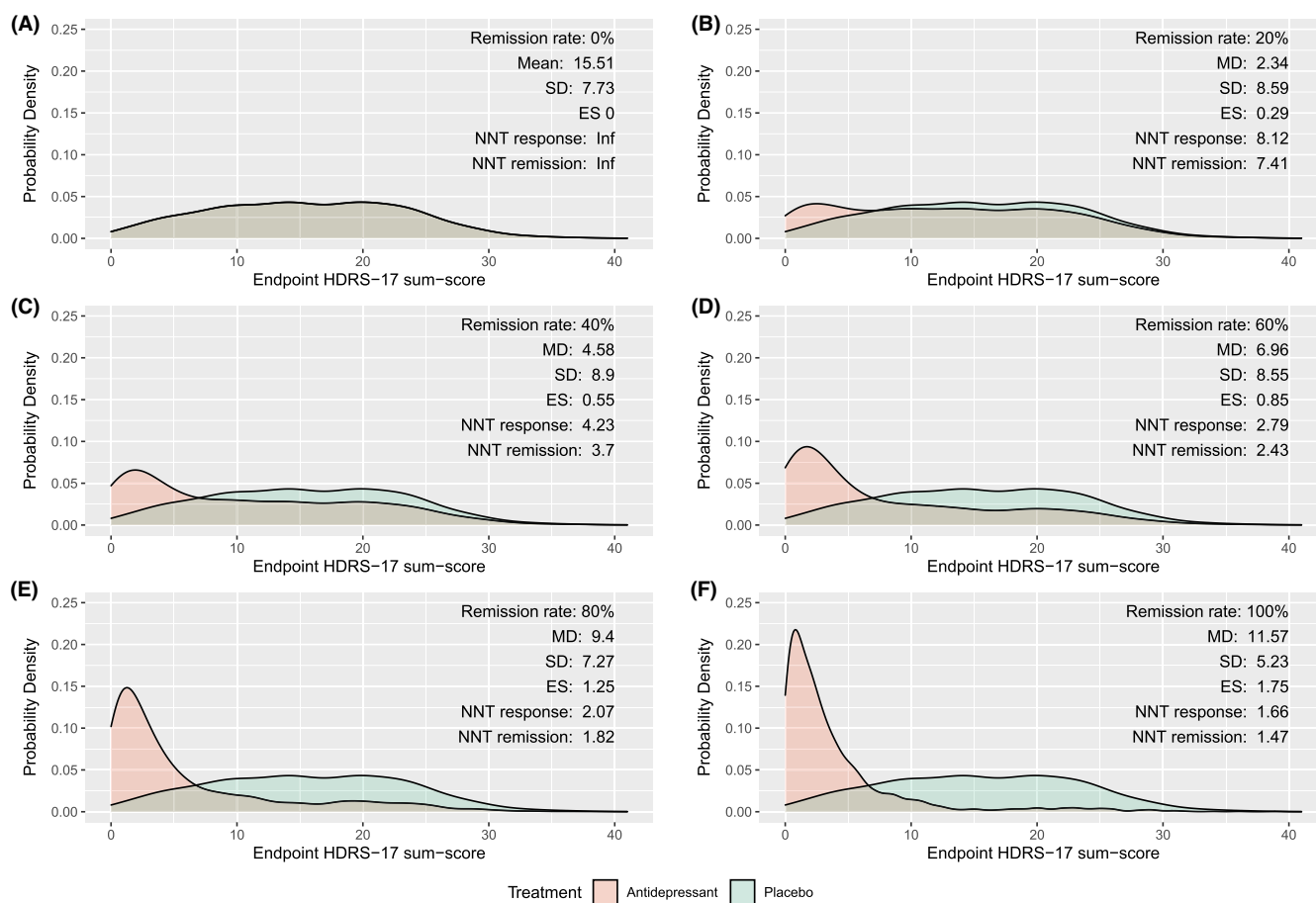


FIGURE 3 Remission-type mechanism, full efficacy in treatment non-completers. Distribution of HDRS-17 scores at endpoint after six weeks of treatment in patients having received simulated antidepressant treatment (pink) or placebo-treatment (light blue), respectively, with a remission rate of 0% (A), 20% (B), 40% (C), 60% (D), 80% (E) and 100% (F). Simulated treatment effects are fully realized in all participants with at least one post-baseline evaluation. 2201 placebo endpoint scores were included in this analysis

the full population, thus making it more difficult to achieve treatment-placebo separation under the simulated scenarios.

It might be argued that our results reveal that most depressive episodes will be significantly improved by 6 weeks also when treated with placebo and that antidepressant interventions are thus of limited value, at least in the acute phase. Such an interpretation is, however, overreliant on mean scores, neglecting that treatment outcomes vary widely also among placebo-treated patients. While a significant fraction of these display very low depression scores at endpoint, there is also a sizeable fraction who do not respond. Reducing the duration and severity of depressive episodes for these patients, as well as decreasing the risk of relapse, should remain a priority.¹⁴⁻¹⁶

In this context, it should also be considered that few other treatments in medicine would be able to surpass the effect size cut-offs sometimes suggested adequate for depression.^{17,18} That antidepressants in fact seem to perform at or above average in reducing core depressive symptoms is noteworthy,^{7,17} especially given that trials in other fields often should be less marred by issues such as poor compliance,

inclusion of misdiagnosed subjects and inexact outcome measures, which are all factors that are inclined to reduce the apparent effect of active treatment in depression trials (but not addressed in the present analyses).

We also assessed the relative importance of the two major factors contributing to the difficulty in achieving a high degree of treatment-placebo separation in depression trials, that is the large and variable symptom reduction in subjects receiving placebo¹⁹ (which can be the result of, for example spontaneous remission, regression towards the mean or an actual placebo effect), and the high rate of dropouts.^{20,21} To this end, the impact of the placebo response was inferred by contrasting different time points while holding the influence of dropouts constant. When contrasting simulations where the full effect of treatment was at hand in all patients who had at least one post-baseline observation, maximum ES for improvement-type simulations changed from 1.84 using week 1 data to 1.08 using week 6 data (Figure 1F, Figure S1F), the corresponding figures being 2.90 and 1.75 for remission-type simulations (Figure 3F, Figure S2F). Similar patterns were seen when contrasting simulations utilizing outcome data at

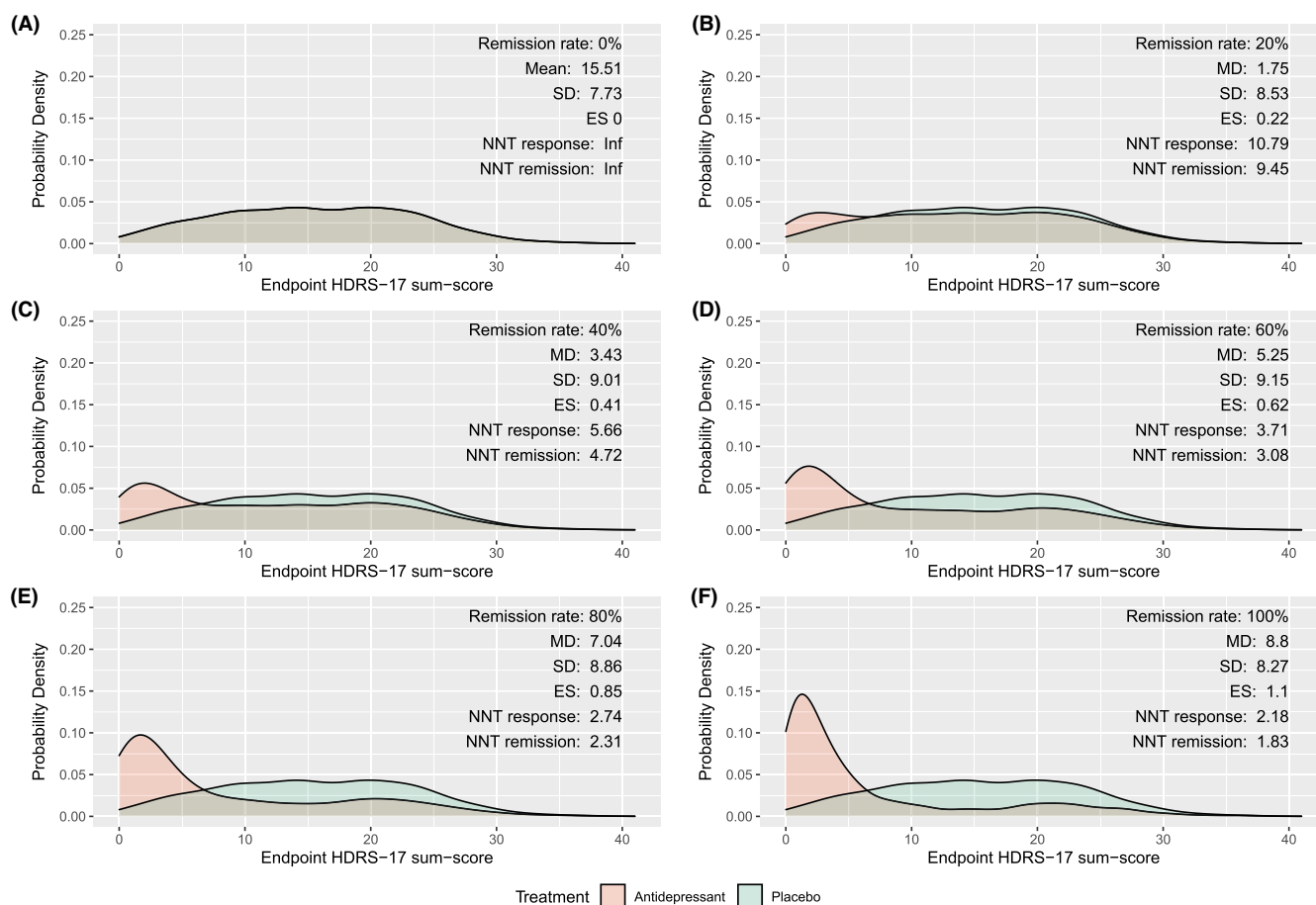


FIGURE 4 Remission-type mechanism, partial efficacy in treatment non-completers. Distribution of HDRS-17 scores at endpoint after six weeks of treatment in patients having received simulated antidepressant treatment (pink) or placebo-treatment (light blue), respectively, with a remission rate of 0% (A), 20% (B), 40% (C), 60% (D), 80% (E) and 100% (F). Simulated treatment effects are linearly related to time and fully realized after 6 weeks of treatment. 2201 placebo endpoint scores were included in this analysis

week 6 and week 12 (Figures 2 and 4, Figures S3 and S4). Needless to say, the major contributor to these differences is the mean placebo endpoint score being much higher at week 1 (19.8) than at week 6 (15.5), and slightly higher at week 6 than at week 12 (13.9); wherefore, there is more room for a treatment effect in earlier evaluations.

Similarly, the impact of dropout can be assessed by comparing models that handle treatment non-completion differently while holding placebo responses constant. Contrasting models where the size of the simulated treatment effect was linearly related to time under treatment to those where full efficacy was assumed to be achieved in all patients showed this constraint to lower the maximally achievable ES at week 6 from 1.08 (Figure 1F) to 0.76 (Figure 2F) for improvement-type models, the corresponding decrease for remission-type models being a change from 1.75 (Figure 3F) to 1.10 (Figure 4F). Similar patterns were seen when contrasting models where the full effect of treatment was realized only after 12 weeks of treatment to those where full efficacy was achieved after 6 weeks (Figures S3 and S6).

Though the debate on clinical relevance has primarily revolved around antidepressants, the difficulties in achieving meaningful separation between treatment arms should apply to all slow-acting therapies for depression (eg transcranial direct-current stimulation and psychotherapies)²²⁻²⁴ and be especially pronounced when one attempts to detect differences between two treatments both of which may display some efficacy.^{10,25}

A notable observation is that SDs were often unrelated to the proportion of treatment responders in improvement-type models although endpoint score distributions for patients receiving simulated treatment were clearly distinguishable from those of patients receiving placebo (eg Figure 1A & C and 2A & D). The observation that comparatively simple transformations can leave standard deviations largely unaffected is in line with a recent publication arguing that meta-analytical results of variability ratios not significantly different from 1 do not imply that there is no treatment effect heterogeneity.²⁶

This study has several limitations. First, it is a simulation study considering a limited range of underlying treatment mechanisms. While this should not impact the general

conclusion regarding the difficulty for slow-acting antidepressant therapies to achieve large treatment-placebo separation, the mechanisms simulated may differ from those in clinical reality; inferences from the undertaken simulations with regard to underlying rates of 'remitters' or 'improvers' should hence be undertaken with caution. Second, we focussed on HDRS-17 ratings since HDRS has been the most commonly used rating instrument for trials of pharmacological antidepressants—to what extent these findings are relevant for other instruments (eg the Montgomery-Åsberg Depression Rating Scale or the different HDRS-17 subscales), other relevant measures (eg functioning and quality of life), or other trial contexts, remains to be investigated. Third, the implication that large ESs are easier to achieve with rapid-acting antidepressants hinges on the assumption that placebo outcomes in short trials will be similar to those in longer trials which may not be the case.^{27,28} Fourth, an optimal slow-acting antidepressant treatment might have better retention rates than the therapies currently available, thereby making our adjustments for treatment non-completion overly pessimistic.

In summary, our results suggest that large effect sizes are practically unattainable when there is (i) a large and variable placebo response, (ii) a high rate of dropout and (iii) a rating-scale floor precluding arbitrarily large improvements. Our simulations suggest that the effect size cut-offs for antidepressant treatments championed by some authors are unlikely to be surpassed even by a theoretically near-perfect, slow-acting antidepressant treatment.^{29,30} We believe these results, in conjunction with the fact that treatment outcomes are highly variable, should prompt a reconsideration of the reliance on mean difference or effect size cut-offs to determine the clinical relevance of interventions in major depression.

CONFLICTS OF INTEREST

FH has received speaker's fees from H. Lundbeck. EE has previously been on advisory boards and/or received speaker's honoraria and/or research grants from Eli Lilly, Servier, GlaxoSmithKline and H Lundbeck. SDØ received the 2020 Lundbeck Foundation Young Investigator Prize. The remaining authors declare no conflicts of interest.

AUTHOR CONTRIBUTION

FH and SDØ conceived and designed the work. EE acquired the data. FH and AL have verified the underlying data and analyses. All authors participated in the analysis and interpretation of the data. FH drafted the manuscript. All authors revised the work for important intellectual content. All authors have given approval for this version to be submitted for publication.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/acps.13340>.

DATA AVAILABILITY STATEMENT

We kindly thank GlaxoSmithKline, Pfizer and H. Lundbeck for making patient-level data available for these analyses. Access to data can be requested from the respective companies and/or from ClinicalStudyDataRequest.com.

ORCID

Fredrik Hieronymus  <https://orcid.org/0000-0003-0930-6068>

Elias Eriksson  <https://orcid.org/0000-0002-4128-2046>

REFERENCES

1. Hieronymus F, Jauhar S, Ostergaard SD, Young AH. One (effect) size does not fit at all: Interpreting clinical significance and effect sizes in depression treatment trials. *J Psychopharmacol*. 2020;34:1074-1078.
2. Hengartner MP, Ploderl M. Statistically significant antidepressant-placebo differences on subjective symptom-rating scales do not prove that the drugs work: effect size and method bias matter!. *Front Psychiatry*. 2018;9:517.
3. Jakobsen JC, Gluud C, Kirsch I. Should antidepressants be used for major depressive disorder? *BMJ Evidence-Based Med*. 2019;25(4):130.
4. Jakobsen JC, Katakam KK, Schou A, et al. Selective serotonin reuptake inhibitors versus placebo in patients with major depressive disorder. A systematic review with meta-analysis and Trial Sequential Analysis. *BMC Psychiatry*. 2017;17:58.
5. Moncrieff J, Kirsch I. Empirically derived criteria cast doubt on the clinical significance of antidepressant-placebo differences. *Contemp Clin Trials*. 2015;43:60-62.
6. Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Medicine*. 2008;5:e45.
7. Hieronymus F, Emilsson JF, Nilsson S, Eriksson E. Consistent superiority of selective serotonin reuptake inhibitors over placebo in reducing depressed mood in patients with major depression. *Mol Psychiatry*. 2016;21:523-530.
8. Zimmerman M, Chelminski I, Posternak M. A review of studies of the Hamilton depression rating scale in healthy controls: implications for the definition of remission in treatment studies of depression. *J Nerv Ment Dis*. 2004;192:595-601.
9. Naslund J, Hieronymus F, Lisinski A, Nilsson S, Eriksson E. Effects of selective serotonin reuptake inhibitors on rating-scale-assessed suicidality in adults with depression. *Br J Psychiatry*. 2018;212:148-154.
10. Hieronymus F, Nilsson S, Eriksson E. A mega-analysis of fixed-dose trials reveals dose-dependency and a rapid onset of action for the antidepressant effect of three selective serotonin reuptake inhibitors. *Transl Psychiatry*. 2016;6:e834.
11. Zimmerman M, Martinez J, Attiullah N, Friedman M, Toba C, Boerescu DA. Why do some depressed outpatients who are not in remission according to the Hamilton depression rating scale nonetheless consider themselves to be in remission? *Depress Anxiety*. 2012;29:891-895.
12. Zimmerman M, Martinez JA, Attiullah N, et al. Why do some depressed outpatients who are in remission according to the Hamilton Depression Rating Scale not consider themselves to be in remission? *J Clin Psychiatry*. 2012;73:790-795.

13. Hieronymus F, Lisinski A, Nilsson S, Eriksson E. Influence of baseline severity on the effects of SSRIs in depression: an item-based, patient-level post-hoc analysis. *Lancet Psychiatry*. 2019;6:745-752.
14. Steinert C, Hofmann M, Kruse J, Leichsenring F. Relapse rates after psychotherapy for depression – stable long-term effects? A meta-analysis. *J Affect Disord*. 2014;168:107-118.
15. Geddes JR, Carney SM, Davies C, et al. Relapse prevention with antidepressant drug treatment in depressive disorders: a systematic review. *Lancet*. 2003;361:653-661.
16. Kato M, Hori H, Inoue T, et al. Discontinuation of antidepressants after remission with antidepressant medication in major depressive disorder: a systematic review and meta-analysis. *Mol Psychiatry*. 2021;26:118-133.
17. Leucht S, Hierl S, Kissling W, Dold M, Davis JM. Putting the efficacy of psychiatric and general medicine medication into perspective: review of meta-analyses. *Br J Psychiatry*. 2012;200:97-106.
18. Fountoulakis KN, Moller HJ. Are antidepressants clinically useful? Conclusion of a decade of debate. *World Psychiatry*. 2014;13:201-202.
19. 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;118:1-8.
20. Salanti G, Chaimani A, Furukawa TA, et al. Impact of placebo arms on outcomes in antidepressant trials: systematic review and meta-regression analysis. *Int J Epidemiol*. 2018;47:1454-1464.
21. Cooper AA, Conklin LR. Dropout from individual psychotherapy for major depression: A meta-analysis of randomized clinical trials. *Clin Psychol Rev*. 2015;40:57-65.
22. Schefft C, Guhn A, Brakemeier EL, Sterzer P, Köhler S. Efficacy of inpatient psychotherapy for major depressive disorder: a meta-analysis of controlled trials. *Acta Psychiatr Scand*. 2019;139:322-335.
23. Moffa AH, Martin D, Alonzo A, et al. Efficacy and acceptability of transcranial direct current stimulation (tDCS) for major depressive disorder: An individual patient data meta-analysis. *Prog Neuropsychopharmacol Biol Psychiatry*. 2020;99:109836.
24. Cuijpers P, Karyotaki E, Weitz E, Andersson G, Hollon SD, van Straten A. The effects of psychotherapies for major depression in adults on remission, recovery and improvement: A meta-analysis. *J Affect Disord*. 2014;159:118-126.
25. Steinert C, Munder T, Rabung S, Hoyer J, Leichsenring F. Psychodynamic therapy: as efficacious as other empirically supported treatments? A meta-analysis testing equivalence of outcomes. *Am J Psychiatry*. 2017;174:943-953.
26. Volkmann C, Volkmann A, Muller CA. On the treatment effect heterogeneity of antidepressants in major depression: A Bayesian meta-analysis and simulation study. *PLoS One*. 2020;15:e0241497.
27. Daly EJ, Singh JB, Fedgchin M, et al. Efficacy and safety of intranasal esketamine adjunctive to oral antidepressant therapy in treatment-resistant depression. *JAMA Psychiatry*. 2018;75:139.
28. Ionescu DF, Fu D-J, Qiu X, et al. Esketamine nasal spray for rapid reduction of depressive symptoms in patients with major depressive disorder who have active suicide ideation with intent: results of a phase 3, double-blind, Randomized Study (ASPIRE II). *Int J Neuropsychopharmacol*. 2021;24:22-31.
29. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet*. 2018;391:1357-1366.
30. Cuijpers P, Turner EH, Mohr DC, et al. Comparison of psychotherapies for adult depression to pill placebo control groups: a meta-analysis. *Psychol Med*. 2014;44:685-695.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Hieronymus F, Lisinski A, Hieronymus M, Näslund J, Eriksson E, Østergaard SD. Determining maximal achievable effect sizes of antidepressant therapies in placebo-controlled trials. *Acta Psychiatr Scand*. 2021;00:1–10. <https://doi.org/10.1111/acps.13340>