



This is a repository copy of *Randomization and placebo effects in clinical trials of major depressive disorder*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/223121/>

Version: Published Version

Article:

Rogers, J.A. orcid.org/0009-0004-2363-1339 and Senn, S. orcid.org/0000-0002-7558-8473 (2025) Randomization and placebo effects in clinical trials of major depressive disorder. *Translational Psychiatry*, 15 (1). 43. ISSN 2158-3188

<https://doi.org/10.1038/s41398-025-03263-0>

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

CORRESPONDENCE OPEN



Randomization and placebo effects in clinical trials of major depressive disorder

© The Author(s) 2025

Translational Psychiatry (2025)15:43; <https://doi.org/10.1038/s41398-025-03263-0>

Research programs in major depressive disorder (MDD) are in some cases shaped by perceived problems arising from the heterogeneity and non-additivity of placebo effects. A recent proposal in this direction has been provided by Gomeni, Hopkins, Bressole-Gomeni, and Fava [1]. In this correspondence we wish to challenge several of the premises that motivate and guide their proposal.

In their opening paragraph, Gomeni, Hopkins, Bressole-Gomeni, and Fava [1] suggest that the validity of statistical inference based on a randomized controlled trial (RCT) depends on covariate balance between the treatment groups. In fact, covariate balance is neither expected nor required for the outcome of any particular randomized allocation in a RCT. Correspondingly, it is not the case that conventional RCT analyses assume, “that the response is only driven by the treatment administered” [1] (p. 1). Rather, it is *expected* that treatment effect estimates from an RCT will be biased (albeit in an unknown direction) *conditional on a given randomization*. Valid confidence intervals for a treatment effect do not require a false pretense that this conditional bias has been eliminated, but instead account for uncertainty in the magnitude and direction of the conditional bias by means of increased interval width relative to the width that would be appropriate in the absence of (unknown) conditional biases [2, 3]. As such, RCT designs are valid in MDD even if placebo effects are large, heterogeneous, and non-additive with treatment effects.

Moreover, the statistical evidence for placebo effect heterogeneity and non-additivity is not particularly strong. Gomeni, Hopkins, Bressole-Gomeni, and Fava emphasize the possibility that treatment effects and placebo effects are non-additive, such that a placebo patient experiencing an improvement relative to baseline would have been unlikely to experience a further incremental improvement if assigned the active treatment. This may be a plausible hypothesis, but it is not supported by the observation that, “a meta-analysis ... showed that a higher placebo response rate statistically significantly correlates with a low risk-ratio of responding to antidepressant versus placebo” [1] (p. 1). On the contrary, a negative correlation between the estimate of the average treatment effect $TE = (TR - PR)$ and the estimate of PR is *always* expected when TR and PR are estimated from independent samples [4, 5]. Since this negative correlation is expected even when the true treatment effect is perfectly additive with the placebo response, the cited meta-analysis result should not be interpreted as evidence that “the level of placebo response has a critical prognostic relevance in the assessment of treatment effect” [1] (p. 1). A detailed analysis of correlations of this nature in MDD trials has been performed by Whitlock, Woodward, and Alexander [5], who concluded, “that the treatment and placebo effects observed in MDD trials are highly correlated, to the degree

expected under the assumption of placebo additivity” (p. 17, emphasis ours), suggesting therefore that, “the recent focus on designing trials that reduce placebo response and/or attempt to remove high placebo responders could be ineffective”.

Concerns with placebo effects are complicated by the challenge of defining placebo effects as empirically estimable quantities. While the placebo *response* is directly observable for any patient randomized to placebo, the placebo *effect* necessarily involves a comparison of potential outcomes, one being observed and the other being counterfactual (this is true even when the response is expressed as a change from baseline: a longitudinal change is not necessarily interpretable as an “effect”). The definitions of placebo response and placebo effect implied by, Gomeni, Hopkins, Bressole-Gomeni, and Fava appear to involve some equivocation: in the authors’ description of their 5 step approach, an artificial neural network (ANN) is developed to identify subjects with “placebo response” in steps 2 and 3, and then employed in step 4 to identify subjects with an “individual probability to have a PE [placebo effect]” [1] (p. 2, emphases ours). When placebo response is defined by dichotomization of an underlying continuous measure, placebo responders are likely to consist especially of patients whose baseline scores are just below the threshold for dichotomization, with “response” occurring as a result of residual variation near the boundary. There is no reason to suppose that placebo *response*, defined in this way, is a promising proxy for placebo *effect*: patients initially near a boundary for dichotomization (the “placebo responders”) would not necessarily be more susceptible to non-specific stimuli (“placebo effect”) and in general would not necessarily have differential treatment effects, as illustrated in Fig. 1.

While the evidential basis for placebo effect heterogeneity and non-additivity is somewhat weak and beset by conceptual challenges, it is not unreasonable to pursue speculative hypotheses that suppose the existence of such effect structures. There are, however, substantial caveats when such hypotheses are both developed and evaluated using the same data set. Relatedly, it is not the case, as claimed by Gomeni, Hopkins, Bressole-Gomeni, and Fava, that their methodology is consistent with the intention-to-treat (ITT) principle (“Two ITT analyses were conducted: ... the second analysis was the propensity weighted analysis” [1], p. 3). The inclusion of all randomized subjects is neither necessary nor sufficient to conform to the ITT principle, which “asserts that the effect of a treatment policy can be best assessed by evaluating on the basis of the intention to treat a subject (i.e. the planned treatment regimen) rather than the actual treatment given” [6]. The proposed methodology’s downweighting of some subjects on the basis of post-randomization data is in fact at odds with the design intention of treating those down-weighted subjects. Relatedly, the more recent ICH E9 R1 articulation of treatment policy estimands emphasizes the importance of clearly identifying the population that one intends to treat *prior to the analysis* [7].

The use of response data from a single dataset to both define and apply a weighting scheme also gives rise to concerns with

Received: 13 March 2024 Revised: 4 January 2025 Accepted: 27 January 2025
Published online: 05 February 2025

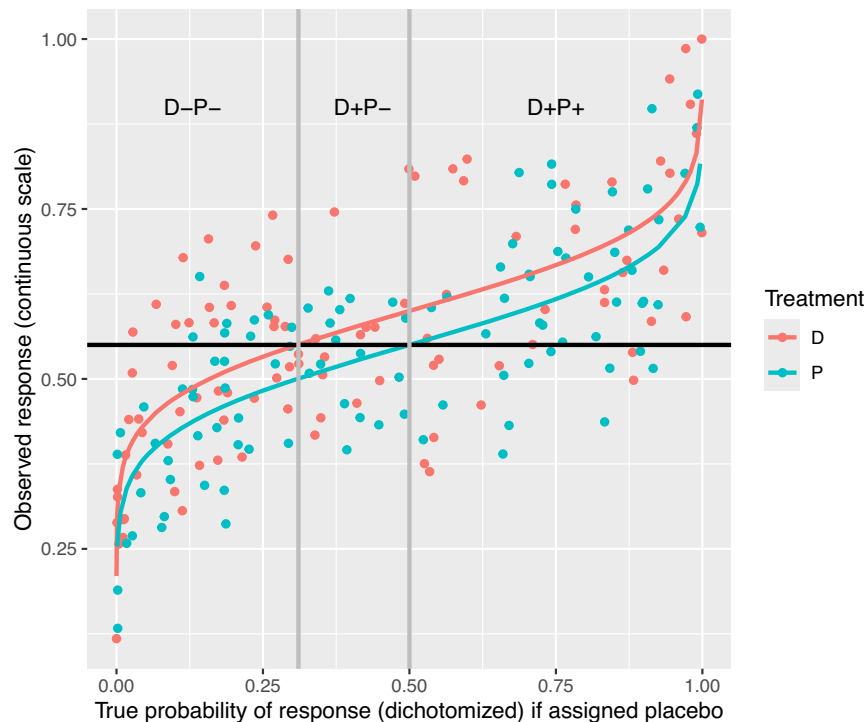


Fig. 1 Data simulated with variable placebo response and additive (constant) treatment effect. Treatment effect and measurement error were simulated on a zero-to-one scale, using a cutoff for response dichotomization of 0.55 (depicted by the horizontal black line) and with parameter values selected to achieve response rates similar to those reported by Gomeni, Hopkins, Bressole-Gomeni, and Fava [1] (~33% “D–P–”, ~25% “D+P–”, ~42% “D+P+”, delineated graphically by vertical grey lines). The probability of placebo response is known exactly in the simulation and increases from left to right, but the magnitude of the treatment effect is identical for all patients, corresponding to the fixed vertical distance between the two sigmoidal lines. As such, the additive nature of the simulated treatment effect is consistent with the meta-analytic findings of Whitlock, Woodward, and Alexander [5]. The figure therefore illustrates that, even if the probability of placebo response could be computed exactly on the basis of baseline and screening data, it would convey no value for trial enrichment. (This situation is, of course, unchanged when the probability of placebo response is instead estimated by an artificial neural net.) Moreover, the figure illustrates why the “D+P–” designation in itself is a misleading artifact of dichotomization and does not constitute a meaningful target for trial enrichment, since it identifies patients whose scores are near the boundary for dichotomization rather than patients with greater magnitudes of treatment effect. R code to generate this plot and the simulated data underlying it is provided as supplemental material.

Type I error control. In this regard, it is important to recognize that use of the term “propensity” in Gomeni, Hopkins, Bressole-Gomeni, and Fava diverges from standard usage. In Rosenbaum and Rubin’s original 1983 publication, the term “propensity score” was used to refer to “the conditional probability of assignment to a particular treatment given a vector of observed covariates” [8] (p. 41, emphasis ours). Standard propensity weights are therefore not a function of observed responses. However, as used in Gomeni, Hopkins, Bressole-Gomeni, and Fava, “propensity” refers not to the probability of treatment assignment, but to the probability of being a placebo responder. This non-standard usage is consequential because it renders irrelevant the investigations of Type I error control under standard propensity weighting (e.g., the investigations Turley, Redden, Case, Katholi, Szychowski, and DuBay [9], which are cited by Gomeni, Hopkins, Bressole-Gomeni, and Fava). Given this specialized meaning of “propensity” in Gomeni, Hopkins, Bressole-Gomeni, and Fava, there is no apparent reason to believe that their proposed methodology controls the Type I error rate at any reasonable level. A convincing demonstration that Type I error rate is controlled would appear to require extensive simulation.

Valid lines of inquiry may exist that are predicated on the heterogeneity and non-additivity of placebo effects in MDD. However, researchers wishing to pursue these lines of inquiry should be aware that the evidence for such hypotheses is equivocal, and that simultaneous hypothesis generation and hypothesis evaluation is likely to convey a significant risk of false positive findings.

James A. Rogers ¹✉ and Stephen Senn ^{2,3}

¹Metrum Research Group, 30 Tower Lane, Suite #140, Avon, CT 06001-4231, USA. ²Sheffield Centre for Health and Related Research (SCHARR) The University of Sheffield, Sheffield, UK. ³Institute of Medical Statistics, Center for Medical Data Science, CeDAS Medical University of Vienna, Vienna, Austria. ✉email: jimr@metrumrg.com

REFERENCES

- Gomeni R, Hopkins S, Bressole-Gomeni F, Fava M. Interpreting clinical trial outcomes complicated by placebo response with an assessment of false-negative and true-negative clinical trials in depression using propensity-weighting. *Transl Psychiatry*. 2023;13:388.
- Senn S. A brief note regarding randomization. *Perspect Biol Med*. 2013;56:452–3.
- Senn S. Seven myths of randomisation in clinical trials. *Stat Med*. 2013;32:1439–50.
- Senn S. Importance of trends in the interpretation of an overall odds ratio in the meta-analysis of clinical trials. *Stat Med*. 1994;13:293–6.
- Whitlock ME, Woodward PW, Alexander RC. Is high placebo response really a problem in depression trials? a critical re-analysis of depression studies. *Innov Clin Neurosci*. 2019;16:12–17.
- Center for Drug Evaluation & Research. E9 statistical principles for clinical trials. *U.S. Food and Drug Administration* <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e9-statistical-principles-clinical-trials> (2020).
- Center for Drug Evaluation & Research. E9(R1) statistical principles for clinical trials: addendum: estimands and sensitivity analysis in clinical trials. *U.S. Food and Drug Administration* <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e9r1-statistical-principles-clinical-trials-addendum-estimands-and-sensitivity-analysis-clinical> (2021).
- Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41–55.

9. Turley FC, Redden D, Case JL, Katholi C, Szychowski J, DuBay D. Comparison of Type I error rates and statistical power of different propensity score methods. *J Stat Comput Simul.* 2018;88:769–84.

AUTHOR CONTRIBUTIONS

JR and SS jointly developed the key conceptual elements of the argument. SS identified relevant prior research on the topic. JR conducted the simulation and wrote the manuscript, with input and review from SS.

COMPETING INTERESTS

JR is employed by Metrum Research Group, and in that capacity has previously provided pharmacometric analysis services for Sunovion, the employer of Seth Hopkins at the time of publication of the manuscript in question. Metrum Research Group and Pharmacometrica, the employer of Roberto Gomeni and Françoise Bressolle-Gomeni, are both providers of pharmacometric analysis services. Potential conflicts of interest for SS are summarized at http://senns.uk/Declaration_Interest.htm.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41398-025-03263-0>.

Correspondence and requests for materials should be addressed to James A. Rogers.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025