

# The Prospective Statistical Power and Positive Predictive Value of Psychotherapies for Treating Depression in adults

François Léonard<sup>1</sup>, Ezio Tirelli<sup>2</sup>



<sup>1</sup> Department of Educational Sciences

<sup>2</sup> Department of Psychology

Presented at the annual meeting of BAPS on May 26, 2023.

## Introduction

Cuijpers et al. (2023) conducted a comprehensive meta-analysis of 409 studies to examine the comparative efficacy of psychotherapies, specifically cognitive behavior therapy (CBT), in relation to control conditions, pharmacotherapies, and combined treatments for depression. The results revealed no statistically significant differences between psychotherapies.

Here a selected subset of studies included in Cuijpers et al. (2023) were scrutinized in terms of potential effect size inflation, statistical power and Positive-Predictive-Value (PPV), the probability of true discovery (Ioannidis, 2005), to gain insights into the methodological quality of studies in the field.

## Method

Studies included in Cuijpers et al. (2023) were sourced from *metapsy* database (Miguel et al., 2022). To establish a more homogeneous subset of studies, two selection criteria were employed: (i) the study must involve adult participants, and (ii) depression must be measured and assessed by a clinician on a continuous scale.

From this subset, a Random-Effects (RE) meta-analysis on the available effect sizes was performed. To account for potential biases arising from small study effects, three correction methods were applied: the Trim-and-Fill, the Rücker's limit method, and a three-parameter selection model. To further assess the strength of the observed effects, median powers were computed for all observed effect sizes and PPV were calculated using an alpha risk level of 0.05.

## Results

107 effect sizes were identified in the database, of which 45.79% were derived from studies that reported multiple effect sizes. Therefore, a decision was made to aggregate these effect sizes using the method outlined by Borenstein et al. (2009). As a result, the aggregation process yielded 79 independent effect sizes. To identify potential influential effects, a Baujat plot was generated (see Figure 1). Subsequently, two effects were excluded from the dataset due to their contribution to between-study heterogeneity, while having minimal impact on the overall pooled effect. Consequently, the final dataset consisted of 77 aggregated effect sizes (105 un-aggregated).

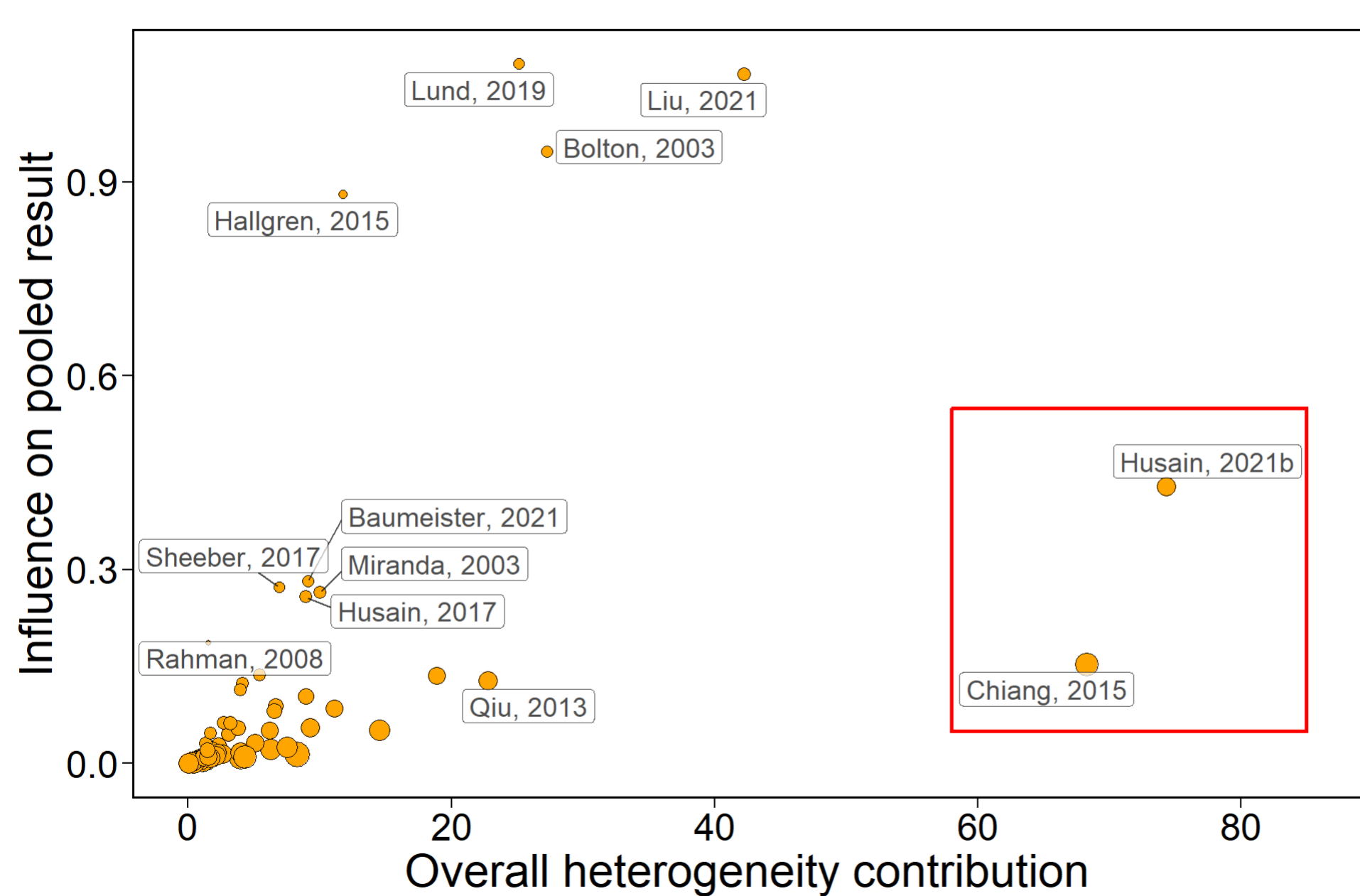


Figure 1: Baujat plot illustrating the aggregated effect sizes. The x-axis represents the contribution of each effect to the overall heterogeneity, while the y-axis represents the influence of each effect on the pooled effect. The red rectangle indicates the two influential effects.

Furthermore, the observed effect sizes from smaller studies exhibited a greater dispersion than those from larger studies, as illustrated in Figure 2. This variation in dispersion could be a sign of a potential effect size inflation.

Regarding the meta-analysis, the RE model was used on the aggregated dataset resulting in a pool estimate of 0.69 (95% CI: [0.57, 0.81]) Hedge's g in favor of psychotherapies. The Trim-and-Fill resulted in an estimate of 0.59 (95% CI: [0.45, 0.73]), the Rücker's limit method resulted in an estimate of 0.39 (95% CI: [0.19, 0.6]), and the selection model resulted in an estimate of 0.66 (95% CI: [0.49, 0.83]). These estimations offer valuable insights into how different correction approaches impact the overall effect size estimation and the robustness of the pooled effect in the meta-analysis.

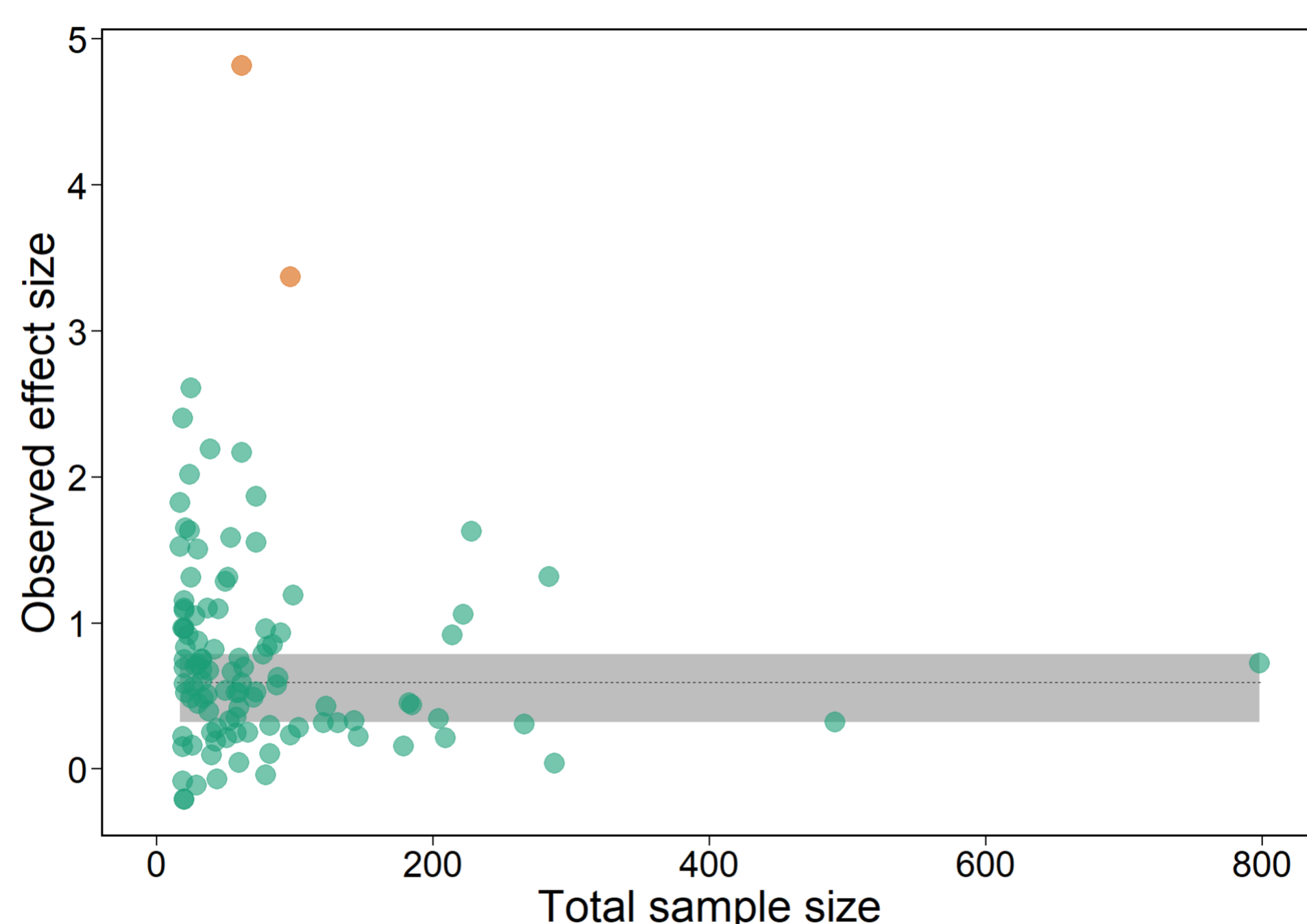


Figure 2: Relationship between total sample size and observed effect size. The two orange dots represent the two influential effects that were excluded from the analysis. The grey area represents the range between the lowest and largest pooled effect sizes. The dashed line represents the mean pooled effect size.

For the four above indicated pooled estimates, medians prospective powers were 0.67, 0.53, 0.28 and 0.63 (see Figure 3 for the distributions of prospective powers).

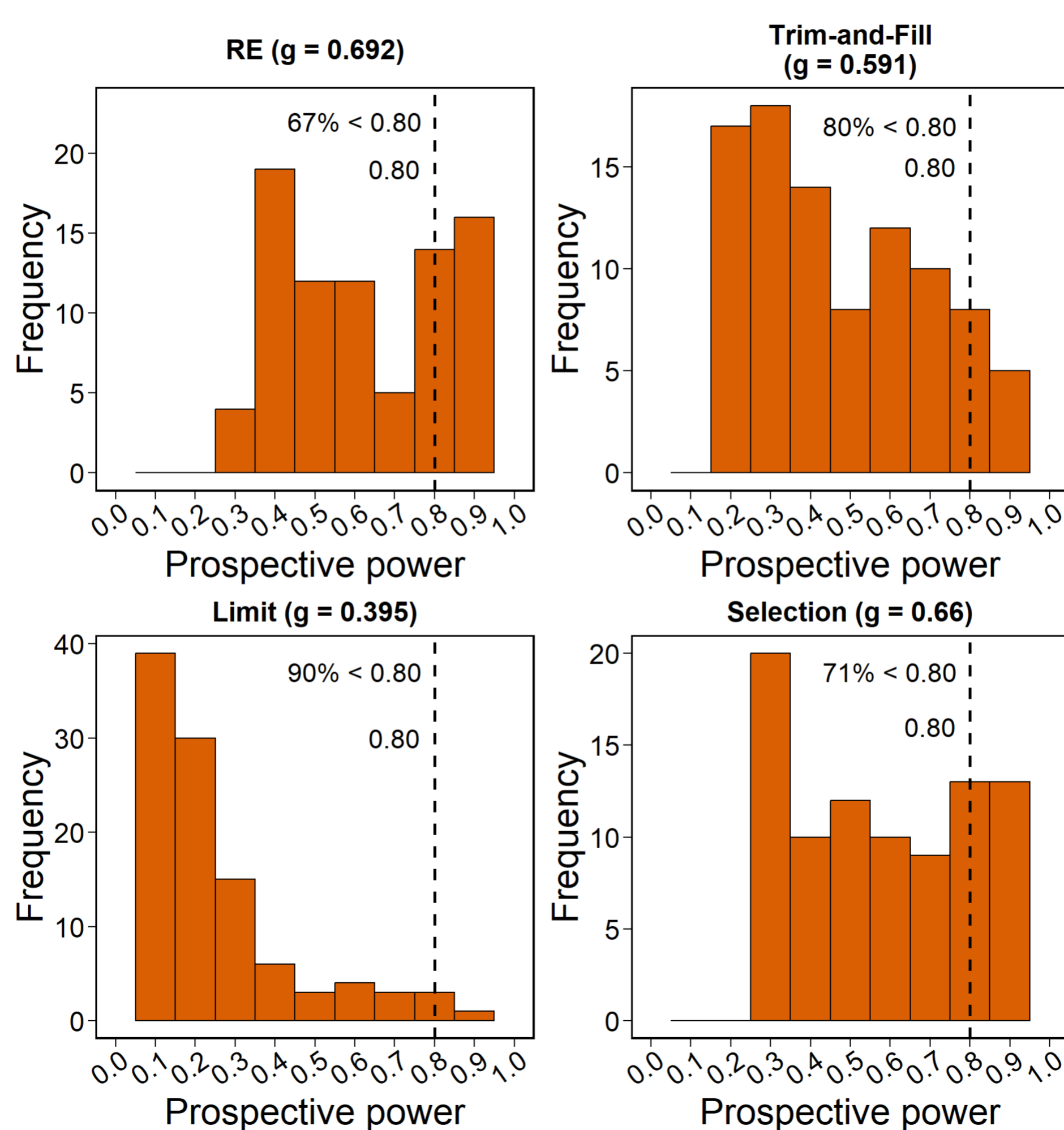


Figure 3: Histograms of prospective powers for four pooled effect sizes. The dashed vertical line indicates the recommended level of power (0.80). Top left panel: powers using the estimate of the RE model ( $g = 0.69$ ). Top right panel: powers using the estimate of the Trim-and-Fill ( $g = 0.59$ ). Bottom left panel: powers using the estimate of the Rücker's limit method ( $g = 0.39$ ). Bottom right panel: powers using the estimate of the selection model ( $g = 0.66$ ).

The percentage of un-aggregated effect sizes with a prospective power lower than 0.80 was 67%, 80%, 90%, and 71% for the RE model, the Trim-and-Fill

method, the Rücker's limit method, and the selection model, respectively.

The PPV was computed for each median prospective powers with hypothetical probabilities of  $H_1$  of being true ( $p(H_1)$ ) set at 0.10 or 0.50. For the estimates derived from the RE model, the PPV were 0.60 and 0.93 for  $p(H_1) = 0.10$  or 0.50. Regarding the estimate obtained from the Trim-and-Fill method, the PPV were 0.54 and 0.91. For the estimate from Rücker's limit method, the PPV were 0.38 and 0.85. Lastly, for the estimate derived from the selection model, the PPV were 0.58 and 0.93.

## Discussion

According to the meta-analysis conducted by Cuijpers et al. (2023), the results suggest that psychotherapies have a moderate to strong effect in treating depression among the adult population. The present study aimed to investigate the methodological quality of a subset of studies included in their meta-analysis by examining potential effect size inflation, prospective power, and PPV.

The subset of studies included here consisted of numerous small-sample studies, and their effect sizes were found to be inflated compared to the mean of the pooled effect sizes. This inflation is likely attributed to the variability and imprecision associated with small-sample studies. To assess the statistical power of the included studies, the pooled estimates from the meta-analytical methods were used. The findings revealed that a significant proportion of studies in the field were underpowered. This means that these studies had a limited ability to detect true positive effects. Consequently, the probability that the significant results observed in these studies represented true discoveries was also diminished.

The high prevalence of underpowered studies highlights the challenge faced by meta-analyses in producing reliable and trustworthy pooled effect sizes. To address these limitations, future clinical studies should be based on larger, adequately powered, and robust methodological designs. In doing so, more reliable evidence of the effectiveness of psychotherapies in treating depression should be obtained, which will improve the quality of research in this field.

## References

- Borenstein, Hedges, Higgins, & Rothstein. (2009). *Introduction to meta-analysis*. John Wiley & Sons.
- Cuijpers, P., Miguel, C., Harrer, M., Plessen, C. Y., Ciharova, M., Ebert, D., & Karyotaki, E. (2023). Cognitive behavior therapy vs. control conditions, other psychotherapies, pharmacotherapies and combined treatment for depression: a comprehensive meta-analysis including 409 trials with 52,702 patients. *World Psychiatry*, 22(1), 105–115. <https://doi.org/10.1002/wps.21069>
- Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, 2(8), 6. <https://doi.org/10.1371/journal.pmed.0020124>
- Miguel, C., Harrer, M., & Plessen, C. Y. (2022). *Metapsy-project/data-depression-psychr: Version 22.0.2*. Zenodo. <https://doi.org/10.5281/zenodo.7255037>