Consolidated Argument

Null Result Penalty Replication

Ryan McWay Emily Kurtz

2025-03-18

Direct Replication

The direct replication was sucessful. But the paper seems almost too good to be true. The point of the paper is the null results are penalized for publication. Yet all the results, even the appendix results, have huge statistically significant effects.

This is strange get the sample. They survey economists and ask them if they would publish a paper. This is measured on a sliding scale of 0 to 100. They provide each person with four of five vignettes. The authors take the vigenttes from real studies that are statistically significant and published. They keep the standard errors the same, but randomize if they shift the coefficient left in the distribution such that the effect is now statistically insignficant.

The get a sample of 480 respondents who complete four vigenettes for 1920 observations. On top of that they cross-randomize 6 other attributes of the vigenettes. Aspects such as gender, prestige, etc. could effect if the finding is publishable beyond statistical significance. This produces 48 treatment assignments using a factorial design. In practice, the authors have 40 observations per treatment assignment to identify off of – 10 respondents. Despite these small clusters, the standard errors are tiny. This makes us suspiciuos.

We examine this in a couple of ways in particular.

1. Variation in the dependent variable
2. Sample composition
3. Quantile Regressions
4. Propensity score matching

The motivation for these robustness checks are to stress test the results in examining if there is potential data manipulation that ensures statistical significance. Our current results suggest that the data is unlikely to have been generated from real world data. The recommendation of this replication is that Chopra et al. (2023) should be replicated using new data with an independent team of researchers.

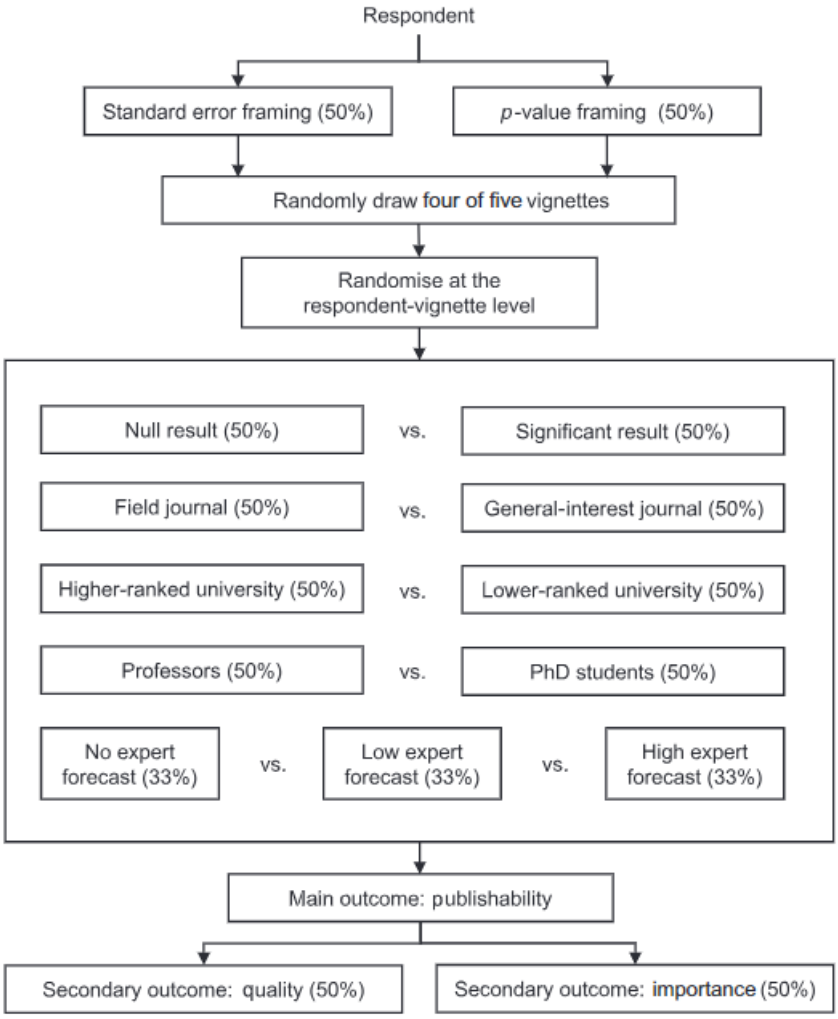

Figure 1: Factorial Design

Variation in the Publishability

Sample Composition

Quantile Regressions

Propensity Score Matching