Evaluation of Bayesian Hypothesis Testing for Direct Replications in a Meta-Analytic Framework

Naike Wang

Texas AM University

Abstract

While the field of psychology is on the verge of a paradigmatic shift toward increased transparency and more replication attempts, a new methodological challenge arises: How do we determine the replication success of multi-lab replication studies? The predominant frequentist method (comparing p-values of original studies with their replications) has been increasingly seen as inadequate. To fill this gap, this study integrates Bayesian methods into the researcher's toolkits by evaluating four Bayesian hypothesis tests (meta-analytic Bayes factor, or MABF) concerning their ability to identify false original findings based on their direct replications. To achieve this, the study adopts a two-phase simulation strategy. The first phase entails generating original findings based on underlying population effect sizes within a simulated research environment shaped by varying degrees of p-hacking, publication bias, and statistical power. The second phase then replicates these findings, applying MABFs for synthesis and analysis. By drawing connections between distinct MABFs that have never been directly compared, this study comprehensively reexamines the potential of Bayesian hypothesis testing as a replication success metric. The study reveals the most robust method(s) for evaluating replication success while also highlighting situations where all methods perform poorly, offering vital methodological guidance for future large-scale, multi-lab replication initiatives.

Keywords: median retrospective power, false positive rate, replication crisis, meta-analysis