Abstract

This study introduces a new replication success classification framework for multi-lab replication studies that goes beyond the traditional binary distinction of replication success and failure. The framework distinguishes between true and false replication successes and failures, revealing the underlying composition of observed replication results and how they are shaped by study-level factors in both the original and replication studies. Using a large-scale simulation, we evaluated four meta-analytic Bayes factor (MABF) methods as tools for determining replication success within this framework. Results showed that the MABF methods were generally more effective at ruling out spurious effects than at confirming small true effects. The one-sided fixed-effect meta-analytic Bayes factor and the inclusion Bayes factor methods most reliably identified genuine replication success. The framework illustrates how bias and methodological choices influence the underlying composition of replication outcomes. These findings provide a more nuanced understanding of replication results and offer practical guidance for designing and evaluating future large-scale replication projects to strengthen research reproducibility.

Keywords: Meta-analysis, Bayesian hypothesis testing, Bayes factor, Replication, Replication crisis