Notation

Statistical Model for Replication Studies

Meta-analysis represents the "results" of an experiment by an effect parameter, which we denote as θ. This "true" effect reflects what the researcher would observe if they has no uncertainty in their estimate; for instance, if they had an infinite sample size. Effect parameters are typically on a standard scale, such as a standardized mean difference, Fisher-transformed correlation coefficient, or log-odds ratio.

In the context of replication studies, we have at least $k \geq 2$ such findings, so let θ_i describe the results of the ith study. For the RPP and RPE programs, this invovles θ_1 for the original study and θ_2 for the replicate. However, we do not actually observe θ_1 and θ_2, and instead we estimate them by T_1 and T_2. These estimates have variances v_1 and v_2. The estimates and their variances are what get reported in typical statistical analyses of experiments.

Two common assumptions are that the estimates are normally distributed with known variances v_i; that is:

$$T_i \overset{indep}{\sim} N(\theta_i, v_i), \quad i = 1, \ldots, k$$

This assumption is almost exactly true for effect sizes such as Fisher's z-transformed correlation coefficient and standardized mean differences, and for others remains a useful large-sample approximation.

For the empirical evaluations in question, we can define replication as the similarity of underlying effect parameters. In that vein, $\theta_1 = \theta_2$ would be a convenient starting point for defining replication. However, there is evidence that even in the face of strong theory and sound scientific practice, that studies may still get slightly different results. Thus, a more practical definition would say studies replicate if $\theta_1 \approx \theta_2$. For further discussion on this distinction see Hedges and Schauer (2018).