

Comparing Type I error and power rates in meta-regression with multiple effect sizes: A study of analytical approaches

Wednesday 23 July 2025 10:45 (15 minutes)

Moderator analyses play a crucial role in meta-analysis, as they help to identify relationships between study characteristics and the effect size magnitude. When multiple effect sizes are reported within studies, various methods can be used to perform moderator analysis or meta-regression. These include three-level models (which may or may not account for variability in moderator effects across studies), Robust Variance Estimation (RVE) methods (with or without the wild bootstrapping technique), and multilevel models combined with RVE. In this study, we conducted a simulation to compare the performance of these methods in terms of Type I error rates and statistical power when performing meta-regressions, focusing specifically on qualitative moderator variables (such as study design or sample type). This focus arises from the common occurrence of unbalanced effect size distributions across moderator categories (i.e., most effect sizes belong to one category, while few belong to others), and it remains unclear which method performs best under these conditions. Additionally, we provide an empirical example of how these differences among methods affect real meta-analyses. To simulate typical meta-analyses, we generated standardized mean differences under varying conditions, such as the number of studies, effect size differences across moderator categories, and average outcome numbers, among others. We analyzed qualitative variables with two or three categories to represent study or effect size characteristics, and the effect sizes were distributed in balanced, unbalanced, or highly unbalanced ways across moderator categories. When simulating three categories, we also used Tukey's multiple comparison correction to assess differences across categories.

Results showed that when the qualitative variable referred to effect size characteristics, the three-level model that did not account for moderator effect variability (the one commonly implemented in practice) had highly inflated Type I error rates, while other methods maintained acceptable rates. Power levels were generally lower when the moderator referred to effect size characteristics, and these were minimally affected by unbalanced effect size distributions across categories. When the moderator referred to study characteristics, all methods exhibited acceptable Type I error rates, but power was inadequate, particularly when effect sizes were highly unbalanced. Across all conditions, three-level models combined with RVE provided the best Type I error-power balance, although power remained very low.

In conclusion, this study suggests that, in the presence of multiple effect sizes within studies, multilevel models should always be applied with RVE correction when conducting meta-regressions. Additionally, further advancements are needed to generally improve power for detecting moderator effects.

Primary authors: FERNÁNDEZ CASTILLA, Belén (Universidad Nacional de Educación a Distancia); LOPEZ LOPEZ, Jose Antonio; RUBIO APARICIO, María (University of Murcia)

Presenters: FERNÁNDEZ CASTILLA, Belén (Universidad Nacional de Educación a Distancia); LOPEZ LOPEZ, Jose Antonio; RUBIO APARICIO, María (University of Murcia)

Session Classification: Symposium : "Recent Developments in Meta-Analysis"