Exploratory Analyses for Missing Data in Meta-Analyses

Introduction

Systematic reviews of substance abuse literature hold great promise for unpacking correlates of effective substance abuse interventions. Methodological tools such as meta-regression can formally test relationships between the type or implementation of an intervention and how effective it is. However, such tools must contend with the real-world difficulties of modern research syntheses, including the fact that it is often impossible to extract relevant information from the literature.

The fact that not every study reports the information required to run a meta-regression means that many meta-analyses run into a missing data problem. Issues with missing data are not new. There is a large literature on methods for handling missing data in primary studies, as well as some work on related issues in meta-analysis. This literature highlights the ways that missingness can bias an analysis, examines conditions under which these biases can be corrected, and proposes various statistical procedures to adjust for bias.

A key assumption of most missing data methods is that the analyst has some idea about which of their data are missing and why. However, while much of the literature has focused on the implications of that assumption, considerably less attention is paid to approaches to examining it in a dataset. In fact, most work on summarizing missingness in a dataset and examining its sources arises in literature on graphical summaries of data.

This is inherently an exploratory data anlysis (EDA), wherein the analyst seeks to identify patterns of missing variables in their data and what those may be correlated with. As such, there is no single procedure or silver bullet for a given dataset. Instead, analysts...

Missing Data and Meta-Analysis

In the context of meta-analysis, "missing data" is a broad term that can be used to describe several different types of scenarios. For instance, data could be missing on individual participants within studies, including their outcomes in the study or other characterisitcs (e.g., their age, race, prior substance use). "Missing data" could also refer to scenarios where information cannot be extracted from a completed study by a meta-analyst. This might occur if a study fails to report enough detail for analysts to back out effect estimates, standard errors, or study-level characteristics. Finally, entire studies or effects may be missing from a meta-analytic dataset. This might occur if effects (or entire studies) are not reported or published. There is empirical evidence that statistically significant results are more likely to be published and hence wind up in a meta-analysis, which can induce *publication bias*, a well-known problem in the field. The studies or effects that are not reported, and thus are not included in a meta-analysis, can be considered missing data.

Precisely how to examine, diagnose, and adjust for missing data will be different depending on what scenario we mean when we say "missing data." For instance, meta-analysts have used "funnel plots" to examine if their systematic review is missing studies or effects due to publication bias. Our focus will be on the second scenario, where information cannot be extracted from some studies. This is a common problem in meta-analysis and one that that can limit the accuracy of any statistical inferences.

Assume we have data on k effect estimates and p variables (including the estimate itself). This can be summarized and stored in a $k \times p$ table where rows correspond to effect estimates and columns correspond to variables concerning those estimates. One column would contain the effect estimates themselves, and another would contain the standard error or estimation error variance of those estimates. The remaining $p - 2$ columns could contain effect- or study-level covariates, including summary demographics (e.g., the percent of a study's sample that were minorities), treatment type (e.g., behavioral therapy versus pharacological interventions), or dosage/duration of an intervention. Some of the cells in this table may be

missing values, and the analyses presented in this article provide ways to summarize and examine patterns of missingness.

Data

To highlight the principles of exploring missingness, we use data from...