Technical Review

Chapter 1. Introduction

Dealing with heterogeneity among study treatment effects, or "the situation in which differences in study outcomes are not readily accounted for by sampling variation,"[1] is one of the most important challenges facing a meta-analyst. Current guidelines for the reporting of meta-analyses in both the randomized controlled trial setting[2] and the observational study setting[3] state that the degree of heterogeneity should be assessed, the sources of heterogeneity should be understood if at all possible, and failing an explanation, this variability should be accounted for, i.e., incorporated, in meta-analytic estimates and policy conclusions.

The National Center for Complementary and Alternative Medicine (NCCAM) recognized the importance of heterogeneity in meta-analysis. With the Agency for Healthcare Research and Quality (AHRQ), NCCAM established the objective of this study to compare and contrast several strategies for understanding heterogeneity via meta-regression methods. They asked the Southern California Evidence-Based Practice Center, in its role as technical support to NCCAM, to conduct the study and to produce this report.

Heterogeneity

Sources of Heterogeneity

Differences among studies may be categorized broadly into those related to the phenomenon being studied and those unrelated. Following the terminology of Thompson,[4] we shall refer to these dimensions as clinical incomparability and design incomparability respectively. Those differences related to the phenomenon being studied are mostly beyond the original investigator's control and constitute clinical incomparability. For example, the treatment may work differently for specific populations, the treatment may have a different effect on mortality measures as compared to morbidity, or the treatment effect may depend on exposure level. The original investigator may focus on a particular patient subgroup to reduce such incomparability.

The investigator may control the design dimension of incomparability. For example, he/she may control whether the study is prospective or retrospective, how long to follow the patients, what outcome to measure given measurement error issues, whether to analyze an odds ratio or a risk difference, and how to analyze that statistical outcome. Choice of study design, or how certain problems such as attrition are dealt with analytically, may induce differential biases in the results as well. Researchers may actually plan differences across studies to induce heterogeneity and increase generalizability, and assessing and understanding such differences is a strength of systematic reviews.

Measuring Heterogeneity

We consider the randomized controlled trial setting, restricting attention to dichotomous study outcomes and choosing as the summary statistic the odds ratio. For example, the outcome might be mortality within a specified follow-up time, and the summary statistic would be the odds of death in the treatment group as compared to the control group. The usual first step is to assess whether heterogeneity exists using a chi-squared test (a Q-statistic).[5] This test is known to have low statistical power,[6] which means that the probability that the null hypothesis of

homogeneity of study treatment effects is rejected given that the alternative hypothesis of heterogeneity is true, is small. Thus non-rejection of the null hypothesis does not necessarily mean that heterogeneity does not exist, and the meta-analyst is well-served to consider that heterogeneity exists regardless and attempt to estimate it.

Addressing Heterogeneity

If heterogeneity is found or suspected to exist, the common approaches used in meta-analysis are to

- Stratify the studies into homogeneous subgroups and then fit a separate fixed effects estimate,[7] e.g., of the pooled odds ratio, in each strata.
- Construct a random effects[7] estimate, e.g., DerSimonian and Laird[8] pooled odds ratio, across all studies. A random effects approach incorporates both within-study and between-study variability. We note that some argue if heterogeneity exists among studies, summary measure across those studies should not be provided.
- Fit a meta-regression model that explains the heterogeneity in terms of study-level covariates. This is the focus of this report.

Meta-regression

A meta-regression can be either a linear or logistic regression model. In most meta-regression approaches, the unit of analysis, that is each observation in the regression model, is a study. Sometimes an arm, e.g., a specific treatment arm or the control arm, or even an arm crossed with outcome, e.g., all patients in a specific treatment who had the outcome, is the unit of analysis. For the moment we will consider the simplest case in which the unit of analysis is a study. The outcome for a study observation might be the log odds ratio for example. Predictors in the regression are at the study-level and might include such factors as the medicine protocol, characteristics of the study population such as average age, or variables describing the study setting such as whether the hospital in which the study is undertaken is a teaching hospital.

The questions that a meta-analyst may answer with a meta-regression include estimating the treatment effect controlling for differences across studies, and determining which study-level covariates account for the heterogeneity. The difficulties faced in a meta-regression are many. Primarily, the degrees-of-freedom available can be small due to the fact most meta-analyses do not include a large number of studies. In addition, covariates tend to be highly collinear, for example all studies in rural areas may administer the medicine in a particular way, while urban hospitals use a different protocol. In such cases, it is impossible to disentangle the effects of individual covariates. The problem of ecological bias[9] is paramount, as the analysis is conducted at the study-level and does not include the underlying patient-level variation. Several publications discuss the pitfalls of meta-regression.[10, 11]

Meta-regression Approaches

We now briefly describe the four major meta-regression approaches presented in the literature.

The first approach is a fixed effects approach which utilizes logistic regression.[12] In this method, a weighted logistic regression of the $2k$ cases per study is fit where k is the number of study arms, and the weight is the number of patients who have or do not have the outcome respectively. Covariates can either be study or arm level, and interactions with treatment can be fit.

The second approach is random effects meta-regression.[13] Generally the log odds ratio is regressed on an intercept and study-level covariates. The terminology "random effects" refers to the fact that a random study effect is included in the regression to take into account the between-study variation. In the simplest case in which only an intercept term is included, this approach reduces to the usual DerSimonian and Laird random effects estimate of the pooled odds ratio.[8]

The third approach is control rate meta-regression.[14, 15] In this setting, the single covariate is outcome rate (e.g., mortality rate) in the control group. The hypothesis is that the control rate is a surrogate for covariate differences between the studies.

The fourth general approach is Bayesian hierarchical modeling, which we have not included in our simulation study. Many references exists on this topic including DuMouchel;[16] Louis and Zelterman;[17] and Smith, Spiegelhalter, and Thomas.[18]

Report Outline

The second chapter of this report describes the methodology we applied, including our systematic review approach, the strategy of producing a common statistical notation that would allow us to compare and contrast various meta-regression approaches, the simulation, and the methodology we used to constitute and work with our expert panel. The third chapter contains results. We report on the systematic review, the resulting bibliography, and the common statistical notation. We describe our preliminary simulation set-up. The expert panel made recommendations that included slight changes to the simulation parameters. These recommendations were implemented and the simulation results reported were based on the revised parameters. The final chapter of the report includes recommendations, conclusions, and future research.