Chapter 3. Results

Article Retrieval Results

The search of the first seven library databases (MEDLINE®, HealthSTAR, EMBASE, MANTIS, SciSearch®, Social SciSearch®, Allied and Complementary Medicine) produced 166 titles. The search of the Current Index to Statistics produced 16 titles, and the search of the Methodology Register of the Cochrane Library produced 135 titles. The titles were not deduplicated across these three searches. Our canvassed experts and referees, the Southern California Evidence-Based Practice Center methodological article database, and searching of reference lists of relevant articles yielded 27 additional titles. We note as an aside that by article we mean a published document including journal articles, books and reports. These combined searches produced 85 relevant articles whose full text was obtained.

Reference List

This final reference list is given in the Bibliography. We note two issues about this bibliography. First, we did not search using terms associated with hierarchical or Bayesian hierarchical modeling, which is a large field of literature. Our experts did identify some hierarchical modeling publications that are relevant to meta-analysis, and we have included those publications in our bibliography. Second, the application of meta-regression is becoming more common in meta-analysis examples. Thus, while our bibliography contains some examples of the application of meta-regression, especially early examples, our bibliography is by no means an exhaustive list of meta-regression applications.

Given these caveats, we categorized the 85 publications into seven categories based on the primary focus of the article (Table 1). The first four categories were the main methods: fixed effects models; random effects models; control rate models; and Bayesian and/or hierarchical models. We also defined an “overview” category that contained articles that surveyed meta-regression methods and/or focused on the unique challenges of such a modeling effort, including for example discussion of ecological bias. Our sixth category consisted of articles that addressed modeling studies that had multiple treatment arms and/or multiple endpoints or outcomes, as such studies present unique challenges. Our seventh category consisted of examples. We note that obviously publications could fall into more than one category, for example most articles that addressed random effects models began with a discussion of fixed effects models, so we categorized studies according to their primary focus.

Common Notation

We restrict attention to dichotomous outcomes, and specify the relationship between treatment, covariates, and outcome at the person level.

First we consider the probability of the outcome in the absence of treatment. For clarity we will restrict our attention to dichotomous outcomes with a logistic link function, i.e., the logistic model is the correct model at the person level. Each person has a baseline effect (in the log odds

scale) in the absence of treatment, ϕ_{ij}, for person *j* in study *i*. The log odds probability of the outcome in the absence of treatment is given by:

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right)=\phi_{ij}$$

For example, ϕ_{ij} would be the log odds probability of mortality in a specified follow-up time if the individual did not receive treatment.

This baseline effect may conditional on characteristics of the patient and the study. Given a study effect ϕ_i, a vector of study level covariates $\vec{z}_j$, and a vector of person-specific covariates $\vec{x}_{ij}$ for person *j* in study *i*, the baseline effect for this individual is:

$$\phi_{ij}=\phi_i+\beta_2\vec{z}_i+\beta_3\vec{x}_{ij} \qquad (1)$$

where the β vector is:

β_2: the effect of a study-level covariate, such as inpatient versus outpatient service delivery, on the outcome

β_3: the effect of a person-level covariate, such as the patient's age, on the outcome

Next we will consider the probability of the outcome for a patient who receives treatment. We denote the treatment effect for person *j* in study *i* to be τ_{ij}. In the simplest case, the log odds probability of the outcome in the presence of treatment is given by:

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right)=\phi_{ij}+\tau_{ij} \qquad (2)$$

In the same way that a person's baseline effect can depend on a study effect ϕ_i, a vector of study level covariates $\vec{z}_j$, and a vector of person-specific covariates $\vec{x}_{ij}$ for person *j* in study *i*, we can write the treatment effect τ_{ij} as

$$\begin{aligned}\tau_{ij}&=\gamma_0+\gamma_1\phi_i+\gamma_2\vec{z}_i+\gamma_3\vec{x}_{ij}+\upsilon_i && \text{if person } ij \text{ receives the treatment}\\ \tau_{ij}&=0 && \text{otherwise}\end{aligned} \qquad (3)$$

where the γ vector is:

γ_0: classic additive treatment effect

γ_1: a treatment effect that depends on the underlying prevalence of the outcome in the absence of treatment

γ_2: a treatment effect that depends on a study-level covariate

γ_3: a treatment effect that depends on a person-level covariate

υ_i: a random effect for study *i*, introducing unexplainable heterogeneity

If we were to estimate the parameters in Model (2) empirically, the model specification would include main effects for the covariates, a treatment indicator, and treatment indicator by

covariate interaction terms. Alternatively, we can substitute study-level indicator variables for the main effects of the study-level covariates. Study-level indicator variables and study-level covariates cannot both be included since they are cofounded. In the first specification, the coefficients of the treatment by covariate interaction terms estimate the γ vector.

The joint distribution of ϕ_i, z_i and $\bar{x}_{ij}$ may be arbitrary. However, to conduct a simulation, we need to specify the joint distribution. Exploration of other distributional assumptions would be straightforward. We consider the Normal special case:

$$(\phi_i, z_i, \mu_i) \sim N\left(\begin{pmatrix}\theta \\ \lambda \\ 0\end{pmatrix}, \Sigma_{\phi z \mu}\right)$$

and

$$x_{ij} \sim N\left(\mu_i, \sigma_{xi}^2\right)$$

where

$$\Sigma_{\phi z \mu} = \begin{pmatrix} \sigma_\phi^2 & \sigma_{\phi z} & \sigma_{\phi \mu} \\ \sigma_{\phi z} & \sigma_z^2 & \sigma_{z\mu} \\ \sigma_{\phi \mu} & \sigma_{z\mu} & \sigma_\mu^2 \end{pmatrix} \tag{4}$$

The zero mean for μ_i, the mean of the x_{ij} in study i, is arbitrary. Study effects, study covariates, and the mean of the person effects may be correlated. The random study effect υ_i could be integrated into this framework and correlated with the other covariates but leaving it independent is convenient: $\upsilon_i \sim N(0, \sigma_\upsilon^2)$.

We will treat β and γ as vectors of constants. However, note that by setting $\beta \sim N(\mu_\beta, \Sigma_\beta)$ and/or $\gamma \sim N(\mu_\gamma, \Sigma_\gamma)$, we could generalize this framework to random coefficients models.

Aggregating to the Study Level

We can write each meta-regression approach as an aggregated version of the person-level model. Our approach is inspired by the aggregation of models literature in econometrics, see for example Theil.[19]

The person-level parameterization that results in this logistic regression (Model (2)) is our fundamental representation of the treatment's effect on outcome and the factors that determine that treatment effect. However, the typical meta-analyst does not often have access to the person-level data. We consider two different aggregations of the person-level data to study-level data that are commonly available.

First, the study may aggregate the outcomes to a two-by-two table, successes and failures in the treatment and control groups separately. In the case of multi-arm trials, this would

correspond to a two by k table where k is the number of arms. Potential explanatory variables may be aggregated to the study level or to the study by arm level.

Second, the study may aggregate to a single treatment effect summary for the study (e.g. an odds ratio or risk ratio.) Potential explanatory variables would typically be aggregated to the study level as well.

"Logistic meta-regression"[12] retains the two by k table and performs a logistic regression with $2k$ cases per study (success/failure by study arm.) This approach allows the use of standard software. The technical problem is the aggregation of explanatory variables. Such variables may be available aggregated at the study level, the arm level, or at both levels. Mis-specification will occur if the aggregation of covariates is only available at the study level. Randomization may limit the effect of this mis-specification. If aggregation is available at both the study and arm levels, the arm-level data is preferred and accommodated by our approach.

Many meta-regression approaches model a single summary statistic per study. For example, Berkey, Hoaglin, Mosteller, et al.[13] annotate a meta-regression as:

$y_i = \log\ (RR_i)$ where RR_i is the relative risk for study i

$$y_i = \alpha \bar{x}_i + \delta_i + \varepsilon_i$$

This is not a fundamentally different structuring of the problem than logistic meta-regression. The log risk-ratio is one possible aggregation of the person-level logistic model outcomes. The ε_i now captures the variability in the binomial process.

However, this additional aggregation step does add a potential source of bias. Recall in the logistic meta-regression approach, as long as arm-level data are available, we have only aggregated to the arm level and only introduced ecological bias from variables that are actually person-level predictors. If further aggregation to the study level is done, additional bias may be introduced due to the mis-specification of functional form, e.g., the risk ratio may not be the most appropriate study-level summary of the treatment effect to use in a meta-regression analysis. We do note that if covariates are specified at the study level in the logistic model, the results are comparable to those obtained in a model that fits the natural logarithm of the odds ratio. This comparability can be important if one is comparing two modeling approaches.

Scenarios in which Meta-regression Might Be Informative

We now consider four common situations in which meta-regression might be applied. We present the person-level specification of each scenario and discuss the most relevant meta-regression methods. In each scenario, if a parameter is nonzero, it is set equal to the same constant for uniformity. To give two examples, in all scenarios the classical additive treatment effect γ_0 is nonzero and is always set equal to g_0; and in the third and fourth scenarios the treatment effect that depends on the underlying prevalence of the outcome in the absence of treatment γ_1 is nonzero and is set equal to g_1.

Scenario 1: Studies have different baseline effects and additive fixed treatment effects.

$$\beta = (0,0)$$
$$\gamma = (g_0, 0, 0, 0)$$

$$\begin{pmatrix} \theta \\ \lambda \end{pmatrix} = \begin{pmatrix} k_1 \\ 0 \end{pmatrix}$$

$$\Sigma_{\phi z \mu} = \begin{pmatrix} c_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\sigma_{\upsilon}^2 = 0$$
$$\sigma_{xi}^2 = 0$$

Simple fixed effects pooling methods, e.g. the Mantel-Haenszel method[22] for combining odds ratios, are appropriate in this scenario. Meta-regression methods may not be efficient but may not be very biased.

Scenario 2: Studies have different baseline effects and an additive random treatment effect.

$$\beta = (0,0)$$
$$\gamma = (g_0, 0, 0, 0)$$

$$\begin{pmatrix} \theta \\ \lambda \end{pmatrix} = \begin{pmatrix} k_1 \\ 0 \end{pmatrix}$$

$$\Sigma_{\phi z \mu} = \begin{pmatrix} c_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\sigma_{\upsilon}^2 = c_7$$
$$\sigma_{xi}^2 = 0$$

Simple random effects pooling methods, e.g., the DerSimonian and Laird[8] method for combining risk ratios, are appropriate in this scenario. Meta-regression methods that incorporate

random effects, e.g., Berkey, Hoaglin, Mosteller, et al.,[13] may be applicable in this scenario but may not be efficient.

Scenario 3: Studies have different baseline effects and the treatment effect depends on the underlying prevalence of the outcome in the absence of treatment.

$$\beta = (0,0)$$
$$\gamma = (g_0, g_1, 0, 0)$$

$$\begin{pmatrix} \theta \\ \lambda \end{pmatrix} = \begin{pmatrix} k_1 \\ 0 \end{pmatrix}$$

$$\Sigma_{\phi z \mu} = \begin{pmatrix} c_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\sigma_{\upsilon}^2 = 0$$
$$\sigma_{xi}^2 = 0$$

Control rate meta-regression methods, e.g., McIntosh[14] and Schmid, Lau, McIntosh, et al.[15] are appropriate in this scenario. Meta-regression methods that model treatment indicator by covariate interaction terms may also be appropriate, although perhaps not as efficient as the control rate approaches.

Scenario 4: Studies have baseline effects and treatment effects that depend on covariates at the study and person levels.

$$\beta = (b_2, b_3)$$
$$\gamma = (g_0, g_1, g_2, g_3)$$

$$\begin{pmatrix} \theta \\ \lambda \end{pmatrix} = \begin{pmatrix} k_1 \\ k_2 \end{pmatrix}$$

$$\Sigma_{\phi z \mu} = \begin{pmatrix} c_1 & c_4 & c_5 \\ c_4 & c_2 & c_6 \\ c_5 & c_6 & c_3 \end{pmatrix}$$

$$\sigma_{\upsilon}^2 = 0$$
$$\sigma_{xi}^2 = c_8$$

Fixed effects meta-regression methods that allow covariates, e.g., logistic meta-regression and the Hasselblad approach[20] are appropriate in this scenario.

Simulation Set-up

Our simulation set-up consists of distributional assumptions and ranges of parameters for which we will generate person-level data. Following the order of presentation for the common notation, we will discuss the parameter values for the baseline effect and for the probability of outcome for treated patients, and then the distributional assumptions. This simulation set-up will allow us to generate cases where the generating mechanism for the data is known exactly with explicit assumptions. Further, our expert panel vetted this set-up.

Baseline Effect

We begin with the baseline effect ϕ_{ij} for person *j* in study *i*, and restate Model (1):

$$\phi_{ij} = \phi_i$$

The baseline effects for all persons in study *i* are equal to a single study effect ϕ_i. The ϕ_i values are drawn from a normal distribution with mean ϕ (described below) and variance one. The variance assumption results in no loss of generality as the simulation can introduce variance via other parameters. For simplification, β_2 representing the effect of a single study-level covariate z_i, and β_3 representing the effect of a single person-level covariate x_{ij}, are set to zero. Note that in the event a patient receives treatment, these covariates will still have an effect on treatment as described in the next section.

Probability of Outcome in the Presence of Treatment

As in Model (2), we denote the treatment effect for person *j* in study *i* to be τ_{ij}, and the log odds probability of the outcome in the presence of treatment is given by:

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \phi_{ij} + \tau_{ij}$$

The treatment effect τ_{ij} depends on a single study level covariate z_i, and a single person-specific covariate x_{ij} for person *j* in study *i* as

$$\tau_{ij} = \gamma_0 + \gamma_1\phi_i + \gamma_2 z_i + \gamma_3 x_{ij} + \upsilon_i \qquad \text{if person } ij \text{ receives the treatment}$$

$$\tau_{ij} = 0 \qquad \text{otherwise}$$

where the γ vector describes relationship between study and person characteristics and treatment effect with γ_0 representing the classic additive treatment effect, γ_1 representing a treatment effect that depends on the underlying prevalence of the outcome in the absence of treatment, γ_2 representing a treatment effect that depends on a study-level covariate, γ_3 representing a treatment effect that depends on a person-level covariate, and υ_i representing a random effect for study *i*, introducing unexplainable heterogeneity. We have reduced from vectors of study-level covariates and person-specific covariates as shown previously in Model (3) to single covariates in each case for simplicity.

Distributional Assumptions

In order to conduct the simulation, we need to specify the joint distribution of s_i, z_i and x_{ij} as presented in Equation (4). We set the distributions and their parameters as follows:

$$(\phi_i, z_i, \mu_i) \sim N\left(\begin{pmatrix}\phi\\0\\0\end{pmatrix}, \Sigma_{\phi z\mu}\right)$$

$$x_{ij} \sim N\left(\mu_i, \sigma^2_{xi}\right) \text{ and } \upsilon_i \sim N\left(0, \sigma^2_{\upsilon i}\right)$$

where

$$\Sigma_{\phi z\mu} = \begin{pmatrix}1 & \rho & \rho\\ \rho & 1 & \rho\\ \rho & \rho & 1\end{pmatrix}$$

The reader should note that γ_0, the additive treatment effect, is defined at the mean value (zero) of z_i and x_{ij}. This set-up corresponds to a meta-analysis that is unbiased for the population, i.e., the analyst has a random sample of studies. Setting the means equal to zero and variances equal to one result in no loss of generality as the simulation can introduce additional complexity via other parameters.

Simulation Parameters

Table 2 contains the parameters used in the simulation. The ϕ values from –0.6 to 6 correspond to odds ratios between 0.55 to 1.82. These values were selected to cover a range of outcome probabilities from 35% to 65%. Table 2 also shows the corresponding odds ratio values for the other coefficient parameters in the simulation. The simulated distributions are multiplied by the coefficient parameters we have selected, so assuming variances of one and means of zero in our distributional assumptions stated previously results in no loss of generality.

Meta-regression Methods Evaluated

We evaluated five methods using the odds ratio as the statistic of interest for the comparability.

Method 1: Fixed effects pooled odds ratio

For comparison purposes, we begin with the "Fixed Effects with No Covariates" method in which we fit a fixed effects pooled log odds ratio. This model may be written as

$$log(OR_i) = \tau + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma_i^2)$$

which is analogous to Method 2 described below with an intercept term only.

Method 2: Logistic meta-regression

In this "Fixed Effects with Covariates" method, we fit a weighted logistic regression[12] of the 2*k* cases per study where *k* is the number of arms, both control and treatment, in the study. Each arm contributes two observations to the regression: those patients in the arm who have the outcome and those patients who do not have the outcome. Thus each study contributes an observation from each cell in the two by *k* table of arm by outcome. The weight for each observation represents the number of patients in that particular cell. Covariates can either be study or arm level, and interactions with treatment can be fit.

This model may be written as

$$\log\left(\frac{p_i}{1-p_i}\right) = \phi_i + \tau_i$$

$$\phi_i = \psi + \beta_2 z_i + \beta_3 \overline{x}_i$$

$$\tau_i = \gamma_0 + \gamma_1 \phi_i + \gamma_2 z_i + \gamma_3 \overline{x}_i \qquad \text{in the treatment arm of study } i$$

$$\tau_i = 0 \qquad \text{otherwise}$$

We implemented this model in SAS.[23]

Table 3 demonstrates the data layout and levels of covariates. Study 1 has a control arm (first two rows) and a single treatment arm (third and fourth rows). For each pair of rows associated with an arm, the first row are those patients without the outcome ("failures" with outcome = 0) and the second row are those patients with the outcome ("successes" with outcome = 1). The number of cases in each row are given, these will serve as the weights in the logistic regression. An example study-level covariate is given. This covariate has the same value for all

observations in a study as it is at the study level. An example might be the average age of the participants in the study across all arms. An arm-level covariate example is also shown, an example might be the average age in each arm, e.g. while the overall average age is 40, the average ages in the control and treatment arms are 45 and 35 respectively.

Method 3: DerSimonian and Laird random effects pooled odds ratio

In this "Random Effects with No Covariates" method, we applied the standard one-step DerSimonian and Laird random effects pooled estimate of the log odds ratio:[8]

$$y_i = \log(OR_i) = \tau + \upsilon_{\mathrm{i}} + \varepsilon_{\mathrm{i}}$$

$$\upsilon_{\mathrm{i}} \sim N(0, \sigma_\upsilon^2)$$

$$\varepsilon_i \sim N(0, \sigma_i^2)$$

with υ_{i} and ε_{i} uncorrelated. We implemented this method in the statistical software package Stata[24] using the "metareg" command with the method of moments estimation option and an intercept term but no covariates (in our experience, this is roughly equivalent to using the "meta" command).[25]

Method 4: Random effects meta-regression of the log odds ratio with covariates

In this "Random Effects with Covariates" method, we fit a random effects meta-regression that regressed the log odds ratio on an intercept and study-level covariates:[25]

$$y_i = \log(OR_i) = \gamma_0 + \gamma_2 z_i + \upsilon_{\mathrm{i}} + \varepsilon_{\mathrm{i}}$$

$$\upsilon_{\mathrm{i}} \sim N(0, \sigma_\upsilon^2)$$

$$\varepsilon_i \sim N(0, \sigma_i^2)$$

with υ_{i} and ε_{i} uncorrelated. We implemented this method in Stata[24] using the "metareg" command with restricted maximum likelihood estimation.[13]

Method 5: Control rate meta-regression

In this "Control Rate" method, we fit a random effects meta-regression that regressed the log odds ratio on an intercept and the control group outcome rate:[14, 15]

$$y_i = \log(OR_i) = \gamma_0 + \gamma_1 \phi_i + \upsilon_{\mathrm{i}} + \varepsilon_{\mathrm{i}}$$

$$\upsilon_{\mathrm{i}} \sim N(0, \sigma_\upsilon^2)$$

$$\varepsilon_i \sim N(0, \sigma_i^2)$$

with υ_{i} and ε_{i} uncorrelated. We implemented this method in S-PLUS,[26] using software courtesy of Drs. McIntosh and Schmid that utilizes the EM algorithm.

How the Simulation Works and is Evaluated

The total number of simulation parameter combinations is 1,944. (We note that this number changed subsequently based on our panel's recommendation.) For each combination of values, we generate one meta-analysis data set and apply each of the five methods. Originally, we proposed the size of this meta-analysis be ten studies, each with 200 patients based on Schmid et al.[15] These authors reported a median number of studies equal to eight in Cochrane meta-analyses and 11.5 in medical journal meta-analyses, with median number of patients equal to 177

and 265 respectively. However, as discussed in the next section, our expert panel recommended that we conduct the simulation over a variety of meta-analysis sizes and study sample sizes.

We compare the methods in terms of bias in the estimation of γ_0, the additive treatment effect. This is the key parameter, typically estimated in meta-analyses. In the tables that follow in this chapter, we present that bias as a percentage of the true parameter:

$$100\left[\frac{ln(OR_{estimated}) - ln(OR_{true})}{ln(OR_{true})}\right]$$

What These Methods Are Estimating in Our Simulations

The population mean treatment effect is the expected value of treatment effects across all patients in all studies. From Model (3):

$$E(\tau_{ij}) = E(\gamma_0 + \gamma_1\phi_i + \gamma_2 z_i + \gamma_3 x_{ij} + \upsilon_i)$$
$$= \gamma_0 + \gamma_1\phi$$
$$\text{since } E(z_i) = E(x_{ij}) = E(\upsilon_i) = 0$$

In the absence of a control rate, i.e., $\gamma_1 = 0$, the population mean treatment effect is simply γ_0. Thus for all models except control rate, we can estimate γ_0 by just averaging across all patients in all studies.

Averaging across all patients in all studies in the control rate model yields $\hat{\gamma}_0 + \hat{\gamma}_1\hat{\phi}$. Thus to estimate γ_0 from this model, we need to subtract $\hat{\gamma}_1\hat{\phi}$ from the average treatment effect.

Panel Recommendations

The panel was enthusiastic about the common notation and preliminary simulation set-up, and noted the usefulness and timeliness of the projected product.

Recommendations Regarding the Simulation

We begin with recommendations regarding the simulation:

- The panel recommended that we vary the number of studies and number of patients within those studies. Our original simulation design fixed these parameters at 10 and 200 respectively. The panel recommended that we evaluate the design with the sample size for studies varying within each meta-analysis, and with meta-analyses of size 3, 10, and 30 studies.

Following this recommendation, we varied the sample sizes within the studies from as few as five patients within a study to as many as 395 patients within a study. The variable "*tilt*" measures the degree of variability of sample sizes across the studies. *Tilt* equal to zero means all studies have sample size 200. *Tilt* equal to one means the studies are uniformly distributed between 5 and 395 patients with an average sample size of 200.

The meta-regression methods we considered were not capable of producing stable parameter estimates when only three studies were available. This outcome is not surprising since there are only two degrees-of-freedom for study effects in meta-regressions with three studies. Thus we decided to include only two levels of number of studies: k = 10 and 30. The addition of the variable *tilt* and two values of k increased our total number of simulation parameter combinations from 1,944 to 7,776.

- We should use symmetry in the simulation parameters to reduce the number of meta-analyses to be evaluated in the simulation, i.e., decrease the size of the design.

We decided that improvements in computational efficiency made it unnecessary to reduce the size of the simulation.

- To make the simulation results most useful and comprehensible, the panel recommended we:
  - Relate the simulation scenarios to realistic (clinical) situations that analysts would commonly find themselves in.

After some discussion, we believe our values of the simulation parameters span the range of common circumstances encountered by the meta-analyst.

  - Define what precisely we mean by bias in our evaluation of the simulation.

We have done this previously in this document—see the formal definition of bias.

  - Define when our model is identifiable, that is when the simulation parameters can actually be estimated by the meta-regression methods.

In this report, we focus on bias in the estimation of γ_0, a parameter which is identifiable for all methods under consideration.

- The panel further recommended that we:
  - Estimate the between-study variation to allow the reader to gauge the degree of heterogeneity present in our simulation scenarios.
  - Consider presenting the common notation in an analysis-of-variance format in addition to a regression one.
  - Consider presenting a table showing each meta-regression method by the parameters it estimates under what conditions.
  - Expand the simulation by allowing the treatment effects to vary by study, and by covariates; including realistically collinear study characteristics; and incorporating random effects.

We were unable to implement these recommendations within the scope of this project.

General Recommendations

The panel had further recommendations for the meta-regression user community:

- Measuring and incorporating heterogeneity in a meta-analysis is not sufficient. The panel recommends that meta-analysts investigate the causes of heterogeneity.
- With respect to meta-regression, a body of techniques for which the panel preferred the term "multilevel modeling," the panel saw the need for further software development. Perhaps more importantly, the panel saw the need for outreach, e.g. in the form of tutorials, to assist new users with learning how to conduct and interpret such analyses. Foremost in the advised strategies should be the use of regression diagnostics and graphics, especially given the limited degrees-of-freedom, high collinearity, and strong possibility for ecological bias in the meta-regression setting.
- Though much of the research that has already been conducted in the usual regression setting to determine how to assess model fit may be transferable to the meta-regression setting, the panel recommended further research in this area. For example, how does one judge whether a meta-regression modeling effort has been well-done? What should an analyst report in a meta-regression analysis, e.g., can guidelines be developed akin to the QUORUM statement?[2]

Recommendations Regarding Future Work

In the Southern California Evidence-Based Practice Center's role as technical support to the National Center for Complementary and Alternative Medicine, we are investigating methodological research topics. Our first topic is the subject of this report: meta-regression. We asked the panel to recommend what methodological research we should undertake in the coming year. We had originally proposed quality assessment of observational studies as our next topic. The panel's recommendation was:

- The panel understood the need for guidance regarding the assessment of quality of observational studies. The panel recommended that if work was to be done in this area, it focus on a specific clinical topic, e.g., a "case study," and empirically investigate the relationship between different quality attributes and treatment outcomes. The panel recommended against developing a global quality scale, and also did not advise considering observational study quality in general.

We agree with the panel that the development of a global quality scale for observational studies is premature. Further work must be done to understand meta-analysis of observational studies. We will consider the panel's recommendation regarding a case study approach.

Simulation Results

The simulation was a complete factorial experiment in that all levels of all simulation parameters appear in combination with all levels of all other simulation parameters without replication at any of the combinations. Rather than repeatedly running the simulation at a particular, usually randomly-drawn, combination of values, we have exhaustively run all combinations. We considered the option of running several replications at each of the design points. Given the purpose of the study we decided that covering a broader range and more

exhaustive combination of parameters would be more informative. One consequence of this approach is that we will need model-based error estimates. Therefore, we analyze the simulation results with analysis-of-variance (ANOVA) methods as described below.

Simulation Analysis

The analysis of the simulation results for each meta-regression method is an ANOVA with a dependent variable of bias, and the independent variables are the simulation parameters. The first decision was what level of interaction among the simulation parameters should be included. Using general F-tests, we considered the addition of all interactions of various orders in forward selection. For example, we compared a model with only main effects to a model that included all two-way interactions. Repeating this process, we compared the two-way interaction model to the three-way interaction model, and the three-way interaction model to the four-way interaction model. For all five meta-regression methods, the three-way interaction model was found to be adequate.

Using the three-way interaction models, we ranked the ANOVA effects for each method by their sums-of-squares. We denoted as practically important model terms those that had a sums-of-squares greater than or equal to 30 in *any* of the five method models. This bounding rule, or "practical criterion," is guaranteed to capture any contribution to bias of 10% or greater on average. Many more terms are statistically significant. The six terms that met this practical criterion and the methods for which they were important are shown in Table 4.

Computer Requirements

The process of simulating the person-level data and aggregating to the study level in SAS 8.0[23] required approximately 22 hours on a 550 MHz dual Pentium PC with 512 megabytes of RAM. Fitting the meta-regression models ranged from 21 minutes for the fixed effects models implemented in SAS 8.0[23] to approximately eight hours for the control rate models implemented in SPLUS[26] on a 700 MHz Pentium with one gigabyte of RAM.

Simulation Table Explanation

Tables 5 – 10 each present the simulation results for an interaction term selected as practically significant in the ANOVA analyses. In each table, we present the estimated bias in the estimate of the population mean treatment effect and the standard error of that estimated bias.

All five meta-regression methods appear in each table regardless of whether that particular method achieved practical significance for the interaction. This facilitates comparison among the methods. The bias is presented as a percentage for the case where γ_o is 0.6, the largest absolute value of the population mean treatment effect in the regression. For example, a –50.0 percent bias would indicate that a γ_o of 0.3 was estimated when the true value of γ_o was actually 0.6.

When presenting the results for a particular interaction term, we hold all omitted simulation parameters equal to their most neutral values. The footnote to each table reminds the reader what these values are. In general, we have selected these values to correspond to values least likely to introduce bias. For example, the largest number of studies, k = 30, is used as more bias exists with a smaller number of studies.

Simulation Tables

Table 5 presents the bias in the estimation of the population mean treatment effect as a function of the treatment effect of a person-level covariate (γ_3) and variability in sample sizes across studies (*tilt*). Note that for $\gamma_3 = 0$, the bias is relatively small. For nonzero values of γ_3, the bias is large when *tilt* = 0. Note that the methods with covariates are much less biased when a covariate is an important predictor, i.e., when γ_3 is nonzero. Methods without covariates can be substantially biased when a covariate is important and the sign of that bias depends on the sign of the covariate.

Table 6 presents the bias as a function of the within-study standard deviation of a person-level covariate (σ_{xi}) and the treatment effect of a person-level covariate (γ_3). Note when $\sigma_{xi} = 0$, the person-level covariate becomes in effect a study-level covariate and the bias monotonically increases in γ_3. When $\sigma_{xi} = 0.5$, the bias monotonically decreases in γ_3. This suggests that aggregating a person-level covariate to the study level can produce substantial bias. This result is consistent with the aggregation bias literature.

Table 7 presents the bias as a function of the treatment effect of a study-level covariate (γ_2) and the number of studies (k). In general, the bias is much larger for the smaller number of studies ($k = 10$). However, the direction of the bias for a small number of studies depends on the value of γ_2. With positive values of γ_2 associated with negative bias and vice versa. Although all the methods have difficulty with small number of studies, the magnitude and direction of the bias is influenced by an important study-level covariate.

Table 8 presents the bias as a function of a treatment effect that depends on the baseline rate (γ_1) and the baseline outcome rate (ϕ). For negative values of γ_1, ϕ is positively correlated with bias, and for positive values of γ_1, ϕ is negatively correlated with bias. This suggests that a nonzero control rate in methods that do not incorporate a control rate can substantially bias the estimate and the direction of that bias depends on the sign of the control rate (γ_1).

Table 9 presents the bias as a function of the treatment effect of a person-level covariate (γ_3) and the baseline outcome rate (ϕ). Although there are some slight variations in the bias as a function of ϕ, this table is dominated by the effect of γ_3 on the bias. For the models for which this interaction is significant, there is a strong negative correlation between γ_3 and the bias. Again, this effect is consistent with the aggregation bias literature.

Table 10 presents the bias as a function of the treatment effect of a study-level covariate (γ_2) and the baseline outcome rate (ϕ). The only model for which this interaction is significant is control rate meta-regression. As γ_2 increases, the relationship between ϕ and bias changes from positive correlation to no correlation, while its magnitude increases. This shows the importance of omitted study-level covariates in the presence of a nonzero control rate. Note that this table is

difficult to interpret as our simulation contained only positive correlations between control rate and the study-level covariate. In future work, we will include negative correlations.