This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

AHRQ is the lead Federal agency charged with supporting research designed to improve the quality of health care, reduce its cost, address patient safety and medical errors, and broaden access to essential services. AHRQ sponsors and conducts research that provides evidence-based information on health care outcomes; quality; and cost, use, and access. The information helps health care decisionmakers—patients and clinicians, health system leaders, and policymakers—make more informed decisions and improve the quality of health care services.

Technical Review
Number 8

Meta-regression Approaches: What, Why, When, and How?

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-97-0001

Prepared by:
Southern California–RAND Evidence-Based Practice Center, Santa Monica, CA

Sally C. Morton, Ph.D.
Task Order Director
John L. Adams, Ph.D.
Marika J. Suttorp, M.S.
Statisticians
Roberta Shanman, M.L.S.
Reference Librarian
Di Valentine, J.D.
Shannon Rhodes, M.F.A.
Assistants

Paul G. Shekelle, M.D., Ph.D.
Director, Southern California—RAND Evidence-Based Practice Center

Sally C. Morton, Ph.D.
Co-Director, Southern California—RAND Evidence-Based Practice Center and Senior Statistician

AHRQ Publication No. 04-0033
March 2004

Suggested Citation:

Morton SC, Adams JL, Suttorp MJ, Shekelle PG. Meta-regression Approaches: What, Why, When, and How? Technical Review 8 (Prepared by Southern California–RAND Evidence-based Practice Center, under Contract No 290-97-0001). AHRQ Publication No. 04-0033. Rockville, MD: Agency for Healthcare Research and Quality. March 2004.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-Based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. This report on *Meta-regression Approaches: What, Why, When and How?* was requested and funded by the National Center for Complementary and Alternative Medicine. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To bring the broadest range of experts into the development of evidence reports and health technology assessments, AHRQ encourages the EPCs to form partnerships and enter into collaborations with other medical and research organizations. The EPCs work with these partner organizations to ensure that the evidence reports and technology assessments they produce will become building blocks for health care quality improvement projects throughout the Nation. The reports undergo peer review prior to their release.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality.

We welcome written comments on this technical review. They may be sent to: Director, Center for Outcomes and Evidence, Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850.

Carolyn M. Clancy, M.D.
Director
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Acting Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Stephen E. Straus, M.D.
Director
National Center for Complementary and Alternative Medicine

Acknowledgments

We gratefully acknowledge the participation and constructive feedback provided by these expert panelists:

- Dr. Jesse Berlin, University of Pennsylvania School of Medicine
- Dr. William DuMouchel, AT&T Research
- Dr. Vic Hasselblad, Duke Clinical Research Institute
- Dr. Joseph Lau, Division of Clinical Care Research, New England Medical Center
- Dr. Thomas Louis, RAND, Washington, DC
- Dr. Martin McIntosh, Biostatistics Department, University of Washington and Fred Hutchinson Cancer Research Center
- Dr. Ingram Olkin, Department of Statistics, Stanford University
- Dr. Allan Sampson, Department of Statistics, University of Pittsburgh
- Dr. Chris Schmid, Tufts University School of Medicine, New England Medical Center

Dr. Jesse Berlin from the University of Pennsylvania School of Medicine and Dr. Thomas Louis from RAND (now at Johns Hopkins) reviewed the draft report. In addition, Southern California EPC staff not on the study team also reviewed the draft report. We gratefully acknowledge the assistance of these referees.

***Note**: Acknowledgments are made with the explicit statement that this does not constitute endorsement of the report.*

Structured Abstract

Objective. The broad objective of this report is to compare and contrast via simulation five meta-regression approaches that model the heterogeneity among study treatment effects: fixed effects with and without covariates; random effects with and without covariates; and control rate meta-regression.

Methodology. We conducted a systematic review of MEDLINE®, HealthSTAR, EMBASE, MANTIS, SciSearch®, Social SciSearch®, Allied and Complementary Medicine, the Current Index to Statistics, and the Methodology Register of the Cochrane Library from the inception of each database through March 2001 using the search terms "metaregress-" or "meta" within two words of "regress-." We supplemented these searches with articles identified by experts, and by searching the reference lists of all relevant articles found. We constructed a statistical notation generally applicable to different meta-regression methods. We convened a one-day panel of nine experts, and elicited their recommendations for the practice of meta-regression and implementation of our simulation study. We implemented a large-scale simulation to compare and contrast the five meta-regression techniques.

Main Results. We identified and categorized 85 publications on meta-regression. We presented scenarios for which meta-regression might be informative, and expressed the most common meta-regression models in a common notation. Our expert panel made several recommendations regarding the simulation parameters. The panel also identified the need for outreach by the methodological community to the user community in advising how to conduct, interpret, and present meta-regression analyses, including the development of software and diagnostic aids to assess models.

The simulation was a complete factorial design including all possible 7,776 combinations of the simulation parameters. The results were evaluated using an analysis-of-variance (ANOVA) model relating the simulation parameters to the bias in the estimation of the additive treatment effect. Across the five different meta-regression methods, six terms in a three-way ANOVA model were found to be practically important as they captured contributions to the bias of 10% or greater on average.

Conclusions. Our simulation results produced specific guidelines for meta-regression practitioners that may be summarized in the key message that the causes of heterogeneity should be explored via the inclusion of covariates at both the person level and study level. Based on our comparison of bias across approaches, either fixed effects or random effects methods can be used to support this exploration. In terms of future simulation research, we need to increase the variability in sample sizes, explore correlations between study outcome rates and covariates at both the study and person level, and evaluate within-study variation for person-level covariate(s). We now have in place a simulation methodology, a common notation, and a supportive panel of national experts to enable and guide our continued work in this area. The research presented in this report has already impacted the application of meta-regression in several alternative medicine settings, and improved our ability to synthesize and understand these therapies.

Contents

Tables

Appendixes for this report are provided electronically at http://www.ahrq.gov/clinic/epcindex.htm. Scroll through the topic list to select this report.

Summary

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," (Colditz GA, 1995) is one of the most important challenges facing a meta-analyst. 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.

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, e.g., of the pooled odds ratio, in each strata.
- Construct a random effects estimate across all studies. A random effects approach incorporates both within-study and between-study variability. We note that some argue that if heterogeneity exists among studies, a summary measure across those studies should not be provided.
- Fit a meta-regression model that explains the heterogeneity in terms of study-level covariates.

The broad objective of this report is to focus on this strategy of modeling the differences between studies, by comparing and contrasting several meta-regression methods.

Methodology

We conducted a systematic review of MEDLINE®, HealthSTAR, EMBASE, MANTIS, SciSearch®, Social SciSearch®, Allied and Complementary Medicine, the Current Index to Statistics, and the Methodology Register of the Cochrane Library through March 2001 using the search terms "metaregress-" or "meta" within two words of "regress-" in order to identify publications on meta-regression. We supplemented these searches with articles identified by experts, and by searching the reference lists of relevant articles.

Given the variety of meta-regression approaches available, our first analytic objective was to propose a common statistical framework, using the knowledge gained from the articles found via our systematic review, in which all meta-regression models could be expressed.

We implemented a simulation study to compare the different meta-regression modeling approaches. Simulation allowed us to set up a scenario (the "true" model), simulate data from that model, estimate parameters using various meta-regression models, and then compare the estimated parameters of each model in terms of bias.

We convened a one-day meeting of nine experts on heterogeneity in meta-analysis, and meta-regression. Prior to the meeting, the experts were sent background materials, including the preliminary parameters we were considering for our simulation study. The experts were also asked to suggest additional meta-regression publications. During the meeting, four of the experts presented half-hour talks spanning different types of meta-regression approaches.

The experts reached agreement on the parameters needed to complete the simulation, and additional analyses to conduct. The meeting was audio-taped and transcribed to assist in the preparation of this report.

Findings

The systematic review produced 85 publications relevant to meta-regression. We categorized the publications into seven categories based on the primary focus of the article. The first four categories were the main meta-regression methods: fixed effects models (4 publications); random effects models (11 publications); control rate models (9 publications); and Bayesian and/or hierarchical models (13 publications). 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 (19 publications). 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 to the meta-analyst (5 publications). Our seventh category consisted of examples (24 publications).

Using the knowledge gained from our review of the retrieved articles, we proposed a common statistical framework in which all meta-regression methods could be expressed. We restricted attention to dichotomous outcomes only. We discussed scenarios in which meta-regression might be informative, and presented the common meta-regression approaches using our proposed notation.

Our expert panel made several recommendations regarding the simulation parameters. The panel also generally identified the need for outreach by the methodological community to the user community in advising how to conduct, interpret, and present meta-regression analyses, including the development of software and diagnostic aids to assess models.

In our simulation study, we evaluated five meta-regression methods: fixed effects with and without covariates; random effects with and without covariates; and control rate meta-regression. The simulation was a complete factorial design including all possible 7,776 combinations of the simulation parameters.

We compared methods in terms of bias in the estimation of the additive treatment effect, which is the parameter typically estimated in meta-analyses. The results were evaluated using an analysis-of-variance (ANOVA) model relating the simulation parameters to the bias. Across the five different meta-regression methods, six terms in a three-way ANOVA model were found to be practically important as they captured contributions to the bias of 10% or greater on average.

Conclusions from a Statistical Perspective

Meta-regression methods will be increasingly used. Their attractiveness lies in their potential to explain differences between studies, thereby helping the clinician and decision-maker determine when, where, and for whom a treatment is beneficial. Our expert panel noted the usefulness and timeliness of this report.

Our panel had several general recommendations regarding meta-analysis and meta-regression. Foremost, the panel echoed the guidance given by others: measuring and incorporating heterogeneity in a meta-analysis is not sufficient. Meta-analysts should investigate and attempt to understand the causes of heterogeneity. The panel identified the need for outreach by the methodological community to the user community in advising how to conduct, interpret, and present meta-regression analyses, including the development of software and diagnostic aids to assess models.

The panel made several recommendations that we were able to include in this report. Some recommendations are delegated to future research. The panel also addressed the next methodological topic for the Southern California Evidence-Based Practice Center given our role as technical support to NCCAM. The panel recommended that if we undertake as our next methodological topic the quality assessment of observational studies, we focus on a specific clinical topic as a "case study." The panel recommended against developing a global scale, and also did not advise considering observational study quality in general.

Our simulation results produced the following guidelines for the meta-regression practitioner:

- In general, failure to incorporate important covariates at either the study or person level, can bias the results of a meta-analysis.
- Despite the importance of including covariates, a model that includes a covariate that is an aggregate of a person-level characteristic rather than a study characteristic can produce biased results. The trade-off between the biases of incorporating an aggregated covariate versus excluding it requires further exploration.
- If the control rate affects treatment, the meta-analysis should incorporate the control rate. However, control rate meta-regression is susceptible to bias via the correlation between the control rate and other omitted covariates. This suggests that extensions of control rate meta-regression to include other covariates may prove useful.
- As always, larger number of studies and larger number of patients per study can reduce bias with proper modeling.

In summary, our key message to practitioners is they should explore the causes of heterogeneity via the inclusion of covariates at both the person level and study level. Either fixed effects or random effects methods can be used to support this exploration. Note that our work presented in this report has not addressed confidence interval construction and statistical significance testing. Further work in this dimension may reveal differences between fixed and random effects approaches.

Conclusions from a Nonstatistical Perspective

Consider a meta-analysis of randomized controlled trials of a treatment to reduce heart disease mortality. Assume that the study-level variable aggregated from person-level data is the average disease severity, e.g., average blood pressure, among persons in each specific trial. Assume further that another study-level variable, whether the trial occurs in a hospital versus an outpatient setting, is also available. The control rate in this example is the rate of heart disease mortality in the control group in each trial.

Based on our simulation study, our first conclusion stated above is *"In general, failure to incorporate important covariates at either the study or person level, can bias the results of a meta-analysis."* Consider in our example the possibility that the treatment effect may not be the same for mild and severe cases. The treatment may have no effect for patients with very high blood pressure (high severity), while it has a strong effect for patients with mildly elevated blood pressure (low severity). Further we would anticipate that a hospital-based trial may in general accumulate sicker patients than an outpatient-based trial. Failure to account for the trial to trial variation in patient severity may lead to the incorrect conclusion that the treatment is less effective in hospital settings than in outpatient settings.

Our second conclusion is *"Despite the importance of including covariates, a model that includes a covariate that is an aggregate of a person-level characteristic rather than a study characteristic can produce biased results. The trade-off between the biases of incorporating an aggregated covariate versus excluding it requires further exploration."* In many statistical problems, it has been observed that using average quantities in place of person-specific quantities can lead to biased results and erroneous conclusions. Many of these problems fall under the general label of "ecological fallacy." Although further research is required, it may be the case that using aggregated variables in meta-analysis, such as average blood pressure, is a useful first step in understanding how the treatment effect differs across different types of patients and settings. However, we must always bear in mind that such conclusions should be considered exploratory rather than confirmatory. When interesting findings are discovered in this manner, person-level data from large trials may be required for confirmation.

Our third conclusion is "*If the control rate affects treatment, the meta-analysis should incorporate the control rate. However, control rate meta-regression is susceptible to bias via the correlation between the control rate and other omitted covariates. This suggests that extensions of control rate meta-regression to include other covariates may prove useful.*" We think that the control rate meta-regression method holds great promise to help us understand the relationship between treatment effect, illness severity and differences in trial protocols. However, the success of control rate meta-regression by its very nature begs the question of what other covariates might help us explore this heterogeneity. With time, we believe improved methods will be available to address these questions. In the meantime, where control-rate meta-regression differs from a simpler meta-regression, policy conclusions should be made tentatively and with caution.

Our fourth conclusion is *"As always, a larger number of studies and larger number of patients per study can reduce bias with proper modeling."* The user of meta-regression should remember that the degrees of freedom he or she has to understand study characteristics are

severely limited by the number of studies. Larger individual studies will always be a scientific gold standard that cannot be completely replaced by the meta-analysis of smaller studies.

Alternative Medicine Meta-analysis

NCCAM and AHRQ were especially interested in the topic of heterogeneity in meta-analysis given its relevance to alternative medicine literature. The challenges faced in synthesizing this literature are very different from those faced in for example the cardiovascular literature. The latter consists mainly of large randomized controlled trials. In contrast, the alternative medicine literature consists mainly of small trials, and nonrandomized studies. In addition to study design heterogeneity, the interventions are heterogeneous as well as the patient populations. Thus, methods for dealing with heterogeneity are particularly relevant.

Staff of the Southern California Evidence-Based Practice Center have applied meta-regression in the alternative medicine setting. In a systematic review of the evidence on ephedra and ephedrine, we used meta-regression to compare weight loss efficacy between groups receiving ephedrine; ephedrine plus caffeine; and ephedra plus herbs containing caffeine versus placebo. In a meta-analysis of spinal manipulation, we developed meta-regression models for acute and chronic back pain patients predicting short-term and long-term pain and function. These models took the unique approach of denoting the spinal manipulation group as the comparison group against which all other treatments, such as sham or physical therapy, were compared. The usual strategy would be to compare versus placebo or control. The knowledge gained via the research presented in this report impacted the application of meta-regression to these alternative medicine questions, and improved our ability to synthesize and understand these therapies.

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 2*k* 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.

Chapter 2. Methodology

Systematic Review to Identify Meta-regression Publications

We searched the following library databases:

- MEDLINE® 1966 - March, 2001
- HealthSTAR 1975 - March, 2001
- EMBASE 1974 - March, 2001
- MANTIS 1880 - 2000
- SciSearch® 1990 - March, 2001
- Social SciSearch® 1974 - March, 2001
- Allied and Complementary Medicine 1985 – 2000

The search terms were "metaregress-" or "meta" within two words of "regress-" (the latter also picks up the hyphenated form of the term, i.e., “meta-regression”). We used the same terms to search the Current Index to Statistics (1974 –1999). We also searched the Methodology Register of the Cochrane Library (version 1, 2001) using the keywords “meta-regression,” “metaregression,” or “regression.” We supplemented these searches with articles from the Southern California Evidence-Based Practice Center’s methodological article database that contains over 500 articles, and canvassed experts, including our expert panel and draft report referees (described below), for additional references. We also searched the reference lists in all relevant articles for additional publications.

A single reviewer (Morton) reviewed all title lists for relevance. The full text for all relevant articles was obtained.

Common Statistical Notation Objective

Given the variety of meta-regression approaches available, our first analytic objective was to propose a common statistical framework using the knowledge gained from the articles found via our systematic review, in which all meta-regression models could be expressed.

Simulation Approach

We decided to implement a simulation study to compare the different meta-regression modeling approaches. Simulation allows us to set up a scenario (the “true” model), simulate data from that model, estimate parameters using various meta-regression models, and then compare the estimated parameters of each model with the true model (bias properties). We defer an evaluation of coverage for future research.

Our first question was: What methods work best under what circumstances? Sub-questions in this domain were:

- What if the treatment effect depends on disease severity?
- What if the studies have a wide range of population risks (e.g. control group mortality rates)?
- What if there are relatively few studies?

Our second question was: What methods are most sensitive to assumptions? Sub-questions in this domain were:

- Do random effects models "protect" against omitted variables?
- If we are uncertain of the factors that modify treatment effects, what is the "safest" method to use?

Our approach was to hypothesize a person-level model, and generate data according to that model. We used this approach as medical intuition applies at the patient level, and treatment is applied at the patient level. We then aggregated the person-level data to the study level. This approach capitalizes on the data aggregation literature.[19] Aggregation in this manner may allow us to work out aggregation bias properties in future research.

Panel Methodology

We convened a one-day meeting of nine nationally-recognized experts on heterogeneity, and meta-regression. Prior to the meeting, the experts were sent a meeting agenda, goals and objectives including key questions, a document containing our common notation, a discussion of our preliminary simulation, and our preliminary bibliography. We asked the experts to suggest additional meta-regression references. The list of experts and items from the meeting are shown in Appendix A.

During the meeting, four of the experts presented half-hour talks. The topics chosen spanned the different types of meta-regression approaches available:

- Meta-analysis of multi-treatment studies[20] (presented by Dr. Vic Hasselblad)
- Control rate meta-regression models[15,21](presented by Dr. Chris Schmid)
- Bayesian meta-analysis[17] (presented by Dr. Thomas Louis)
- Methodological challenges in meta-regression[10, 11] (presented by Dr. Jesse Berlin)

The common notation and the preliminary simulation results were discussed in detail. During the last part of the meeting, the attendees broke into smaller groups to discuss the three topics of heterogeneity; meta-regression; and the simulation. Each group then reported on their group discussion to the entire panel. The experts reached agreement on the parameters needed to complete the simulation, and additional analyses to conduct. In addition, the experts provided advice on the second methodological topic the Southern California Evidence-Based Practice Center should address. Quality assessment for nonrandomized studies had been proposed. The meeting was audio-taped and transcribed to assist in the preparation of this report.

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

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

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

$$x_{ij} \sim N(\mu_i, \sigma^2_{xi})$$

where

$$\Sigma_{\phi z \mu} = \begin{pmatrix}\sigma^2_\phi & \sigma_{\phi z} & \sigma_{\phi \mu}\\ \sigma_{\phi z} & \sigma^2_z & \sigma_{z\mu}\\ \sigma_{\phi \mu} & \sigma_{z\mu} & \sigma^2_\mu\end{pmatrix} \qquad (4)$$

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: $\upsilon_i \sim N(0, \sigma^2_\upsilon)$.

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

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

$\tau_{ij} = 0$ 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:

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

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

$$y_i = \log(OR_i) = \tau + \upsilon_i + \varepsilon_i$$
$$\upsilon_i \sim N(0, \sigma_\upsilon^2)$$
$$\varepsilon_i \sim N(0, \sigma_i^2)$$

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]

$$y_i = \log(OR_i) = \gamma_0 + \gamma_2 z_i + \upsilon_i + \varepsilon_i$$
$$\upsilon_i \sim N(0, \sigma_\upsilon^2)$$
$$\varepsilon_i \sim N(0, \sigma_i^2)$$

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]

$$y_i = \log(OR_i) = \gamma_0 + \gamma_1 \phi_i + \upsilon_i + \varepsilon_i$$
$$\upsilon_i \sim N(0, \sigma_\upsilon^2)$$
$$\varepsilon_i \sim N(0, \sigma_i^2)$$

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):

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

Chapter 4. Conclusions

Meta-regression methods will be increasingly used in the future. Their attractiveness lies in their potential to explain differences between studies, thereby helping the clinician and decision-maker determine when, where, and for whom a treatment is beneficial. Our experts noted the usefulness and timeliness of this report.

Our panel had several general recommendations regarding meta-analysis and meta-regression. Foremost, the panel echoed the guidance given by others: measuring and incorporating heterogeneity in a meta-analysis is not sufficient. Meta-analysts should investigate and attempt to understand the causes of heterogeneity, and meta-regression is an appealing technique to use. The panel saw the need for outreach by the methodological community to the user community in advising how to conduct, interpret, and present meta-regression analyses. Further software development should include model diagnostics, and graphical approaches.

The panel made several recommendations that we were able to include in this report. Some recommendations are delegated to future research. The panel also addressed the next methodological topic for the Southern California Evidence-Based Practice Center given our role as technical support to the NCCAM. The panel recommended that if we undertake as our next methodological topic the quality assessment of observational studies, we focus on a specific clinical topic as a "case study." The panel recommended against developing a global scale, and also did not advise considering observational study quality in general.

Our simulation compared five meta-regression methods ranging from fixed effects to random effects, either including or not including covariates, to the control rate method. Our simulation results produced the following guidelines for the meta-regression practitioner:

- In general, failure to incorporate important covariates at either the study or person level, can bias the results of a meta-analysis.
- Despite the importance of including covariates, a model that includes a covariate that is an aggregate of a person-level characteristic rather than a study characteristic can produce biased results. The trade-off between the biases of incorporating an aggregated covariate versus excluding it requires further exploration.
- If the control rate affects treatment, the meta-analysis should incorporate the control rate. However, control rate meta-regression is susceptible to bias via the correlation between the control rate and other omitted covariates. This suggests that extensions of control rate meta-regression to include other covariates may prove useful.
- As always, larger number of studies and larger number of patients per study can reduce bias with proper modeling.

In summary, our key message to practitioners is they should explore the causes of heterogeneity via the inclusion of covariates at both the person level and study level. Either fixed effects or random effects methods can be used to support this exploration. Note that our work presented in this report has not addressed confidence interval construction and statistical significance testing. Further work in this dimension may reveal differences between fixed and random effects approaches.

The research to date has revealed promising directions for expanding the simulation. In particular, we need to further implement the expert panel's advice that more variability in sample sizes be included. Furthermore, correlations, both negative and positive, between study baseline outcome rates and covariates at both the study and person level need to be explored over an expanded range of values. Similarly, a broader range of within-study variation for the person-level covariate(s) should be evaluated. The ability of the various meta-regression methods to identify which covariates, study or person-level, are able to predict the treatment effect, and how well they can do so, should also be a question addressed by an expanded simulation study. We now have in place a simulation methodology, a common notation, and a supportive panel of national experts to enable and guide our continued work in this area.

NCCAM and AHRQ were especially interested in the topic of heterogeneity in meta-analysis given its relevance to alternative medicine literature. The challenges faced in synthesizing this literature are very different from those faced in for example the cardiovascular literature. The latter consists mainly of large randomized controlled trials. In contrast, the alternative medicine literature consists mainly of small trials, and nonrandomized studies. In addition to study design heterogeneity, the interventions are heterogeneous as well as the patient populations. Thus, methods for dealing with heterogeneity are particularly relevant.

Staff of the Southern California Evidence-Based Practice Center have applied meta-regression in the alternative medicine setting. In systematic review of the evidence on ephedra and ephedrine,[27, 28] we used meta-regression to compare weight loss efficacy between groups receiving ephedrine; ephedrine plus caffeine; and ephedra plus herbs containing caffeine versus placebo. In a meta-analysis of spinal manipulation,[29] we developed meta-regression models for acute and chronic back pain patients predicting short-term and long-term pain and function. These models took the unique approach of denoting the spinal manipulation group as the comparison group against which all other treatments, such as sham or physical therapy, were compared. The usual strategy would be to compare versus placebo or control. The knowledge gained via the research presented in this report impacted the application of meta-regression to these alternative medicine questions, and improved our ability to synthesize and understand these therapies.

References

1. Colditz GA, Burdick E, Mosteller F. Heterogeneity in meta-analysis of data from epidemiologic studies: a commentary. Am J Epidemiol 1995;142(4):371-382.

2. Moher D, Chalmers TC, Cook DJ, Eastwood S, Olkin I, Drummond R, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Lancet 1999;354(9193):1896-1900.

3. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000;283(15):2008-2012.

4. Thompson SG. Controversies in meta-analysis: the case of the trials of serum cholesterol reduction. Stat Methods Med Res 1993;2(2):173-192.

5. Hedges LV, Olkin I. Statistical methods for meta-analysis. San Diego, CA: Academic Press, Inc.; 1985.

6. Hardy RJ, Thompson SG. Detecting and describing heterogeneity in meta-analysis. Stat Med 1998;17(8):841-856.

7. Laird NM, Mosteller F. Some statistical methods for combining experimental results. Int J Technol Assess Health Care 1990;6(1):5-30.

8. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7(3):177-188.

9. Morgenstern H. Uses of ecologic analysis in epidemiologic research. Am J Public Health 1982;72(12):1336-1344.

10. Berlin JA, Antman EM. Advantages and limitations of metaanalytic regressions of clinical trials data. Online J Curr Clin Trials 1994;Doc No 134.

11. Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI. Individual patient versus group-level data metaregressions for the investigation of subgroup effects: ecological bias rears its ugly head. Stat Med 2002;21(2): 371-387.

12. Stuck AE, Siu AL, Wieland GD, Adams J, Rubenstein LZ. Comprehensive geriatric assessment: a meta-analysis of controlled trials. Lancet 1993;342(8878):1032-1036.

13. Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random-effects regression model for meta-analysis. Stat Med 1995;14(4):395-411.

14. McIntosh MW. The population risk as an explanatory variable in research synthesis of clinical trials. Stat Med 1996;15(16):1713-1728.

15. Schmid CH, Lau J, McIntosh MW, Cappelleri JC. An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. Stat Med 1998;17(17):1923-1942.

16. DuMouchel WH. Bayesian Meta-analysis. In: Berry DA. Statistical Methodology in the Pharmaceutical Sciences. New York: Marcel Dekker; 1999.

17. Louis TA, Zelterman D. Bayesian Approaches to Research Synthesis. In: Cooper H, Hedges LV. The Handbook of Research Synthesis. New York: Russell Sage Foundation; 2000. p. 411-422.

18. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. Stat Med 1995;14(24):2685-2699.

19. Theil H. Principles of Econometrics. New York: John Wiley and Sons; 1971.

20. Hasselblad V. Meta-analysis of multitreatment studies. Med Decis Making 1998;18(1):37-43.

21. Schmid CH. Exploring heterogeneity in randomized trials via meta-analysis. Drug Inf J 1999;33:211-224.

22. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1958;22(4):719-748.

23. SAS/STAT [computer program]. Version 8. Cary, NC: SAS Institute Inc.; 1999.

24. Stata Statistical Software: Release 6.0 [computer program]. Version 6.0. College Station, TX: Stata Corporation; 1999.

25. Sharp SJ. Meta-analysis regression. STATA Technical Bulletin 1998;42_:16-22.

26. S-PLUS [computer program]. Version 6.0. Seattle, WA: Data Analysis Division, MathSoft Inc.; 2000.

27. Shekelle PG, Morton SC, Maglione M, et al. Ephedra and ephedrine for weight loss and athletic performance enhancement: clinical efficacy and side effects. Evidence Report/Technology Assessment No. 76 (Prepared by the Southern California Evidence-Based Practice Center under Contract No. 290-97-0001). AHRQ Publication No. 03-E022. Rockville, MD: Agency for Healthcare Research and Quality. March 2003.

28. Shekelle PG, Hardy ML, Morton SC, Maglione M, Mojica WA, Suttorp MJ, et al. Efficacy and safety of ephedra and ephedrine for weight loss and athletic performance: a meta-analysis. JAMA 2003;289(12):1537-45.

29. Assendelft WJ, Morton SC, Yu EI, Suttorp MJ, Shekelle PG. Spinal manipulative therapy for low back pain. Ann Intern Med 2003;138(11):871-881.

Table 1. Distribution of Identified Meta-regression Publications

Publication Category	Number of Articles
Fixed Effects Models	4
Random Effects Models	11
Control Rate Models	9
Bayesian and/or Hierarchical Models	13
Overviews and Challenges	19
Multiple Treatments and/or Outcomes	5
Examples	24
Total	85

Table 2. Simulation Parameters

Parameter	Meaning	Simulation Values			Range of Odds Ratios
ϕ	baseline outcome rate	-0.6	0.0	0.6	0.55 to 1.82
γ_0	additive treatment effect	0.0	0.6		1 to 1.82
γ_1	treatment effect that depends on baseline rate	-0.6	0.0	0.6	0.55 to 1.82
γ_2	treatment effect of a study-level covariate	-0.6	0.0	0.6	0.55 to 1.82
γ_3	treatment effect of a person-level covariate	-0.6	0.0	0.6	0.55 to 1.82
σ_{xi}	within-study standard deviation of a person-level covariate	0.0	0.5		not applicable
$\sigma_{\upsilon i}$	a study-level random effect	0.0	0.5		not applicable
ρ	the correlation between baseline, study covariate and person covariate	0.0	0.3	0.6	not applicable
k	Number of studies	3	10	30	not applicable
tilt	Variability in sample sizes across studies	0	1		not applicable

Table 3. Example of Data Layout for Logistic Meta-regression

Study	Treatment	Outcome	Number of cases (weight)	Study-Level Covariate	Arm-Level Covariate
1	0	0	25	40	45
1	0	1	25	40	45
1	1	0	20	40	35
1	1	1	30	40	35
2	0	0	40	38	39
… and so on					

Table 4. Simulation Results Tables

Table Number	Interaction	Meaning	Method(s) for which this interaction was practically important	Sum-of-Squares
5	γ_3 by *tilt*	treatment effect of a person-level covariate by variability in sample sizes across studies	Fixed Effects Control Rate Random Effects	88.55 55.99 47.17
6	σ_{xi} by γ_3	within-study standard deviation of a person-level covariate by treatment effect of a person-level covariate	Fixed Effects + Covariates	58.64
7	γ_2 by k	treatment effect of a study-level covariate by number of studies	Random Effects Control Rate Fixed Effects	47.08 40.70 31.33
8	γ_1 by ϕ	treatment effect that depends on baseline rate by baseline outcome rate	Random Effects + Covariates Fixed Effects Fixed Effects + Covariates	44.24 43.69 42.50
9	γ_3 by ϕ	treatment effect of a person-level covariate by baseline outcome rate	Control Rate Random Effects Fixed Effects	36.18 32.59 30.02
10	γ_2 by ϕ	treatment effect of a study-level covariate by baseline outcome rate	Control Rate	33.69

Table 5. Percent bias in γ_o associated with the γ_3 by *tilt* interaction

		γ_3 = -0.6		γ_3 = 0		γ_3 = 0.6	
Method	**Std. Error**	*tilt* = 0	*tilt* = 1	*tilt* = 0	*tilt* = 1	*tilt* = 0	*tilt* = 1
Fixed Effects[a]	1.9	21.8	-14.1	1.6	-3.3	-30.1	-6.2
Fixed Effects + Covariates	5.0	-9.7	5.8	2.2	-2.1	6.9	-4.7
Random Effects[a]	1.6	23.2	-9.9	2.3	-7.6	-18.1	-3.9
Random Effects + Covariates	1.9	-3.3	-11.8	3.4	-7.1	6.8	NR[b]
Control Rate[a]	1.6	23.6	-9.9	3.8	-8.4	-17.3	-2.7

The following parameters are held constant: ϕ = 0, γ_0 = 0.6, γ_1 = 0, γ_2 = 0, σ_{xi} = 0, σ_{vi} = 0, ρ = 0, k = 30.

[a] This table corresponds to an interaction that was practically important for these methods.

[b] NR=not reportable. This method does not converge for this combination of simulation parameters.

Table 6. Percent bias in γ_o associated with the σ_{xi} by γ_3 interaction

		σ_{xi} = 0			σ_{xi} = 0.5		
Method	**Std. Error**	γ_3 = -0.6	γ_3 = 0	γ_3 = 0.6	γ_3 = -0.6	γ_3 = 0	γ_3 = 0.6
Fixed Effects	1.9	21.8	1.6	-30.1	19.5	0.9	-32.8
Fixed Effects + Covariates [a]	5.0	-9.7	2.2	6.9	58.0	-10.2	-8.3
Random Effects	1.6	23.2	2.3	-18.1	20.9	1.8	-21.2
Random Effects + Covariates	1.9	-3.3	3.4	6.8	-5.2	2.2	1.4
Control Rate	1.6	23.6	3.8	-17.3	21.2	3.2	-20.0

The following parameters are held constant: ϕ = 0, γ_0 = 0.6, γ_1 = 0, γ_2 = 0, σ_{vi} = 0, ρ = 0, k = 30, *tilt* = 0.

[a] This table corresponds to an interaction that was practically important for these methods.

Table 7. Percent bias in γ_o associated with the γ_2 by k interaction

Method	Std. Error	γ_2 = -0.6		γ_2 = 0		γ_2 = 0.6	
		k = 10	k = 30	k = 10	k = 30	k = 10	k = 30
Fixed Effects [a]	1.9	38.9	-8.1	14.9	1.6	-25.3	2.3
Fixed Effects + Covariates	5.0	35.2	9.8	34.3	2.2	32.2	-2.6
Random Effects [a]	1.6	50.8	0.6	19.5	2.3	-11.7	9.8
Random Effects + Covariates	1.9	39.3	8.3	39.7	3.4	33.0	-2.3
Control Rate [a]	1.6	49.7	1.7	20.9	3.8	-10.6	9.4

The following parameters are held constant: ϕ = 0, γ_0 = 0.6, γ_1 = 0, γ_3 = 0, σ_{xi} = 0, σ_{vi} = 0, ρ = 0, *tilt* = 0.

[a] This table corresponds to an interaction that was practically important for these methods.

Table 8. Percent bias in γ_o associated with the γ_1 by ϕ interaction

Method	Std. Error	γ_1 = -0.6			γ_1 = 0			γ_1 = 0.6		
		ϕ = -0.6	ϕ = 0	ϕ = 0.6	ϕ = -0.6	ϕ = 0	ϕ = 0.6	ϕ = -0.6	ϕ = 0	ϕ = 0.6
Fixed Effects[a]	1.9	-19.0	-6.2	14.4	-2.5	1.6	7.7	26.4	5.4	-8.5
Fixed Effects + Covariates[a]	5.0	-33.4	-7.1	45.0	-9.7	2.2	23.6	30.3	1.9	-15.0
Random Effects	1.6	-12.5	-5.2	8.0	-3.6	2.3	6.0	25.6	5.2	-3.8
Random Effects + Covariates[a]	1.9	-16.3	-0.8	21.2	-3.1	3.4	7.9	29.4	1.8	-13.6
Control Rate	1.6	-8.8	-2.9	6.3	-2.8	3.8	7.9	12.7	-0.6	6.8

The following parameters are held constant: γ_0 = 0.6, γ_2 = 0, γ_3 = 0, σ_{xi} = 0, σ_{vi} = 0, ρ = 0, k = 30, *tilt* = 0.

[a] This table corresponds to an interaction that was practically important for these methods.

Table 9. Percent bias in γ_o associated with the γ_3 by ϕ interaction

		γ_3 = -0.6			γ_3 = 0			γ_3 = 0.6		
Method	**Std. Error**	ϕ = -0.6	ϕ = 0	ϕ = 0.6	ϕ = -0.6	ϕ = 0	ϕ = 0.6	ϕ = -0.6	ϕ = 0	ϕ = 0.6
Fixed Effects[a]	1.9	19.6	21.8	23.6	-2.5	1.6	7.7	-23.5	-30.1	-28.4
Fixed Effects + Covariates	5.0	-25.4	-9.7	21.2	-9.7	2.2	23.6	5.5	6.9	7.2
Random Effects[a]	1.6	18.6	23.2	24.9	-3.6	2.3	6.0	-21.6	-18.1	-17.7
Random Effects + Covariates	1.9	-10.6	-3.3	4.0	-3.1	3.4	7.9	3.2	6.8	9.8
Control Rate[a]	1.6	18.9	23.6	25.4	-2.8	3.8	7.9	-21.1	-17.3	-16.9

The following parameters are held constant: γ_0 = 0.6, γ_1 = 0, γ_2 = 0, σ_{xi} = 0, σ_{vi} = 0, ρ = 0, k = 30, *tilt* = 0.

[a] This table corresponds to an interaction that was practically important for these methods.

Table 10. Percent bias in γ_o associated with the γ_2 by ϕ interaction

		γ_2 = -0.6			γ_2 = 0			γ_2 = 0.6		
Method	**Std. Error**	ϕ = -0.6	ϕ = 0	ϕ = 0.6	ϕ = -0.6	ϕ = 0	ϕ = 0.6	ϕ = -0.6	ϕ = 0	ϕ = 0.6
Fixed Effects	1.9	-8.7	-8.1	-7.8	-2.5	1.6	7.7	12.7	2.3	3.7
Fixed Effects + Covariates	5.0	-8.4	9.8	32.7	-9.7	2.2	23.6	-13.7	-2.6	22.0
Random Effects	1.6	-8.2	0.6	1.5	-3.6	2.3	6.0	11.2	9.8	13.2
Random Effects + Covariates	1.9	-2.9	8.3	12.5	-3.1	3.4	7.9	-4.1	-2.3	6.8
Control Rate[a]	1.6	-7.6	1.7	2.4	-2.8	3.8	7.9	10.9	9.4	12.7

The following parameters are held constant: γ_0 = 0.6, γ_1 = 0, γ_3 = 0, σ_{xi} = 0, σ_{vi} = 0, ρ = 0, k = 30, *tilt* = 0.

[a] This table corresponds to an interaction that was practically important for these methods.

Bibliography

Aitkin M. Meta-analysis by random effect modelling in generalized linear models. Stat Med 1999 Sep;18(17-18):2343-51.

Allison DB, Mentore JL, Heo M, et al. Antipsychotic-induced weight gain: a comprehensive research synthesis. Am J Psychiatry 1999 Nov;156(11):1686-96.

Arends LR, Hoes AW, Lubsen J, et al. Baseline risk as predictor of treatment benefit: three clinical meta-re-analyses. Stat Med 2000 Dec;19(24):3497-518.

Barza M, Ioannidis JP, Cappelleri JC, et al. Single or multiple daily doses of aminoglycosides: a meta-analysis. BMJ 1996 Feb;312(7027):338-45.

Becker BJ, Schram CM. Examining Explanatory Models Through Research Synthesis. In: Cooper H, Hedges LV, editors. The Handbook of Research Synthesis. New York: Russell Sage Foundation; 1994. p. 357-81.

Berkey CS, Hoaglin DC, Mosteller F, et al. A random-effects regression model for meta-analysis. Stat Med 1995 Feb;14(4):395-411.

Berkey CS, Anderson JJ, Hoaglin DC. Multiple-outcome meta-analysis of clinical trials. Stat Med 1996 Mar;15(5):537-57.

Berkey CS, Hoaglin DC, Antczak-Bouckoms A, et al. Meta-analysis of multiple outcomes by regression with random effects. Stat Med 1998 Nov;17(22):2537-50.

Berlin JA, Longnecker MP, Greenland S. Meta-analysis of Epidemiologic Dose-Response Data. Epidemiology 1993 May;4(3):218-28.

Berlin JA, Antman EM. Advantages and limitations of meta-analytic regressions of clinical trials data. Online J Curr Clin Trials 1994 Jun;Doc No 134.

Berlin JA, Santanna J, Schmid CH, et al. Individual Patient versus Group-level Data Meta-regressions for the Investigation of Subgroup Effects: Ecological Bias Rears its Ugly Head. Stat Med 2001. In press.

Bucher HC, Griffith LE, Guyatt GH. Systematic review on the risk and benefit of different cholesterol-lowering interventions. Arterioscler Thromb Vasc Biol 1999 Feb;19(2):187-95.

Carroll RJ, Stefanski LA. Measurement error, instrumental variables and corrections for attenuation with applications to meta-analyses. Stat Med 1994 Jun;13(12):1265-82.

Cook RJ, Walter SD. A logistic model for trend in 2 X 2 X K tables with applications to meta-analyses. Biometrics 1997 Mar;53(1):352-7.

Cooper H, Hedges LV, editors. The handbook of research synthesis. New York, NY: Russell Sage Foundation; 1994.

Corrao G, Bagnardi V, Zambon A, et al. Exploring the dose-response relationship between alcohol consumption and the risk of several alcohol-related conditions: a meta-analysis. Addiction 1999 Oct;94(10):1551-73.

Daniels MJ, Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. Stat Med 1997 Sep;16(17):1965-82.

DuMouchel WH, Herbert D. Combining information from multiple dose/time/Fractionation studies (Bayesian hierarchical meta-analysis revisited). Proceedings of the 4th International Conf. on Time, Dose and Fractionation in Radiation Oncology; 1993; Madison, WI.

DuMouchel WH, Harris JE. Bayes Methods for Combining the Results of Cancer Studies in Humans and Other Species. JASA 1983 Jun;78(382):293-308.

DuMouchel WH, Groer PG. A Bayesian methodology for combining radiation studies. 1988. p. 348-66.

DuMouchel WH, Groer PG. A Bayesian methodology for scaling radiation studies from animals to man. Health Phys 1989;57 Suppl 1:411-8.

DuMouchel WH, Waternaux C, Kinney D. Hierarchical Bayesian Linear Models for Assessing the Effect of Extreme Cold Weather on Schizophrenic Births. In: Berry DA, Stangl DK, editors. Bayesian Biostatistics. 1996.

DuMouchel WH. Predictive Cross-Validation of Bayesian Meta-analyses. In: Bernardo JM, Berger JO, Dawid AP, et al., editors. Bayesian Statistics 5. Oxford University Press; 1996. p. 107-27.

DuMouchel WH. Repeated Measures Meta-analysis. Proc. 51st Session International Statistical Institute; 1997; Istanbul, Turkey.

DuMouchel WH, Normand SL. Computer-modeling and Graphical Strategies for Meta-analysis. In: Stangl DK, Berry DB, editors. Meta-analysis in Medicine and Health Policy. 2000. p. 127-78.

Egger M, Davey Smith G, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. Cornwall, UK: BMJ; 2001.

Fay MP, Graubard BI, Freedman LS, et al. Conditional logistic regression with sandwich estimators: application to a meta-analysis. Biometrics 1998 Mar;54(1):195-208.

Freemantle N, Cleland J, Young P, et al. Beta Blockade after myocardial infarction: systematic review and meta regression analysis. BMJ 1999 Jun;318(7200):1730-7.

Freemantle N, Anderson IM, Young P. Predictive value of pharmacological activity for the relative efficacy of antidepressant drugs. Br J Psychiatry 2000 Oct;177:292-302.

Frost C, Clarke R, Beacon H. Use of hierarchical models for meta-analysis: experience in the metabolic ward studies of diet and blood cholesterol. Stat Med 1999 Jul;18(13):1657-76.

Gareen IF, Greenland S, Horgenstern H. Intrauterine Devices and Pelvic Inflammatory Disease: Meta-analyses of Published Studies, 1974-1990. Epidemiology 2000 Sep;11(5):589-97.

Geddes J, Freemantle N, Harrison P, et al. Atypical antipyschotics in the treatment of schizophrenia: systematic overview and meta-regression analysis. BMJ 2000 Dec;321(7273):1371-6.

Gleser LJ, Olkin I. Stochastically Dependent Effect Sizes. In: Cooper H, Olkin I, editors. The Handbook of Research Synthesis. New York: Russell Sage Foundation; 1994. p. 339-55.

Goldstein H, Yang M. Meta-analysis using multi-level models with an application to the study of class size effects. Applied Statistics 2000 Apr;49(3):399-412.

Greenland S, Robins J. Invited commentary: ecologic studies--biases, misconceptions, and counterexamples. Am J Epidemiol 1994 Apr;139(8):747-60.

Greenwood CMT, Midgley JP, Matthews AG, et al. Statistical Issues in a Metaregression Analysis of Randomized Trials: Impact on the Dietary Sodium Intake and Blood Pressure Relationship. Biometrics 1999 Jun;55:630.

Hasselblad V. Meta-analysis of multitreatment studies. Med Decis Making 1998 Jan;18(1):37-43.

Hedges LV. A random effects model for effect sizes. Psychol Bull 1983;93:388-95.

Hedges LV, Olkin I. Regression models in research synthesis. The American Statistician 1983;37:137-40.

Hedges LV, Olkin I. Statistical methods for meta-analysis. San Diego, CA: Academic Press, Inc.; 1985.

Hedges LV, Vevea JL. Fixed and random-effects models in meta-analysis. Psychological Methods 1998;3:486-504.

Houghton T, Freemantle N, Cleland JG. Are beta-blockers effective in patients who develop heart failure soon after myocardial infarction? A meta-regression analysis of randomised trials. Eur J Heart Fail 2000 Sep;2(3):333-40.

Ioannidis JP, Cappelleri JC, Lau J, et al. Early or deferred zidovudine therapy in HIV-infected patients without an AIDS-defining illness. Ann Intern Med 1995 Jun;122(11):856-66.

Ioannidis JP, Lau J. The impact of high-risk patients on the results of clinical trials. J Clin Epidemiol 1997 Oct;50(10):1089-98.

Ioannidis JP, Lau J. Pooling research results: benefits and limitations of meta-analysis. Jt Comm J Qual Improv 1999 Sep;25(9):462-9.

Johnson ES, Lanes SF, Wentworth CE, et al. A metaregression analysis of the dose-response effect of aspirin on stroke. Arch Intern Med 1999 Jun;159(11):1248-53.

Jones DR. Meta-analysis of observational epidemiological studies: a review. J R Soc Med 1992 Mar;85(3):165-8.

Kasiske BL, Kalil RS, Ma JZ, et al. Effect of antihypertensive therapy on the kidney in patients with diabetes: a meta-regression analysis. Ann Intern Med 1993 Jan;118(2):129-38.

Larholt KM, Gelber RD. Personal communication to Morton SC. 2000. Heterogeneity in meta-analysis: A simulation study of fixed and random effect methods. Available from Dr. AR Sampson.

Larholt KM, Tsiatis AA, Gelber RD. Personal communication to Morton SC. 2000. Variability of Coverage Probabilities When Applying a Random Effects Methodology for Meta-analysis. Available from Dr. AR Sampson.

Larholt KM, Sampson AR. Effects of heteroscedasticity upon certain analyses when regression lines are not parallel. Biometrics 1995 Jun;51(2):731-7.

Levey AS, Schmid CH, Lau J. Antilymphocyte antibodies, renal transplantation, and meta-analysis. Ann Intern Med 1998 May;128(10):863-5.

Li Y, Powers TE, Roth HD. Random-effects linear regression meta-analysis models with application to the nitrogen dioxide health effects studies. Air Waste 1994 Mar;44(3):261-70.

Louis TA, Zelterman D. Bayesian Approaches to Research Synthesis. In: Cooper H, Hedges LV, editors. The Handbook of Research Synthesis. New York: Russell Sage Foundation; 2000. p. 411-22.

McIntosh MW. The population risk as an explanatory variable in research synthesis of clinical trials. Stat Med 1996 Aug;15(16):1713-28.

Morgenstern H. Uses of ecologic analysis in epidemiologic research. Am J Public Health 1982 Dec;72(12):1336-44.

Morris CN, Normand SL. Hierarchical Models for Combining and for Meta-analyses. In: Bernardo JM, Berger JO, Dawid AP, et al., editors. Bayesian Biostatistcs 4. Oxford: University Press; 1992. p. 321-44.

Mulrow C, Cook D, editors. Systematic reviews: synthesis of best evidence for health care decisions. Philadelphia, PA: American College of Physicians; 1998.

Pauler DK, Wakefield J. Modeling and Implementation Issues in Bayesian and Meta-analysis. In: Stangl DK, Berry DB, editors. Meta-analysis in Medicine and Health Policy. New York: Marcel Dekker, Inc.; 2000. p. 205-30.

Perry BH, Sampson AR, Schwab BH, et al. Personal communication to Morton SC. 2000. A meta-analytic approach to an integrated summary of efficacy: A case study of becaplermin gel. Available from Dr. AR Sampson.

Pladevall-Vila M, Delclos GL, Varas C, et al. Controversy of oral contraceptives and risk of rheumatoid arthritis: meta-analysis of conflicting studies and review of conflicting meta-analyses with special emphasis on analysis of heterogeneity. Am J Epidemiol 1996 Jul;144(1):1-14.

Platt RW, Leroux BG, Breslow N. Generalized linear mixed models for meta-analysis. Stat Med 1999 Mar;18(6):643-54.

Prevost TC, Abrams KR, Jones DR. Hierarchical models in generalized synthesis of evidence: an example based on studies of breast cancer screening. Stat Med 2000 Dec;19(23):3359-76.

Raudenbush SW. Random Effects Models. In: Cooper H, Hedges LV, editors. The Handbook of Research Synthesis. New York: Russell Sage Foundation; 1994. p. 301-21.

Rimm EB, Williams P, Fosher K, et al. Moderate alcohol intake and lower risk of coronary heart disease: meta-analysis of effects on lipids and haemostatic factors. BMJ 1999 Dec;319(7224):1523-8.

Schmid CH, Lau J, McIntosh MW, et al. An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. Stat Med 1998 Sep;17(17):1923-42.

Schmid CH. Exploring Heterogeneity in Randomized Trials Via Meta-analysis. Drug Information Journal 1999;33:211-24.

Sharp SJ, Thompson SG, Altman DG. The relation between treatment benefit and underlying risk in meta-analysis. BMJ 1996 Sep;313(7059):735-8.

Sharp SJ. Meta-analysis regression. STATA Technical Bulletin 1998;42:16-22.

Sharp SJ, Thompson SG. Analysing the relationship between treatment effect and underlying risk in meta-analysis: comparison and development of approaches. Stat Med 2000;19:3251-74.

Shea S, DuMouchel WH, Bahamonde L. A meta-analysis of 16 randomized controlled trials to evaluate computer-based clinical reminder systems for preventive care in the ambulatory setting. J Am Med Inform Assoc 2000 Nov;3(6):399-409.

Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. Stat Med 1995 Dec;14(24):2685-99.

Smith TC, Abrams KR, Jones DR. Hierarchical Models in Generalised Synthesis of Evidence: An Example Based on Studies of Breast Cancer Screening. 16th International Society for Clinical Biostatistics Conference; 1995 Jul 31; Barcelona, Spain. Leicester (UK): University of Leicester; 1995. p. 3-32.

Stuck AE, Siu AL, Wieland GD, et al. Comprehensive geriatric assessment: a meta-analysis of controlled trials. Lancet 1993 Oct;342(8878):1032-6.

Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F; Methods for Meta-Analysis in Medical Research. Chichester, UK: John Wiley & Sons; 2000. Wiley Series in Probability and Statistics.

Sutton AJ, Jones DR, Abrams KR, et al. Systematic reviews and meta-analysis: a structured review of the methodological literature. J Health Serv Res Policy 1999 Jan;4(1):49-55.

Thompson SG, Smith TC, Sharp SJ. Investigating underlying risk as a source of heterogeneity in meta-analysis. Stat Med 1997 Dec;16(23):2741-58.

Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. Stat Med 1999 Oct;18(20):2693-708.

Turner RM, Omar RZ, Yang M, et al. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. Stat Med 2000 Dec;19(24):3417-32.

Viswesvaran C, Sanchez JI. Moderator Search in Meta-analysis: A Review and Cautionary Note on Existing Approaches. Educational and Psychological Measurement 1998 Feb;58(1):77-87.

Walter SD. Variation in baseline risk as an explanation of heterogeneity in meta-analysis. Stat Med 1997 Dec;16(24):2883-900.

Wieland D, Stuck AE, Siu AL, et al. Meta-analytic methods for health services research--an example from geriatrics. Eval Health Prof 1995 Sep;18(3):252-82.

Zeegers MPA, Tan FES, Dorant E, et al. The impact of characteristics of cigarette smoking on urinary tract cancer risk, a meta-analysis of epidemiologic studies. Cancer 2000 Aug;89(3):630-9.

Zhong L, Goldberg MS, Parent ME, et al. Exposure to environmental tobacco smoke and the risk of lung cancer: a meta-analysis. Lung Cancer 2000 Jan;27(1):3-18.

Zhou XH, Brizendine EJ, Pritz MB. Methods for combining rates from several studies. Stat Med 1999 Mar;18(5):557-66.

Appendices

Appendix A. Expert Panel and Acknowledgements

Dr. Jesse Berlin, University of Pennsylvania School of Medicine

Dr. William DuMouchel, AT&T Research

Dr. Vic Hasselblad, Duke Clinical Research Institute

Dr. Joseph Lau, Division of Clinical Care Research, New England Medical Center

Dr. Thomas Louis, RAND, Washington, DC

Dr. Martin McIntosh, Biostatistics Department, University of Washington and Fred Hutchinson Cancer Research Center

Dr. Ingram Olkin, Department of Statistics, Stanford University

Dr. Allan Sampson, Department of Statistics, University of Pittsburgh

Dr. Chris Schmid, Tufts University School of Medicine, New England Medical Center

We gratefully acknowledge the participation of and constructive feedback by these expert panelists. Acknowledgments are made with the explicit statement that this does not constitute endorsement of the report.

Dr. Jesse Berlin from the University of Pennsylvania School of Medicine, and Dr. Thomas Louis from RAND (now at Johns Hopkins) reviewed the draft report. In addition, Southern California Evidence-based Practice Center staff not on the study team also reviewed the draft report. We gratefully acknowledge the assistance of these referees. Acknowledgments are made with the explicit statement that this does not constitute endorsement of the report.

Appendix B. List of Southern California Evidence-Based Practice Center Attendees at the Expert Panel Meeting

Dr. John Adams

Dr. Ian Coulter

Dr. Mary Hardy

Dr. Sally Morton

Ms. Elizabeth Roth

Dr. Paul Shekelle

Ms. Marika Suttorp

Appendix C. Expert Panel Meeting Agenda

9:00 am	Welcome and Overview	
9:30 am	Unifying Notation for Models	John Adams
10:00 am	Meta-regression by Example	Vic Hasselblad
10:25 am	Control Rate Models	Chris Schmid
10:50 am	BREAK	
11:05 am	Bayesian Meta-analysis	Tom Louis
11:30 am	Preliminary Simulation Results	John Adams
12:00 pm	Lunch	
1:00 pm	Methodological Challenges in Meta-regression	Jesse Berlin
	Discussion	
3:00 pm	BREAK	
3:15 pm	Focus on Key Questions	Sally Morton
4:30 pm	Future Work	Sally Morton
5:00 pm	Adjourn	

Appendix D. Objectives and Key Questions of the Expert Panel Meeting

Objectives of the Meta-regression Meeting

Our meeting objectives are to

a. Obtain the reaction of the experts to our research so far, and their suggestions on how to proceed both in terms of the meta-regression topic and future methods topics.
b. Answer the key questions described below.
c. Stimulate a dialogue of the potentials and limitations of meta-regression to inform the evidence-based practice of medicine.

Key Questions to Address At the Meeting

Quantifying Heterogeneity

1. Is the quantification of heterogeneity important?
2. If the quantification is important, how should it be done in different circumstances?
3. What, if any, is the role of formal statistical testing for heterogeneity?

Accounting For and Understanding Heterogeneity

4. Is the quantification of heterogeneity sufficient, or does one need to account for (incorporate) it?
5. Is accounting for heterogeneity sufficient, or does one need to understand its causes?
6. What methods should be used to account for and/or understand heterogeneity in different settings?

Regarding Meta-regression

7. Which meta-regression approaches should be used in which circumstances?
8. Are there meta-regression "best practices," and do they vary by situation?
9. How should meta-regression results be reported and interpreted?
10. How does meta-regression compare to other common meta-analysis approaches (e.g., constructing random effects estimates within separate study strata)?
11. What are the key concerns about meta-regression?
12. What research needs to be conducted to improve the application of meta-regression?
13. How does meta-regression of study-level statistics relate to modeling patient-level data?

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'd like your guidance regarding which methodological research topic we should pursue next, and what methodology should be used (e.g., an expert panel, simulation, etc.)? Given our work in this first year, we now feel that continued investigation of meta-regression might be the most fruitful course. We had originally proposed quality assessment of observational studies as our next topic but several research groups are already working on this topic. Conceivably we could contribute by convening an expert panel akin to the QUORUM[2] or MOOSE[3] efforts.

Appendix E. Responses to Referees

We note that one referee's comments were handwritten margin notes on the manuscript. Those that were copyediting comments we considered and made necessary changes. We do not record those changes here. We summarize this referee's margin notes that were substantive below and respond. The second referee provided written comments we include those below and respond. We reordered and interwove the comments of both reviewers to follow the order of material in the report.

General: The report needs a prose edit and reduction of repetition.

Response: We have edited the manuscript and the referee's edits were very helpful in this regard.

Page v, paragraph 1 (Objective): here and elsewhere you use the term "heterogeneity" and define it only as differences between studies. As an aside, grammatically, I think it should be "among" studies, but "between" seems to be a holdover from the (incorrect) language of ANOVA. In any case, you might want to be specific and talk about variability in study results. You really are talking about treatment effects throughout most of the document, so why not just say variability in estimated treatment effects?

Response: We have made this change throughout the report.

Pg x: I'd switch the order of presentation here. I'd first say that the sources of heterogeneity should be investigated. Failing any explanation, the heterogeneity should then (one might argue) be incorporated into (accounted for in) the analysis.

Response: We have made this change throughout the report.

As a bit of an aside, there are those who would argue (I believe Sander Greenland might be one) that it might not be wise to present a single summary result in the presence of unexplained heterogeneity. The average might not be representative of any of the individual studies, according to this concern.

Response: We have made a comment on this point-of-view in the Introduction section.

At the end of the first paragraph on page x, you say that "differences among studies, …, are a strength …". Here you apparently mean differences in design and study populations. Be explicit.

Response: We have made this change.

Page xiii, para 2: You use the word "covariates" in line 4. Here and maybe in one or two other places it's not always clear whether these covariates are predictors of risk or of the treatment effect.

Response: The phrase "covariates" appeared in has been deleted.

I'd say that a study with N=5 would probably be rejected from most meta-analyses, although I could be persuaded otherwise.

Response: Our panel advised that we drop the k (number of studies)=3 value in our simulation study. The only values of this parameter included are 10 and 30 as explained in the Panel Recommendations section in Results.

Page 2, bottom: presumably here you mean that certain dependencies are beyond the *original* investigator's control (as opposed to the meta-analyst, who is also an investigator).

Response: We have changed the term to "original investigator" for clarity as suggested.

Page 2, bottom: Some aspects are within the investigator's control such as the selection of patients, etc.

Response: We have clarified that some aspects of clinical comparability can be addressed by the original investigator.

Page 3, 2nd paragraph: Add follow-up of patients and measurement error as aspects the investigator can consider/control.

Response: We have added these as examples.

Page 3, 3rd paragraph: Add the word "important" before "heterogeneity at all."

Response: This paragraph has been deleted.

Page 4: You should recommend estimating the heterogeneity to assess its importance.

Response: We have made this change.

Page 4: I'd head the section "Addressing Heterogeneity" as you are not just "incorporating" it.

Response: We have made this change.

Page 6: You need to get formal in the presentation of the models.

Response: The models are presented mathematically in Results section so we have not introduced the mathematical notation this early in the report. To remove redundancy, we have shortened this first discussion of the methods.

Page 7: It can also be that the model is wrong inducing a correlation between the treatment effect and baseline.

Response: Model mis-specification is discussed in the Results section with respect to the simulation results.

Page 8: Opening paragraph has been stated too many times.

Response: We have deleted this opening paragraph.

Page 9: Will the readers understand the notation "SCM."

Response: We have changed this to "Morton" to indicate which author did the title screening.

Page 11: Comment in margin says "Of course, if the patient level is not logistic and has covariates that vary over patients, then the aggregated model is not logistic. This needs to be considered."

Response: Model mis-specification is discussed in the Results section with respect to the simulation results.

Page 14-15: You've missed a lot of literature. For example the Cooper & Hedges Handbook, many Hedges and Olkin articles, etc. You need to explain why these and others aren't in your list.

Response: We have included eight new references including the references suggested by the reviewer. We do note that the Bayesian hierarchical modeling literature was not included in our search. We have included several books that do survey the field.

Page 15: Mortality p=1 unless you give a finite follow-up window.

Response: We have clarified throughout the report that we mean mortality within a specified follow-up time.

Page 16: Not sure where else to ask this: When you speak of covariates (as in model 2) – is the ultimate goal to estimate the true treatment effect adjusted for imbalances in covariate values between treatment arms? Why else would you bother modeling "risk"?

Response: This is the goal of the modeling.

Page 17: I prefer "i" for patient and "j" for study.

Response: We have decided not to change the notation.

Page 17, bottom: You describe how you would fit model 2, but don't mention indicator variables for "study." Would you not include these indicators?

Response: We have revised the text to be clear that study-level indicator variables and study-level covariates cannot both be included in a model since they are confounded.

Page 20: The reviewer wrote margin comments that the technical problem was "big," mis-specification "will" occur, and other related comments.

Response: We have edited the prose in this paragraph to take into account this is a major technical problem, that the model will be mis-specified if only aggregated variables are available, and that randomization only has a limited effect on this problem.

Page 20: You make the claim here that it is better to specify the covariate values at the level of the treatment arm than at the study level. In light of my question above (Page 16 comment above), I actually share your opinion. However, you might note here, as you sort of do later on, that when you specify the covariates at the study level in the logistic model, you then get

answers that are directly comparable (and nearly the same in numerical value) to those obtained from fitting models with ln OR as the outcome variable. This is an important point if one is comparing the two modeling approaches (logistic versus linear models of ln OR).

Response: We have added this comment to the discussion of this issue.

Page 21: The reviewer made margin comments that he would prefer that the subjunctive in "might be informative" and "might be applied" not be used.

Response: We prefer to keep this discussion in the subjunctive voice.

Page 25: Why variance=1 in the distribution of the baseline effects? Also on page 27, a reviewer asked the related question of "is assuming that the variance is one sufficiently general?"

These assumptions do not result in a loss of generality. The simulated distributions are multiplied by the coefficients we have selected (usually -0.6,0,0,6). This means the combined effect of the product has variance β^2. *So the variance is controlled by the coefficient, and there is no loss of generality. Similarly, setting the means of the covariate distribution to zero does not reduce the generality of the model. Means of the base outcome rate and/or the treatment effect can be introduced via* β_0 *or* γ_0 *respectively. We have added a discussion of these issues in the text in both the Simulation Set-up and Simulation Parameters sections.*

Page 25: The tau term looks like a random effect. Use some other letter.

Response: We prefer to keep this notation given we had to choose unique notation for a large number of other parameters as well.

Page 27: I know I'm not supposed to proofread, but want to make sure I understood what I read to some small degree. In the first equation on this page you have the vector ending in $:_i$. Should that be x_i?

Response: We believe this was a problem with the mathematical formula translating across machine platforms. The notation is correct in the printed copy.

Page 27, bottom: Does the fact that the additive treatment effect is defined at the mean value (zero) of z and x imply that these covariates should generally be centered at their means, or is this just an artifact of the particular values of the covariates (-x, 0, x) you have chosen for the simulations?

Response: This is an artifact of the simulation values chosen as the referee correctly notes, and is related to the fact we do not lose generality by assuming the mean is zero for these covariates.

Page 28, bottom: For the values of sbar, I think you may mean that they correspond to *odds* of 0.55 to 1.82. I thought g0 was the odds ratio. The logic would then extend to the values of the probabilities you specify.

Response: Sbar (now denoted as ϕ) is the intercept in a logistic regression and therefore is on the log odds ratio scale so the text as written is correct. All coefficient parameters in the simulation are log odds ratios.

Page 28, Table 2. Add a column explaining the parameter values as you did for sbar.

Response: We have added such a column to Table 2.

Page 28: Is the random effects model with no covariates really standard?

Response: We have deleted this sentence as well as the one following it regarding which methods are standard.

Page 29, bottom: I see what you mean. Arm level covariates will be used in a study-level model. When would you use arm-level.

Response: Sometimes publications report covariates at the arm level. For example, the average age in the placebo group.

Page 30, table 3: Maybe put earlier.

Response: We prefer to keep the table at this location.

Page 30, bottom: You can adjust (a bit) for aggregation if you also know the SD of age, etc.

Response: We have not included this comment in the discussion at it is outside the scope of the logistic meta-regression approach.

Page 30: You note on the bottom of the page that the method could be employed with two cases per study, if using study-level covariates. In fact, one could (in SAS or in STATA) fit a logistic model with two data points per study, in which each contains a numerator and a denominator for each treatment arm within the study. This need not require specifying covariates only at the study level.

Response: We have deleted this note to avoid confusion.

Page 31: Do you get the same answers if you force the RE model with no covariates to use REML as you do with the method of moments estimator? It seems to me I've seen reports of bias in the DerSimonian and Laird method, that might be related to their use of method of moments.

Response: We have also observed some slight bias. We have clarified that in our experience the results of these two approaches are roughly similar.

Page 31, bottom: Do you want to explain why you used maximum likelihood estimation?

Response: We believe this is the most relevant approach as it was generally used in the examples we found in the literature.

Page 32: Isn't the number of simulation parameter combinations over 7000?

Response: We began with a design that generated 1944 combinations. Based on our panel's recommendation, we expanded our simulation. We have clarified this in the text.

Page 32: Sentence at the bottom is repetitive.

Response: We have deleted this sentence.

Page 33: You mention calculating bias as a percentage of the true parameter. There are lots of ways one can do that and be "correct". For example, one might take:
(ln OR true – ln OR estimated) / (ln OR true) and multiply that to get a percent.
Is this what you did? I didn't see where you stated explicitly what you did.

NOTE: The second reviewer was also confused about our definition of bias.

Response: We used the definition of bias hypothesized by the reviewer except that the numerator is ln OR estimated – ln OR true so that positive bias means we are overestimating the true parameter value.. We have added this equation to the report to be explicit.

Page 34: Need to reorganize to make the number of simulation parameters clear.

Response: We believe the edit regarding the reviewer page 32 comment (see above) will make clear that the panel expanded our design and thereby increased the number of parameter combinations.

Page 37: paragraph at the bottom needs to be made more clear.

Response: This paragraph has been expanded to read:

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.

Page 38: Do you want some RMSE displays. They would be a great addition. A plot would be good to see. NOTE: A similar comment was made in the margin on page 41.

Response: We were unable to determine how to clearly display the simulation results in a plot. Therefore we prefer to keep the results in tabular form. We have further clarified and summarized the results for each table.

Pages 42-44: The referee made several comments about the notation and summary comments.

Response: We have revised this section to use greek notation throughout and also clarified the summary of each table. We have also edited our conclusions to contain summary conclusions across methods for the practitioner.

Page 42, second paragraph: If the results do not converge, then do not present them.

Response: We investigated this issue and found that control rate meta-regression is not estimating the same parameter as the other four methods. We explain this in a new subsection "What These Methods are Estimating" that appears just before "Panel Recommendations." Once we adjusted our bias calculation to take this difference into account, the anomalies we had observed previously in the control rate meta-regression results disappeared. The results across all methods are more intuitive and consistent. As a result of this change, the Results Tables (new Tables 5-10) have changed.

Table 5: Consider rounding all percents to whole numbers. Decomposing by a main effects model, put the main effects as row/column values and display residuals.

Response: We prefer to keep the level of precision as is. The suggested display is another way of looking at the same results. We believe with the expanded and clarified discussion of the tables that interpretation will be easier for the reader so we prefer to leave the results tables in the original format.

Results: By way of limiting the number of possibilities you present (and perhaps simplifying the tables), I wonder if you need the results for all values of g1 (which, if I'm remembering, relates the treatment effect on the log odds ratio scale to the logit of the baseline risk. For a beneficial treatment, i.e., a negative value of g0, presumably the most common situation would be to have a larger treatment effect (more negative) as the value of the baseline risk gets larger (more positive – or really less negative). I believe this translates into the negative value for g1. Perhaps you could present only the values of bias for g1 negative and g1 = 0. Just a suggestion.

Response: We have decided to keep the entire range of simulation values for symmetry.

Table 4: I found myself wanting the meaning of the meaning (not to get philosophical – but a longer explanation of the "meaning" would be helpful.) For example, the first one seems to mean that the bias depends on the degree of association between baseline risk and treatment effect, but that dependence, in turn, depends on the value of the baseline risk.

Response: We have further clarified and summarized the results for each table.

Results: I liked the summaries for the practitioner describing the tabulated results.

Response: We have further clarified and summarized the results for each table.

Results: In the presentation of the simulation results, you rely on the reader to do the "mapping" back to the parameter descriptions in Table 2. It may just be my own view, but I found that even after doing that mapping, it was hard to get an intuitive sense for the simulation results in the tables. Knowing, for example, that the value of g1 is 0.6 still left the results out of context for me. I'd rather see what the true odds ratio is for a given combination of parameters, what the estimated value is, and then what the percent bias is. For example, in Table 5, for sbar = -0.6, I know it's simply an exercise to plug in the values of all the parameters to obtain the treatment effect for, say, g1= 0.6, but perhaps that work could be done for the reader.
NOTE: The other referee requested that we stay with greek letters throughout the text and tables.

Response: We have used greek letters for notation throughout the report, rather than presenting the mathematics in greek notation and switching back to Roman acronyms for the tables, for ease in interpretation. The number of parameter values, the fact that the results in the tables are for a certain number of fixed parameter values while the interaction of interest varies, makes the translation suggested difficult.

Conclusions: One thing you didn't seem to address explicitly, but I would have been curious about, was the power of fixed-effects versus random-effects regression models to detect underlying associations between, say, the log OR and either study-level or individual-level covariates. One might argue, for example, that part of the goal of a meta-regression is not just to estimate a single treatment effect (especially in the presence of heterogeneity of treatment effects known to depend on a covariate), but, in fact, to identify those predictors of treatment effect. Which method does better at that? I would expect the fixed-effects approach to be more powerful, but am willing to be convinced otherwise.

Response: Determining the relative ability of methods to detect specific predictor effects is beyond the scope of this work. However, we have added this suggestion to the future research agenda paragraph in the Conclusions section.