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.