AN EXAMINATION OF THE EFFICACY OF CLASSICAL AND BAYESIAN META-ANALYSIS APPROACHES FOR ADDRESSING IMPORTANT META-ANALYSIS OBJECTIVES

by

Jill Lucas Findley

A dissertation submitted to the Graduate Faculty in Educational Psychology
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy, The City University of New York

2011

This manuscript has been read and accepted for the
Graduate Faculty in Educational Psychology in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

David Rindskopf, Ph.D.

Date — Chair of Examining Committee

Mario Kelly, Ed.D.

Date — Executive Officer

David Rindskopf, Ph.D.

Jay Verkuilen, Ph.D.

Hannah Rothstein, Ph.D.
Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

AN EXAMINATION OF THE EFFICACY OF CLASSICAL AND BAYESIAN META-ANALYSIS APPROACHES FOR ADDRESSING IMPORTANT META-ANALYSIS OBJECTIVES

by
Jill Lucas Findley

Advisor: David M. Rindskopf, Ph.D.

This paper examines the efficacy of classical versus Bayesian meta-analytic models for addressing the five important meta-analytic objectives that were proposed by Higgins, Thompson, and Spiegelhalter (2009). In addition, it presents and examines a sixth important meta-analytic objective within the classical and Bayesian frameworks – a consideration of how meta-analytic inferences may change depending upon the uncertainty in the estimate of the amount of heterogeneity. In order to meet this sixth objective, this study uses a classification system which follows the guidelines proposed by Rothstein, Sutton, and Borenstein (2005) for describing the impact of publication bias. Here, the impact of the way meta-analytic results may change depending upon the uncertainty in the heterogeneity is classified with the use of qualitative indicators akin to those used by Rothstein et al. (2005). Thus, the discrepancy between the best-fitting meta-analytic model and the meta-analytic models used for heterogeneity sensitivity analyses is described as: (a) “minimal”, when the fitted meta-analytic models and the estimates remain similar; (b) “modest”, when the fitted meta-analytic models remain the same, but the estimates change to a moderate degree; and (c) “severe”, when the fitted meta-analytic models and estimates differ substantially from each other.

This research suggests that Bayesian hierarchical linear modeling offers the most complete and accurate approach for addressing all relevant meta-analytic objectives. The project uses five different meta-analytic datasets as illustrative examples. It also provides examples of the code for the classical models for the *metafor* package, the Bayesian code for the WinBUGS package, and the S-PLUS code for the Bayesian *hblm* function. Given the

complexity and nuances associated with Bayesian model development, a Bayesian quality assurance meta-analysis checklist was refined for this research project.

The use of meta-analytic trace plots produced with the *hblm* function, which depict the dependency of meta-analytic results on the values of the standard deviation of the between-study variance, is shown to summarize the essence of a fully Bayesian meta-analysis. In a single picture, these plots summarize four out of five of Higgins et al.'s (2009) important meta-analytic objectives. Furthermore, meta-analytic trace plots also provide the additional, important (though underappreciated) advantage of representing how meta-analytic estimates change depending upon the uncertainty in the estimate of the amount of heterogeneity. This paper suggests that the future design of meta-analytic trace plots should also include inlaid curves that depict the estimates for the predicted effect in a new study so that all six important meta-analytic objectives could be addressed in a single graphic display.

Acknowledgments

I would like to extend a heartfelt thank you to my esteemed dissertation chairman, Dr. David Rindskopf, for providing me with valuable wisdom, guidance, patience, support, and humor throughout the entire research project.

This project would not have been possible without the amazing childcare/preschool offered to my two young children, Tommy and Jack, by the Graduate Center of the City University of New York. This nurturing preschool (for both children and parents) is the hidden gem of all preschools in New York City. I am so fortunate that my children had the opportunity to have wonderful childcare such as this, while their mom had the good fortune to pursue her doctorate in the research library.

I also extend my sincere gratitude to Dr. Hannah Rothstein, a member of my dissertation committee, for introducing the topic of meta-analysis to me, inspiring my passion for research synthesis, offering mentorship, and providing me with such thoughtful insight. Dr. Rothstein taught my favorite course (Meta-Analysis) at the Graduate Center. I also wish to thank Dr. Jay Verkuilen for being so approachable, helping me with the statistical program, R, and having such great expertise; Dr. Patricia Eckardt for all of the support that she offered me as a graduate student when I had recently become a new mother and then in quick succession, a mother yet again; and Dr. Howard Everson for his amicability and genuine interest in my topic.

My passion for research began when I was enrolled in my first master's program at Columbia University under the tutelage of Dr. Debra Clayton-Krasinski. I had a wonderful research partner at Columbia University – Janice Perdek Coffey – who inspired me to be both a better student and a better person. I am thankful to have had that opportunity to study with Janice. In addition, another friend from Columbia University, Mladen Laudanovic, was especially helpful to me with this research project by answering questions about computer programming and Bayesian statistical modeling. I am also very appreciative to the Information

Technology staff at the Graduate Center of the City University of New York, in particular Glenda Perreira and Andrea Allard, for providing me with computer training .

I am so incredibly thankful to my sweet husband, Barry Findley, for offering me so much love, devotion, and support. My two beautiful children, Tommy and Jack, who are the little loves of my life, have inspired me to complete this project as well. I have deep gratitude to my family and friends for always providing love and encouragement – foremost, my truly amazing, loving, and inspirational parents, Daniel and Marcy Lucas, and my sisters, Kim Lucas Langway and Amy Lucas. Finally, there has been one indispensible childhood friend who has always offered me so much friendship and support as we grew up together – Annie Kemper Marchant. Everyone should be as lucky as I to be able to go through life with a best friend like Annie.

Table of Contents

List of Tables

List of Figures

Chapter I: Introduction

Results of meta-analyses—the quantitative synthesis of data from a series of studies—are becoming increasingly important in evaluating the research evidence regarding the effectiveness of educational, psychological, and health care interventions. Meta-analytic results allow for more powerful estimates of treatment effects than those estimates provided by individual studies considered in isolation (Borenstein, Hedges, Higgins, & Rothstein, 2009). Clinical, educational, and medical decision-making is based increasingly on evidence-based practices and the totality of the relevant accumulated evidence that meta-analyses provide (Sutton & Higgins, 2008). Meta-analytic results help inform practitioners of evidence-based medicine, policymakers, and regulatory bodies, about the overall efficacy of different treatment interventions.

The citation impact of meta-analytic studies is profound; meta-analyses are the most frequently cited type of research design in the medical literature (Patsopoulos, Analatos, & Ioannidis, 2005). In fact, in recognition of the growing importance of meta-analyses, the United States government's American Recovery and Reinvestment Act of 2009 appropriated funds for comparative effectiveness research (CER), which synthesizes research that compares treatment outcomes and efficacies. The same CER likewise considers the evidence for prevention, treatment, and diagnosis of diseases and other health conditions (H.R.1, S.1, 111th U.S. Congress, first session, 2009). A widely accepted goal of research is to produce cumulative knowledge that is generalizable, and meta-analyses, in particular random/mixed-effects models and Bayesian hierarchical linear models, provide a means of addressing this goal through quantitative integration of the cumulative research on a topic.

In a meta-analysis, data (usually in the form of published summary statistics) from studies) are converted with statistical techniques into a standardized measure of effect sizes such as standardized mean differences, odds ratios, or correlation coefficients. Converting study results into a common standard metric allows a research synthetist to make comparisons

of effect sizes easily across studies (Lipsey & Wilson, 2001). A noteworthy advantage of meta-analysis is that it yields a summary effect size estimate that has considerably more power to detect effects than that of any of the individual studies. This power permits meta-analysts to uncover more meaningful effects when study results concur and to discover study-level characteristics that can help explain differences in effects among studies (Lipsey & Wilson, 2001).

Different computational approaches—such as fixed-effects (FE), random/mixed-effects models, and Bayesian methods—can be used to compute meta-analytic summary effects; however, each of these methods has different assumptions that affect the average effect size estimate and the width of the corresponding confidence interval. It is important for researchers to understand the theoretical and statistical distinctions between FE, random-effects (RE), and Bayesian models because differences among these models will affect the certainty of meta-analytic findings (Schmidt, Oh, & Hayes, 2009).

Choosing the appropriate meta-analytic statistical model involves thoughtful consideration by the systematic reviewer regarding the population of studies to which the researcher wants to generalize the results of the meta-analyses. However, confusion and debate remains among researchers over selecting appropriate modeling strategies for making such statistical inferences (Hedges, 2009). Borenstein et al. (2009) recommend basing the model choice upon whether the included meta-analytic studies share a common effect size and goals in performing the meta-analysis. Shadish and Haddock (2009) recognize that there is often no single correct answer to the question regarding model choice, while others imply that FE models are rarely tenable or appropriate (Higgins, Thompson, & Spiegelhalter, 2009; Borenstein et al., 2009; National Research Council, 1992). Advocates of Bayesian methods for research synthesis believe that Bayesian modeling ends the debate over FE versus RE models because, as will be discussed in a later section, Bayesian models can explicitly model heterogeneity (Sutton, Abrams, Jones, Sheldon, &Song, 2000).

If systematic reviews and meta-analyses are well-constructed statistically with sound modeling principles and validity, consumers and stakeholders can be assured that policy decisions based on meta-analytic results are well-informed and derive from best-available evidence models. Meta-analysis is a constantly evolving research tool for evidence synthesis (Schmidt, 2008). Because of this, researchers need exposure to the newer methodological techniques and refinements so that they can apply these contemporary methods for valid research synthesis. Many meta-analyses are conducted by non-statisticians researchers (Sutton & Higgins, 2008), who need to become familiar with progressive meta-analytic techniques such as Bayesian methods, which offer improvement and greater flexibility over more traditional meta-analytic methods. Despite the numerous advantages and flexibility that Bayesian methods offer to meta-analytic data, many researchers still regard them as complicated and difficult to implement due to the computational complexity of the models and the fact that the majority of published meta-analyses have historically relied on the classical FE and RE models.

It is likely that some previously published classical FE and RE systematic reviews did not apply such methodologies for data analyses, and, as a result, the meta-analytic summaries may have less accuracy or may simply focus on interpreting the overall summary effect without considering heterogeneity of effects. For these reasons, many meta-analytic experts have recognized the need for a streamlined and unified Bayesian modeling approach that adequately integrates FE, RE, mixed-effect, and Bayesian hierarchical models (e.g., DuMouchel & Normand, 2000).

A paradigm shift has recently occurred in meta-analytic methods literature from an emphasis on focusing on the weighted mean effect size alone to an emphasis on exploring between-study heterogeneity (Higgins, Thompson & Spiegelhalter, 2009; Sutton & Higgins, 2008). Often, the single parameter of $\hat{\mu}$, the weighted mean effect size, when considered in isolation, does not provide an adequate summary because of a heterogeneous distribution of

effects (Raudenbush & Bryk, 1985). Higgins et al. (2009) emphasize that not only should the estimation of the between-study variance be given equal importance in summarizing meta-analytic results, but that consideration should also be given to the uncertainty in the estimate of the between-study variance. However, this recommended emphasis shift toward heterogeneity estimation and exploration is not always evident or adequately addressed in many recently published meta-analyses.

This research project focused on reanalyzing selected meta-analytic data sets using more sophisticated modeling techniques and refined methodologies and comparing the stability of the overall meta-analytic results among different models. The view taken here and the one more fully discussed in the section, "Description of a Fully Bayesian Model for Meta-analysis," is that Bayesian hierarchical linear models offer the most flexible and appropriate meta-analytic modeling strategies for addressing meta-analytic objectives.

Given the recent embrace of evidence-based healthcare and the related explosion in the use of meta-analysis (Sutton & Abrams, 2001), an instructive and illustrative research project such as this is valuable because many systematic reviewers remain unfamiliar with the manner in which Bayesian techniques are applied to meta-analytic data and the substantive implications of using such techniques to investigate statistical heterogeneity. This study:

(a) describes a fully Bayesian model and the methods used to employ the model,

(b) illustrates and compares the results among Bayesian and classical meta-analytic models in terms of their ability to address important meta-analytic objectives,

(c) provides instructions and computer programming code so that research synthetists can replicate a Bayesian analysis with their own meta-analytic data,

(d) develops a Bayesian meta-analysis quality assurance checklist, and

(e) demonstrates how Bayesian hierarchical linear modeling offers a superior ability to address all relevant meta-analytic objectives as applied to examples from five

previously published meta-analyses from the education, psychology, and healthcare literature.

The second chapter of this paper, “Review of the Literature,” provides an overview of the classical FE, random/mixed-effects meta-analytic models and estimation methods, along with the limitations of these models. In addition, this section discusses some of the recommended practices for estimating and exploring heterogeneity within a meta-analysis. It highlights the advantages of using random/mixed-effects models for addressing relevant meta-analytic hypotheses.

The second section of Chapter II, “A Fully Bayesian Hierarchical Model for Meta-Analysis,” explains a fully Bayesian hierarchical linear model, reviews the assumptions for Bayesian models, discusses the use of prior distributions and sensitivity analyses, and highlights the advantages that Bayesian models offer for addressing relevant meta-analytic hypotheses over classical methods.

Chapter III describes the methodology for this research project and reviews the five meta-analytic data sets that were used as illustrative examples. Chapter IV discusses the results, comments on the usefulness of the meta-analytic software programs used in this project, and suggests directions for future research.

Statistical Notation

Statistical notation used for meta-analyses varies among different sources. This paper will use the following notation, which is summarized in Table 1. The notation in the equations cited throughout this paper has been changed to reflect the unified notation system used in this paper.

Table 1

Statistical Notation

Notation	Other common notations	Definition
k	N	Number of effect size estimates
Y_i	T_i	The observed effect size estimate from study *i*
s_i^2	v_i, V_{Yi}, σ_i^2	Within-study sampling variance of Y_i
w_i		Weight assigned to study *i* in a FE model
w_i^*		Weight assigned to study *i* in a RE model
θ		Unknown parameter or parameters
δ_i	μ_i, ζ_i	Random effect of study *i*;
τ^2	σ_θ^2	RE variance component; between-study variance;
τ	σ_θ	Standard deviation of the between-study variance
μ		Weighted mean effect size (for a simple RE model)
$\hat{B}_0$		Model intercept for the weighted mean effect size
$\hat{d}$		Weighted mean effect size in terms of standardized mean differences for a simple FE /RE model
X		The matrix of study-level covariates
x_i		The i^{th} row of X
$\beta_1, \beta_2 \ldots$		Study-level covariates
$x_i\beta$		Linear combination: $x_i\beta = \beta_0 + \beta_1 x_{i1} + x_{iJ}$
V_μ	V_M, *V..*	Variance of FE summary effect
V_μ^*		Variance of RE summary effect
se_μ	SE_μ, SE_M	Standard error of mean effect size estimate
e_i		Error of estimation at Level 1
$\theta_i^*(\tau)$		Empirical Bayesian (EB) study-specific estimate
θ_i^*		Fully Bayesian (FB) study-specific estimate
θ_{new}		Predicted effect in a new study
θ_i^{**}		Posterior variance of the FB shrinkage estimator
m		The mean for μ in a Bayesian diffuse prior distribution, (typically $m = 0$)
d^2		The variance of μ in a Bayesian diffuse prior distribution (typically $d^2 \to \infty$)
$\pi(\tau)$		Prior distribution for the square root of the between-study variance.
s_0		Median of a 'DuMouchel' prior on τ
s_0^2		Harmonic mean of the sampling variance for a DuMouchel prior on τ

Chapter II: Review of the Literature

Classical Meta-analytic Models

In a meta-analysis, also known as a quantitative research synthesis, quantitative methods are used to combine statistically the results of an ensemble of similar research studies into a weighted mean and explore the consistency of the findings. Current classical meta-analytic methods allow a researcher to: a) estimate the magnitude of the effect size with increased power beyond that of an individual study, b) estimate and evaluate the consistency of study outcomes across a series of studies, c) identify study-level characteristics that are associated with differences in study outcomes, d) delineate which treatment groups or subgroups benefit particularly from an intervention, e) estimate a prediction intervals for an effect in a new study, f) quantify and construct a 95% confidence interval (CI) for the heterogeneity.

Fixed Effect (FE) Models.

Many published meta-analyses in education and psychology have traditionally relied on FE models (Schmidt et al., 2009; Schmidt, 2008; Stangl & Berry, 2000), although the use of these models is often not adequately justified (Shadish, Cook, & Campbell, 2002), frequently oversimplified (Lau, Ioannidis, & Schmid, 1998), and often not reasonable given the heterogeneity that inherently exists among studies in meta-analyses. Historically, many systematic reviewers have preferred the FE model because FE models offer simpler computational formulas and are easier to conceptualize (National Research Council, 1992). With an FE model the *a priori* statistical assumption is that there is a single, underlying, true effect size, μ, which is shared by all *k* separate studies. The assumption of the FE model is that the effect size is fixed and homogeneous across studies: $\theta_i = \ldots = \theta_k = \mu$, where θ_i is the population effect of the i^{th} study with an ensemble of *k* independent studies. FE models assume

that the variance observed across studies can be attributed solely to sampling variability and that τ, the standard deviation of the between-study variation in true effect sizes, is equal to zero.

In an FE model the observed effect size, Y_i, for study *i*, is represented by the population mean, μ, plus the within-study sampling error: $Y_i = \mu + \varepsilon_i$ (Borenstein, et al. 2009). The overall summary meta-analytic effect size is calculated by averaging effect sizes according to the weight assigned to each study. In an FE meta-analysis the weight (i.e., the precision) assigned to each individual study is the inverse of the sampling variance (squared standard error): $w_i = \frac{1}{s_i^2}$, where s_i^2 is the within-study error variance for the i^{th}study, which is inversely proportional to the within-study sample size (Shadish & Haddock, 2009).

The overall FE treatment effect, $\hat{\mu}$, is estimated as a weighted average:

$$\hat{\mu} = \frac{\sum_{i=1}^{k} w_i Y_i}{\sum_{i=1}^{k} w_i} = \sum_i \frac{w_i}{\sum_i w_i} Y_i \,, \qquad (1)$$

where the *i*th study reports an observed effect size of Y_i with a corresponding assigned weight of w_i (Shadish & Haddock, 2009), the numerator in the middle term equals the sum of the products of each effect size multiplied by its weight, and the denominator is the sum of the all the individual weights (Borenstein, et al., 2009).

The variance v_μ of the weighted mean effect size is estimated as the reciprocal of the sum of the individual study weights, $v_\mu = \frac{1}{\sum_{i=1}^{k} w_i}$, and the square root of v_μ is the estimated standard error of the mean effect size, $SE_u = \sqrt{\frac{1}{\sum_{i=1}^{k} w_i}}$ (Borenstein et al., 2009).

It is important to note that the FE formula for the weights assigned to each study and the standard error of estimate of the average effect size *does not* include a term for the variance observed between studies, because this term is assumed to be zero. This important distinction between the FE and the RE model will be reviewed in a following section because the decision to exclude the between-study variance component affects the size of the standard error of the mean effect sizes, uncertainty estimates, and meta-analytic conclusions.

The null hypothesis of the FE meta-analytic model is that there is zero population effect in every study, $\mu = 0$. This hypothesis can be tested with the statistic, $Z = \frac{|\hat{\mu}|}{SE_{\mu}}$, or the 95 percent confidence interval width, $\mu \pm (1.96)SE_{\mu}$, can be examined to see if it includes zero.

Disadvantages of FE Models.

Synthesizing the results of studies that vary in design, populations sampled, and treatment protocols will inevitably result in a compilation of effect sizes that has an inherent element of diversity (Higgins & Thompson, 2002), and it can be argued that there is always going to be some variation across studies (National Research Council, 1992), making FE estimates invalid. When there is a large amount of variation present between studies, termed *heterogeneity*, FE models may be particularly imprecise in their estimation resulting in threats to the validity of the meta-analytic results. Higgins et al. (2009) note that the FE assumption of homogeneity of effect sizes is often untenable for studies in biomedicine and social sciences because these studies are likely to differ from each other on numerous dimensions such as populations, settings, treatments, and outcomes, and they recommend avoiding the use of FE models. Many researchers have recognized the limitations with FE meta-analytic methods and have advocated the use of other methodological options such as random effect models (Schmidt et al., 2009; Kisamore & Brannick, 2008), meta-regressions, Bayesian meta-analyses and meta-regressions, and meta-analyses of individual-level data to explore heterogeneity (e.g., National Research Council, 1992; DuMouchel, 1994; Sutton, 2002; Ioannidis, Patsopoulos, & Rothstein, 2008).

If the FE assumption of homogeneity is incorrect, the standard error of the mean effect size may be underestimated (i.e., the standard error will be too small), which leads to too-narrow confidence intervals and an overstatement of the precision of the FE meta-analytic results (National Research Council, 1992; Overton, 1998; DuMouchel, 1994; Kisamore & Brannick, 2008; Schimdt, Oh, & Hayes, 2009). Schmidt et al. (2009) empirically demonstrated

this in their re-analysis of 68 FE meta-analyses from five meta-analytic studies that were published in the *Psychological Bulletin* between 1988 and 2006. They discovered that the FE confidence interval widths around the meta-analytic mean effect size were actually much narrower than those confidence intervals from the RE model (the width of a 95% RE confidence interval was, on average, equivalent to 56% of the width of an FE confidence interval).

The standard test for the presence of homogeneity

Most meta-analytic studies report the values for the standard statistical test for the presence of homogeneity, which tests the hypothesis that τ , the standard deviation of the between-study variation in true effects, is equal to 0: $H_0: \tau = 0$. This equation can also be expressed as a test of the hypothesis of no variation (other than sampling error) among individual effect size estimates, θ_i, which is expressed as $H_0: \theta_i = \theta_i = \theta_i$(Konstantopoulos & Hedges, 2009). The null hypothesis is that all studies are evaluating the same effect and share the same underlying population effect size. The hypothesis of homogeneity of effect sizes is tested with *Q*, the homogeneity test statistic, the weighted sum of squares of the individual study effect sizes about the weighted mean effect size. Q has a chi-square distribution with $(k-1)$ degrees of freedom,

$$Q = \sum_{i=1}^{k} w_i \, (Y_i - \hat{\mu})^2 \qquad (2)$$

(Konstantopoulos & Hedges, 2009), under the null hypothesis of homogeneity of effect sizes. When this hypothesis is rejected, the observed variance is considered to be significantly greater than that expected by chance under the assumption that all of the studies share the same effect size (Shadish & Haddock, 2009).

In the past, some research synthetists relied on the results of the statistical homogeneity test as the criterion for selecting an FE or RE model and defaulted upon their meta-analytic model based solely upon the statistical significance of the test of homogeneity. If the null hypothesis of homogeneity were accepted (i.e., the *p* value for *Q* was non-significant), then,

according to this default-model decision-making framework, the researcher would proceed with an FE analysis assuming that there was one common underlying population-effect size; if the hypothesis was rejected then the researcher would default to an RE model (DuMouchel, 1994). However, it is incorrect to assume that the between-study variation component is equal to zero because the hypothesis was not rejected (DuMouchel & Normand, 2000). Numerous authors have discouraged defaulting upon a meta-analytic model based upon *Q* significance (National Research Council, 1992; Hardy & Thompson, 1998; Matt & Cook, 2009; Borenstein et al., 2009). A non-significant *Q* does not necessarily indicate that the studies are homogeneous (Spiegelhalter et al., 2004). Higgins et al. (2009) emphasize that it is preferable to measure the amount of heterogeneity, because testing whether or not heterogeneity exists addresses an unimportant question.

The use of the *Q*-statistic test has been widely recognized to have low power and thus lacks sensitivity to detect real between-study variation, especially when there is a small number (*k*) of studies in a meta-analysis (National Research Council, 1992; DuMouchel, 1994; Hardy & Thompson, 1998; Higgins et al., 2003; Jackson 2006; Shadish & Haddock, 2009; Donner & Paul, 1992; DuMouchel & Normand, 2000). Furthermore, the test has excessively high power to detect unimportant heterogeneity with many studies (Higgins & Thompson, 2002; Higgins et al., 2003). A further limitation of relying solely on the *Q*-statistic is that it only detects total overall heterogeneity and does not identify which particular individual studies make the greatest contribution to this variation.

Random Effects Models

Formal exploration of the between-study variation with random/mixed-effects modeling has increasingly been recognized as a necessary and worthwhile meta-analytic endeavor, because explanation of the variation will often result in a better and more thorough understanding of the treatment effect under investigation. Random/mixed-effects model explicitly account for the heterogeneity with a parameter that represents the between-study

variation. The underlying RE statistical assumptions are that the population from which the effect size is drawn is actually characterized by a distribution of effect sizes (Shadish et al., 2002) and that the true effects presented in each individual study are not identical but rather sampled from a distribution of true effects (Borenstein et al. 2009). Sutton and Abrams (2001) express the assumptions of the RE meta-analytic models as: $Y_i \sim N(\theta_i, \sigma_i^2)$, where Y_i is assumed to come from a normal distribution with a known sampling variance, and $\theta_i \sim N(u, \tau^2)$, where the true underlying effect sizes, θ_i, are assumed to come from a normal distribution of effect sizes with mean u and variance τ^2, which represents the between-study variation of each θ_i around μ. However, in practice, this normal distribution assumption for the underlying effects in individual studies is a strong assumption, which is often made without supporting evidence in favor of the assumption (Higgins, et al., 2009).

The simple (i.e., no covariates) RE model is expressed as

$$Y_i = \mu + \delta_i + \varepsilon_i, \quad (3)$$

where μ is the overall mean, δ_i, is the deviation of study *i*'s true effect from the grand mean, and ε_i is the error deviation of study *i*'s observed effect size from the true effect size (Borenstein et al., 2009).

The different assumptions of the FE and RE models (i.e., FE assumes $\tau = 0$ and RE allows $\tau > 0$) result in differing formulas for the standard error of the mean effect size, an important statistic that is used both in confidence interval computation and for significance testing of the mean (Schmidt et al., 2009).

Raudenbush (2009) recognizes RE models as advantageous because these modeling procedures help a) quantify heterogeneity in true effect sizes, b) include the between-study variation in confidence interval estimates, c) extend easily to investigate the ability of study-level variables (covariate) to account for variation, d) derive improved estimates of effect sizes in individual studies, and e) conceptualize the random effect in a manner that is consistent with the scientific goal of generalization.

RE estimates of summary effect and standard error.

In order to compute the overall RE weighted mean effect, the weighting scheme, w_i^*, that is assigned to each study is inversely proportional to its within study and between-study variance

$$w_i^* = \frac{1}{\sigma_i^2 + \tau^2} \quad (4)$$

Borenstein et al., 2009). These weights are then used to compute the overall summary mean effect, μ, where

$$\mu = \frac{\sum_{i=1}^{k} w_i^* Y_i}{\sum_{i=1}^{k} w_i^*} = \sum_i \frac{w_i^*}{\sum_i w_i^*} Y_i. \quad (5)$$

Here μ is equal to the sum of the products of the RE weights multiplied by each effect size divided by the sum of the weights (Borenstein, et al., 2009). The summary effect, μ, has a variance, V_μ^*, that is estimated as

$$V_\mu^* = \frac{1}{\sum_{i=1}^{k} w_i^*} \quad (6)$$

and the standard error of μ, $SE_{\mu*}$, is estimated as $SE_{\mu*} = \sqrt{V_\mu^*}$ (Borenstein et al., 2009). An RE 95% confidence interval about μ is expressed as $\mu \pm 1.96(SE_{\mu*})$ based on the normal distribution.

Knapp and Hartung adjustment to standard errors and test statistics.

In meta-analyses where *k* is small, it may be incorrect to assume that: (a) the estimate of τ^2 is equal to its estimated value and (b) $x_i\beta = x_i\hat{\beta}$, because such assumptions lead to confidence intervals for the regression coefficient that may be too small and hypothesis tests that are overly liberal (Raudenbush, 2009). Knapp and Hartung (2003) proposed a new method for adjusting the standard errors and test statistics of the regression coefficients in RE models in order to consider the uncertainty in the estimate of τ^2. With this method, a *t*-distribution with *k* – *p* degrees of freedom (*p* represents the number of regression parameters) is used to formulate the confidence intervals for the regression coefficients. The Knapp and Hartung method (2003)

allows the regression coefficients to have wider confidence intervals and permits more conservative hypothesis testing, while holding the nominal significance level. Viechtbauer (2010) indicates that the Knapp and Hartung adjustment will usually, but not always, lead to more conservative *p* values for the regression coefficients and reduced Type I error rates.

Prediction intervals.

Higgins, Thompson, and Spiegelhalter (2009) propose that one of the five important RE meta-analytic objectives is to calculate the predictive distribution for a new study by giving consideration to the whole distribution of effects. Such an objective is important because knowledge accumulated through research synthesis should be able to be applied to future applications. The formula for the prediction interval is an *approximate* interval that recognizes the imprecision in the estimate of τ^2 with the use of a *t*-distribution with *k* – 2 degrees of freedom:

$$\hat{\mu} \pm t_{k-2}^{\alpha}\sqrt{\{\tau^2 + \widehat{SE}(\hat{\mu})^2\}}. \quad (7)$$

This prediction interval allows for consideration of the uncertainty in the estimates of μ and τ and addresses the actual dispersion of the effect sizes (Borenstein et al., 2009). Prediction intervals can be distinguished from confidence intervals because confidence intervals quantify only the accuracy of the mean, reflect only the error of estimation in the mean, and approach a width of zero as the number of studies approaches infinity (Borenstein, et al. 2009). Many published meta-analyses to date have not made use of predictive distributions for the effect in a new study.

Mixed-Effects Model (the two-level hierarchical linear random-effects model)

Raudenbush and Bryk (2002) represent the RE meta-analytic model as a two-level hierarchical linear model, because meta-analytic data has an inherent hierarchical structure where the subjects are nested within studies. Raudenbush and Bryk describe the model as a two-stage sampling design where the sampling mechanism results in two components of

variance (e.g., random effect variance at the study level and estimation variance at the subject level). The random effect variance is the variance that arises as a result of sampling a random sample from a larger universe of studies that vary in their true effect sizes. The estimation variance occurs because each study's effect size estimate is based on a limited number of subjects (Raudenbush, 2009).

Level-1 model.

Raudenbush expresses the Level-1 (Within Studies) model as

$$Y_i = \theta_i + e_i, \quad (8)$$

where Y_i is the observed effect size estimate for study *i,* θ_i represents the true effect size for each of the $i = 1, \ldots, k$ studies, and e_i is the sampling error. The sampling errors, e_i are assumed to be statistically independent from each other, and they come from a normal distribution with a mean of zero and a known variance σ_i^2 , where σ_i^2 reflects the within-study sampling variance and the sample size of study *i*: $e_i \sim N(0, \sigma_i^2)$.

Level-2 model.

The Level-2 model includes study-level covariates (also termed effect modifiers, explanatory variables, and treatment interactions) that can be added to the meta-analytic model to help explain some of the heterogeneity, so that the estimate of τ^2 represents the remaining variation in θ_i that is not explained by the study covariates. In the Level-2 model, the true unknown effect size depends on both fixed study characteristics and the level-2 random effect (Raudenbush & Bryk, 2002). The Level-2 (Between-Studies) prediction model is expressed by Raudenbush (2009) as more general than other models:

$$\theta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_p x_{ip} + \delta_i. \quad (9)$$

Here β_0 represents the model intercept; $x_{i1}, \ldots, x_{ip}$ represents the coding of the study-level characteristics; $\beta_1, \ldots \beta_P$ are the regression coefficients, which can be used to predict differences in the individual study effect sizes θ_i; and δ_i is the random effect of the i^{th} study. The random

effect, $\delta_i \sim N(0, \tau^2)$ is usually assumed to come from a normal distribution with mean zero and variance, τ^2 (Raudenbush, 2009).

These two models can be combined into a mixed-effects linear model (also referred to as a hierarchical linear model or a generalized linear mixed model)

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_p x_{ip} + \delta_i + e_i \quad (10)$$

with the assumption that $\delta_i + e_i \sim N(0, v_i^*)$, where $v_i^* = \tau^2 + \sigma_i^2$ represents the total variance of the observed effect size (Raudenbush, 2009).

Estimation methods for the between-study variance.

Two types of point estimation methods that are commonly used to estimate τ^2 in random/mixed-effect meta-analytic models include DerSimonian and Laird (1986) and Restricted Maximum Likelihood Estimation (REML).

DerSimonian and Laird Method (Method of Moments).

The DerSimonian and Laird (1986) RE method is a method of moments based approach that is used as an estimator for τ^2. It is used widely in RE meta-analyses and it is provided as an option for estimating the between-study variance in popular meta-analytic software programs such as the commercially available, e.g., Comprehensive Meta-Analysis 2 (Borenstein, Hedges, Higgins, & Rothstein, 2004), and the openly available, e.g., Review Manager (Cochrane Collaboration, 2008) program and *metafor* (Viechtbauer, 2009) package. DerSimonian and Laird (D-L) RE models are often used for point estimation of the amount of heterogeneity in meta-analyses because of their simplicity (a non-iterative procedure) and ability to incorporate between-study differences in the overall effect size estimate (DerSimonian & Kacker, 2007). The results from this approach *do not* depend on the distributional assumption for the RE (Higgins et al., 2009) and, thus, unlike the REML method, the D-L approach offers a moment based approach that does not require distributional assumptions about the probability density of τ^2. However, a limitation of D-L is that the computation of prediction intervals for an effect in a

new study will not follow naturally from an estimation approach that does not require a distributional assumption for the random effect (Higgins, et al., 2009).

Method of moments provides consistent estimates, although they lack efficiency (Raudenbush, 2009). The standard error of the overall D -L effect size estimate is likely to be an underestimate of the true standard error because of the uncertainty that arises with this procedure (DerSimonian & Kacker, 2007). In the D-L method, τ^2 is estimated as

$$\tau^2 = \frac{Q-df}{C}, \quad (11)$$

where *Q* in the numerator of the equation is

$$Q = \sum_{i=1}^{k} w_i Y_i^2 - \frac{\left(\sum_{i=1}^{k} w_i Y_i\right)^2}{\sum_{i=1}^{k} w_i} \quad (12)$$

and *C* is

$$C = \sum w_i - \frac{\sum w_i^2}{\sum w_i} \quad (13)$$

(Borenstein et al., 2009).

Restricted maximum likelihood estimation.

REML is based on the normal likelihood and it provides an estimator of the heterogeneity, assuming a normal distribution for the random effect (Higgins et al., 2009). REML is a maximum-likelihood approach that is based on the likelihood of the variance, τ^2, given the data (Raudenbush, 2009). REML offers the advantage of considering the uncertainty about the unknown regression coefficients and thus yields nearly unbiased estimates and the possibility of improved standard errors (Raudenbush, 2009). Viechtbauer (2005) recommends the use of the REML estimator for the heterogeneity, because REML provides the best balance between unbiasedness and efficiency, as shown in his illustrative and simulated comparisons of REML, maximum-likelihood, D-L, Hunter-Schmidt, and Hedges estimators of τ^2. Furthermore, because REML is based on a normal distribution for the random effect, this method can be used to make inferences regarding the predicted effect in a new study with the computation of an

approximate prediction interval. REML also allows for the meta-analysis to address the important meta-analytic goal of estimating the study-specific true effect sizes, as will be discussed in a following section.

Estimators of study-specific true effect sizes.

Empirical Bayes (EB) estimators, can be used as part of a random/mixed-effects meta-analysis to represent an improved estimation of study-specific effects for each individual study by combining the fixed effects with the estimated contribution of the random effects (Raudenbush & Bryk, 1985). To date, many published meta-analyses have not made use of these improved study-specific estimates. An EB estimate assumes that τ is estimated empirically from the data, hence the nomenclature for *empirical* Bayes estimate (Raudenbush, 2009). EB estimates are known as *shrinkage estimates*, because they mathematically tend to shrink outlying values toward the overall mean (Morris, 1983) particularly with small values of τ (DuMouchel & Normand, 2000). Raudenbush, Bryk, and Congdon (2004) describe EB estimates as the optimal composite of an estimate because the estimate is based both on the data from the individual study and on the data from the ensemble of similar studies.

Raudenbush (2009) describes how EB estimates borrow strength in the form of reliability information from the entire ensemble of studies, and in cases of studies that have low reliability, the EB estimators borrow the greatest amount of strength from the other studies and experience the most shrinkage toward the overall mean. EB estimates describe the study-specific expectations of the parameter, θ_i, given y *and* τ. Raudenbush (2009) terms these EB estimates either *unconditional shrinkage estimators* or *conditional shrinkage estimators*, depending upon whether or not the estimate is based on an unconditional (i.e., without the inclusion of Level 2 study-level covariates) or conditional (with the inclusion of Level 2 study-level covariates) model. Raudenbush uses the following equation to estimate the average value of θ_i based on the simple RE (unconditional) model, $\theta_i = \mu + \delta_i$:

$$E(\theta_i|\, Y_i, \mu, \tau^2) = \theta_i^*(\tau) = \hat{\lambda}_i Y_i + (1 - \hat{\lambda}_i)\mu \quad (14)$$

When Level 2 study-level covariates are included in the model, the conditional model for shrinkage estimators is expressed as:

$$E(\theta_i|\, Y_i, X_i\, \mu, , \tau^2) = \theta_i^*(\tau) = \hat{\lambda}_i Y_i + (1 - \hat{\lambda}_i)(\mathrm{X}\beta) \quad (15)$$

and is based on the conditional model $\theta_i = x_i\beta + \delta_i$, (Raudenbush, 2009). In these equations, the measure of the reliability of Y_i as an estimate of θ_i is represented by $\hat{\lambda}_i$, the reliability of study *i*'s effect size. (Raudenbush & Bryk, 1985). The estimate of the between study variance, τ^2, is used to compute the reliability of study *i* as follows

$$\hat{\lambda}_i = {\tau^2}\Big/{\tau^2 + \sigma_i^2}. \quad (16)$$

Shrinkage estimates represent a compromise between varying amounts of heterogeneity and reliability. If studies were homogeneous (i.e., τ^2 were close to 0), then μ would be the best estimator for each θ_i (Raudenbush, 2009), because each study would be estimating the same common-effect size parameter and the meta-analytic model is similar to an FE model. However, if studies are heterogeneous (i.e., when τ^2 is large) then studies are not likely to borrow much strength from each other and each study is likely to be nearly equivalent its own individual value (Raudenbush, 2009).

Describing Heterogeneity.

According to Higgins and Thompson (2002), one of the most troublesome aspects that reviewers contend with when conducting a systematic review is addressing the statistical heterogeneity. An empirical study by Ionnidis et al. (2008) supports this premise. They analyzed Issue 4 of the *Cochrane Database of Systematic Reviews* (2005) and found that the typical reason that of 135 of the 1739 systematic which included forest plots failed to include the associated meta-analytic summary statistic, was typically due to large heterogeneity. Ioannidis et al. (2008) suggest that it may be preferable to investigate the differences among the studies

in a meta-analysis with RE, meta-regressions, Bayesian meta-analysis, and Bayesian meta-regressions. Regardless of the interpretative difficulties regarding heterogeneity, it is crucial for researchers to provide heterogeneity descriptive statistics, quantitative information, and investigation of sources of heterogeneity (Higgins & Thompson, 2002; Hardy & Thompson, 1998). Heterogeneity can arise from artifactual factors (i.e., reliability of measures) and from real between-study variation due to study features such as differences in a) populations, b) interventions and co-interventions, c) outcome measurement and timing, and d) research design and conduct (Glasziou & Sanders, 2002; Higgins & Thompson, 2002). The following section will review some of the recommended indexes and practices for estimating and describing heterogeneity.

Quantification of the uncertainty in the estimate of the heterogeneity.

Confidence intervals for the heterogeneity are useful for assessing the amount of precision in meta-analytic estimates, although the presence of such intervals remains relatively uncommon and their usefulness unappreciated in meta-analysis (Viechtbauer, 2007). Such intervals can be used for sensitivity analyses in order to reflect how estimates change as a function of increasing or decreasing values of τ^2 Viechtbauer (2007). Within the classical framework several methods have been proposed for quantifying the uncertainty in the heterogeneity such as Hardy and Thompson's (1996) profile likelihood based approach, Biggerstaff and Tweedie's (1997) approximation of the *Q* distribution, and bootstrap estimates of standard errors (Switzer, Paese, & Drasgow, 1992). Viechtbauer (2007) proposed the use of the *Q*-profile method as a preferred method for constructing confidence intervals about the heterogeneity. The *Q*-profile method is an iterative method that obtains the upper and lower critical values of the χ^2 distribution with *k* -1 degrees of freedom by profiling the generalized *Q*-statistic. Here, the generalized *Q*-statistic, $Q(\tau^2)$, is

$$Q(\tau^2) = \sum \frac{(Y_i - \hat{\mu})^2}{\tau^2 + \sigma_i^2}. \quad (17)$$

The upper and lower values of the 95% CI for τ^2 are found iteratively by repeatedly calculating $Q(\tau^2)$ over increasing τ^2 values (Viechtbauer, 2007). Viechtbauer (2007) systematically compared the *Q*-profile method to other methods for constructing confidence intervals about τ^2 (i.e., Biggerstaff and Tweedie, Profile Likelihood with ML /REML, Wald-type with ML/ REML, Sidik-Johnson, and bootstrap methods) and determined that the *Q*-profile method produced the most accurate coverage probabilities for the confidence intervals.

Inspection of Forest Plots for Visual Summaries of Dispersion.

Higgins et al. (2009) recommend the use of forest plots for a preliminary inspection of heterogeneity. Forest Plots offer a useful visual tool for reporting, interpreting, and summarizing the consistency of results across studies (Sutton & Higgins, 2008), and their use has been supported as an excellent practice for presenting meta-analytic findings (Borman & Grigg, 2009; Borenstein et al., 2009). Forest plots display the point estimate (i.e., effect size) for each individual study with a box centered within the corresponding horizontal lines, which represent the width of the confidence intervals (usually a 95% confidence interval) associated with each point estimate (Lewis & Clarke, 2001). The overall summary meta-analytic effect estimate is displayed with a diamond at the bottom of the plot along with its confidence interval. The width of the summary effect estimated is depicted by the width of its diamond (Lewis & Clarke, 2001) or with horizontal lines. Ideally, the size of the boxes that represent the point estimates should be proportional to the precision of that study so that it is easy to visualize which studies contributed the most weight to the overall meta-analytic findings (Lewis & Clarke, 2001; Borenstein et. al, 2009). Inspection of a forest plot allows the reader to conceptualize the extent of between-study variation by examining the confidence interval width and overlapping of individual studies and identify studies that may be outlying data points.

The I^2 Statistic.

Higgins and Thompson (2002) proposed using the descriptive statistic I^2, to describe the proportion of total variance in the estimates of effects between studies that is due to heterogeneity, rather than to chance. I^2 is independent of both the effect size metric (i.e., odds ratio or standardized mean differences) and the number of included meta-analytic studies (Higgins & Thompson, 2002). This makes it a preferable and suitable summary measure for quantifying the amount of heterogeneity (Sutton, 2002; Higgins & Thompson, 2002). I^2 is calculated as $I^2 = 100\% * \left(\frac{Q-(k-1)}{Q}\right)$. Higgins, Thompson, Deeks, & Altman (2003) proposed the following tentative suggestions for interpreting the values of the I^2 statistic: $I^2 = 75\%$ represents large heterogeneity; I^2 = 50% represents medium heterogeneity; and I^2 = 25% represents a small amount of heterogeneity.

Use of meta-regression (conditional models) to explain heterogeneity.

As discussed, the RE model can be expanded to a conditional model that includes study-level covariates (moderator variables) to help explain some of the RE variance. Information about study features such as characteristics of subjects, study design, intervention, research design, and methodology can be included in the expanded RE model as covariates. In such meta-regression models, τ^2 represents the residual variance after controlling for covariates. A variance-explained statistic can be calculated by comparing the conditional variance to the total variance with the following formula: $\frac{\tau^2(total)-\tau^2(given\ X)}{\tau^2(total)}$ (Raudenbush, 2009).

Disadvantages of RE models.

In a random/mixed-effects approach the random-effect variance is treated as if it were known even though it is estimated, and thus when τ is estimated from the data, the *uncertainty* of the estimate is not considered. The overall mean meta-analytic effect size estimate and regression coefficients are weighted estimates that are dependent upon the uncertainty in the

variance of the random effects (Raudenbush, 2009). Not considering the underlying uncertainty in the estimate of the RE variance may result in threats to the validity of statistical inference from the meta-analysis (Raudenbush, 2009). Meta-analytic estimates are often based on a limited number of data points (i.e., studies) and this further compromises the validity of the between-study variance estimate and the confidence intervals about μ. When the number of included studies (k) is small, when the sample size within studies (n) is small, or when the sampling variance v_i, is large, methods that estimate τ as a fixed value underestimate the standard errors and the corresponding confidence limits, which may result in inaccurate overall estimates of μ (DuMouchel, 1994; Raudenbush, 2009).

Furthermore, the RE assumption that Y_i is normally distributed about θ_i with a known variance is only a reasonable assumption when the sample size for each individual study is moderate to large, and this assumption is difficult to assess when the number of included studies is small (Raudenbush, 2009). In practice it is often not plausible to assume that the random effects are normally distributed with constant variance (Hardy & Thompson, 1998).

A Fully Bayesian Hierarchical Linear Model for Meta-analysis

The fully Bayesian (FB) approach to meta-analysis is a flexible and intuitive modeling approach that is becoming increasingly popular with researchers due to recent advances in computational methods and the relative advantages that Bayesian methods offer over more traditional meta-analytic methods (Sutton, 2000; Sutton & Abrams, 2001). Bayesian methods offer the advantage of encouraging the use of a unified and model-based approach to evidence synthesis (Sutton & Higgins, 2008; DuMouchel & Normand, 2000).

The Bayesian statistical philosophy is essentially about updating probabilities in light of new evidence and thus it translates well into the practice of quantitative research synthesis and updating of meta-analyses. The Bayesian approach distinguishes itself from traditional meta-analytic methods because Bayesian analyses emphasize estimation and prediction of parameters and uncertainty assessments (National Research Council, 1992), and the Bayesian approach fully takes into account the uncertainty about all the unknown model parameters (Raudenbush & Bryk, 2002).

When the uncertainty of τ is not considered in a meta-analytic model, as is the case in the classical random/mixed-effects models, it is possible that a treatment effect may be incorrectly identified as significant (DuMouchel & Normand, 2000). Comparative studies have shown that the problem with the EB RE approach is that, because the uncertainty in the RE variance estimate is not considered, τ may not be estimated accurately when there are a small number of studies (Spiegelhalter, et al. 2004).

Exchangeability.

The essential concept in the Bayesian approach to research synthesis is the notion of exchangeability of study effects (Higgins, et al. 2009). Within a Bayesian framework, study effects are considered to be similar to each other but not identical (Spiegelhalter et al., 2004; Higgins et al., 2009). Although the Bayesian model is similar to an RE model, it differs conceptually from the classical (frequentist) RE model in the exchangeability assumption and

the justification for the process that generates the random effects (Raudenbush, 2009). According to the Bayesian perspective, the random variation of the true effect sizes reflects the investigator's lack of knowledge (uncertainty) about the process that generates the random effects, while a traditional RE model specifies the sampling mechanism of the sampling studies from a larger population of studies as the source of random effects variance (Raudenbush, 2009).

In Bayesian statistics every unknown model parameter has its own probability distribution. This allows for direct probability statements (i.e., computation of the probability that an effect is greater than zero) and uncertainty estimates to be made about the data. Bayesian models may incorporate other relevant information about parameters that is external to the actual meta-analytic data but available to the researcher (Schmid, 2001). The researcher's probability beliefs about the external evidence can be modeled with a 'prior' quantitative summary of the variance that reflects the researcher's uncertainty about the mean of the true effect size. This prior evidence is then formally combined with the observed meta-analytic data (known as the likelihood) via the application of Bayes' theorem and merged into the *current state of knowledge* (Sutton et al., 2000) regarding the meta-analytic outcome or intervention. Bayesian methods address the question of how beliefs about an outcome change in light of the evidence generated by the new study or meta-analysis (Sutton et al., 2000), which makes Bayesian methods particularly suitable for updating meta-analytic data.

The fully Bayesian approach differs from the empirical Bayes approach because the empirical Bayes model assumes that τ^2 is equal to its estimated value (Raudenbush, 2009). With a FB approach uncertainty in the estimation is reflected in the prior probability distributions of the model parameters. An important implication of the use of fully Bayesian methods is that the widths of the FB confidence intervals for the fixed parameters may tend to be wider than those intervals from a classical analysis because Bayesian methods consider and model the full uncertainty of all of the parameters. Furthermore, the FB estimates of between-study variability

may be more appropriate than those estimates from non-Bayesian methods when the number of included meta-analytic studies is small.

The National Research Council (1992) expresses Bayes' theorem informally as the mathematical combination of prior probability distributions regarding belief about the certainty of parameter values with the actual data (the likelihood function) to yield a *posterior* distribution via the application of Bayes' theorem, which is expressed as

$$P(\theta|Data) = c\,P(\theta)P(Data|\theta). \quad (18)$$

Here $P(\theta)$ represents the prior probability density function regarding the unknown quantities of interest; $P(Data|\theta)$ denotes the ***likelihood function***, which represents information about the unknown quantity provided by the current data; and $P(\theta|Data)$ represents the combination of these two sources of information into the joint posterior probability distribution. The constant of proportionality, c, is $\int P(\theta)P(Data|\theta)d\theta$. This constant is required with Bayes' theorem so that $P(\theta|data)$ is a proper probability density function that integrates to one (National Research Council, 1992; Sutton et al., 2000).

With a Bayesian approach, the research synthetist can effectively consider and include small studies and extreme results (Smith, Spiegelhalter, & Thomas, 1995), while at the same time allowing for moderate violation of the statistical assumption that effect size estimates have normal distributions with known variances (DuMouchel,1994), which can be a restrictive assumption with some types of data. Bayesian models provide more accurate estimates of study-specific parameters, θ_i, by incorporating the information from all of the studies in a meta-analysis (i.e., by borrowing strength from the other studies) in order to provide a better estimate of each individual study's effect size.

The parameters that require estimation within a Bayesian conceptualization are: μ, τ, θ_i, and β (DuMouchel, 1994). DuMouchel (1994) uses the following equations for a Bayesian meta-analysis. The hierarchical model without covariates is expressed as

$$Y_i = \mu + \delta_i + \varepsilon_i \quad (19)$$

where the observed effect size estimate derived from the i^{th} study is denoted by Y_i. The effect size estimates from each study are assumed to be normally distributed with a known variance σ_i^2 , which is conditional on the true parameter value, $y_i|\ \theta_i \sim \mathrm{N}(\theta_i, s_i^2)$. The study-specific parameters, which are the expectation of Y_i, are represented by θ_i, where $\theta_i = \mu + \delta_i$. The random effect, δ_i, is assumed to be normally distributed with a mean of 0 and variance τ^2 represented in notation as $\delta_i \sim$ *Normal* $(0, \tau^2)$. The sampling error associated with Y_i is represented by ε_i. It is assumed to be normally distributed with a mean 0 and a *known* sampling variance: $\varepsilon_i \sim N(0, s_i^2)$. Both the random effects, δ_i, and the within-study sampling errors, ε_i, are assumed to be independent of each other and independent across studies.

Equation (19) can be easily be generalized and expanded to include study-level covariates (moderator variables) which represent fixed characteristics of the studies and are used to explain variation between studies:

$$y_i = [X_i\ \beta + \delta_i] + \varepsilon_i \quad (20)$$

$$\theta_i = X_i\ \beta + \delta_i. \quad (21)$$

DuMouchel (1994) uses the term $X_i\,\beta$ to replace μ; where $X_i\beta$ represents a linear combination:

$$X_i\ \beta = \beta_0 + \beta_1 x_{i1} + \ldots\ldots + \beta_J x_{iJ}. \quad (22)$$

There are three sources of variation to be estimated in a hierarchical Bayesian meta-analysis: a) s_i, the within-study random sampling error which is usually assumed to be known, b) β, the between-study differences that can be explained by fixed study-level characteristics at the second level of the hierarchical model, and c) τ, the standard deviation of the unexplained random variation due to differences between studies (DuMouchel, 1994; DuMouchel & Normand, 2000). In Bayesian hierarchical models it is τ, the standard deviation of the random effects variance, that plays a crucial role in assessing the uncertainty about μ and in predicting future θs (DuMouchel, 1994; Spiegelhalter, Abrams, & Myles, 2004).

Prior Distributions.

Bayesian hierarchical models specify probability distributions, termed prior distributions, for the unknown model parameters. Raudenbush and Bryk (2002) define a prior distribution as a distribution that expresses a researcher's beliefs about the values of parameters *prior* to any new data collection. However, Spiegelhalter et al. (2004) comment that the temporal relationship that the name *prior* conveys is actually a common misconception because it is plausible to choose a prior distribution after knowing the results of the study. The purpose of the prior is to summarize the evidence and the evidential uncertainty that is external to the study or meta-analysis under consideration (Spiegelhalter et al., 2004), and this summarization of the data with a prior distribution can be formulated subsequently after reviewing the data.

In a Bayesian model prior distributions are assigned for μ, τ, and β. The Bayesian hierarchical linear model can be expressed in notation as:

$$Y_i \sim N\left[\theta_i, s_i^2\right] \; i = 1, \ldots, k \tag{23}$$

$$\theta_i \sim N[\mu, \tau^2] \tag{24}$$

$$\mu \sim [-,-] \;\; \tau^2 \sim [-,-] \; \beta \sim [-,-], \tag{25}$$

where $\mu \sim [-,-]$, $\tau^2 \sim [-,-]$ and $\beta \sim [-,-]$ represent that various prior probability distributions that need to be specified (Sutton & Abrams, 2001).

Prior Distributions for the random effects variation.

It is through the specification of a prior distribution for τ that the Bayesian framework provides a technique for investigating the similarity of studies and the extent to which studies can borrow information from the entire ensemble of studies (Greenhouse & Iyengar, 2009). The choice of the prior distributions affects both the width of the credible interval estimates for μ and the amount of shrinkage imposed on the θ_i^* (Pauler & Wakefield, 2000) as well as the size and width of the credible intervals for θ_{new}. However, there is no single, generally-accepted, correct prior distribution that is used as a default or reference prior in Bayesian meta-analyses

(Spiegelhalter et al., 2004). For this reason a Bayesian analysis often includes specification from a community of prior distributions.

Reference priors for the random effect variation.

The five prior distributions that have been recognized as potential candidates for reference prior distributions on the RE variation (Spiegelhalter et al., 2004) will be discussed in the next section.

DuMouchel Prior for τ.

DuMouchel (1994) suggests the use of a diffuse, highly-dispersed proper prior distribution for τ:

$$\pi(\tau) = \frac{s_0}{(s_0+\tau)^2}, \quad (26)$$

where s_0 represents the median of the posterior density. In this log-logistic prior distribution, s_0^2 represents the harmonic mean of the *K* sampling variances, s_i^2,

$$s_0^2 = {}^{K}\!/_{\sum s_i^{-2}}, \quad (27)$$

and s_i^{-2} represents the precision of each study. DuMouchel's prior is often advantageous because DuMouchel's prior is extremely dispersed (DuMouchel, 1994), and this dispersion protects against the tendency of the uniform prior to have a posterior distribution that is highly skewed toward large values of τ (Sutton & Abrams, 2001). This prior distribution for τ also has the advantage of having a maximum at zero with a decreasing function of τ, which permits values of τ to be near 0 as a strong possibility (DuMouchel & Normand, 2000). Furthermore, the DuMouchel prior provides some protection against the unwanted influence that imprecise studies may have on s_i , because s_0 tends to be weighted toward smaller sizes of s_i when the study standard deviations are not equal (DuMouchel & Normand, 2000).

Prior distributions can be specified in such a way that so that FE and RE models become special cases of the HBLM (DuMouchel & Normand, 2000). For example, the meta-analytic model can reduce to the equivalent of an FE model when the prior for τ is set near the

value of $\tau = 0$, so that there is assumed to be one underlying common effect. Alternatively, when the number of studies in the meta-analysis is very large or when the prior for τ is concentrated around the estimate of τ, then the meta-analytic model becomes equivalent to a RE model. For sensitivity analyses, DuMouchel (1994) recommends scaling prior median by a factor of 3 in both directions (i.e., $\tau_0 = \frac{s_0}{3}$ and $\tau_0 = 3s_0$) where the first and third quartiles for τ are equivalent to $\frac{s_0}{3}$ and $3s_0$. DuMouchel's recommendation of scaling by a factor of 3 allows the prior distribution to represent a continuum of FE versus RE beliefs about τ and the pooling of data. For example, the first, second, and third quartiles represent a 25% probability of essentially complete shrinkage ($\tau_0 = \frac{s_0}{3}$), a 50% probability of that studies borrow strength from each other, and a 25% probability of virtually no shrinkage ($\tau_0 = 3s_0$) (Greenhouse & Iyengar, 2009).

Uniform prior distribution for τ^2.

A uniform prior distribution for the between-study variance such as $\tau^2 \sim Uniform\ (0, 1000)$, gives preference to large values of τ^2 (Spiegelhalter, et al. 2004). Here, the mean is equal to 0 and the variance is equal to 100.

Uniform prior for τ.

A uniform prior can be specified on τ such as $\tau \sim Uniform\ (0, 5)$, where small values of τ are as equally likely as large values of τ (Spiegelhalter et al., 2004). Here, the mean is equal to zero and the variance is equal to 5.

Inverse Gamma Prior on τ^2.

Gamma distributions (or inverse gamma distributions on τ^2) are convenient and flexible mathematical distributions that are a popular choice for prior distributions of the precision (Spiegelhalter et al., 2004). Assuming an inverse gamma distribution for the between-study heterogeneity is beneficial because this distribution can accommodate a wide range of *a priori* informational situations that may be available regarding prior beliefs (Abrams & Sanso, 1998).

The use of a gamma distribution on the precision is equivalent to the use of an inverse gamma distribution for τ^2 since the precision is equal to the inverse of the variance. The following inverse gamma prior on the precision, $\frac{1}{\tau^2} \sim$ Gamma(.001, .001) is a proper distribution that gives greater weight to values of τ near 0 (Spiegelhalter et al., 2004).

A Subjective Distribution for τ: Half-normal distribution.

Spiegelhalter et al., 2004 classify the half-normal prior distribution about τ as a subjective distribution because it allows for some subjective input from the researcher as to the location of a feasible upper bound for τ. This distribution is constrained to be positive, so it represents the positive half of a normal distribution with a mode at 0 and steadily decreases in value (Spiegelhalter et al., 2004): $HN\left[\left({\tau_U}/{1.96}\right)^2\right]$. Here, τ_U represents the upper 95% boundary of τ.

Prior Distributions for the overall mean and the effect of covariates.

A non-informative prior for μ and β.

For the fixed effect parameters of μ and β, flat and diffuse priors are the standard (Pauler & Wakefield, 2000) in the absence of any prior information (DuMouchel & Normand, 2000). For μ and β the prior distribution typically comes from a normal distribution with a mean of m and a variance of d^2:

$$\mu \sim N\ (m, d^2\ \rightarrow\ \infty) \tag{28}$$

$$\beta_j\ \sim N(m_j, {d_j}^2). \tag{29}$$

Often the following values are chosen for the prior distribution of β: a) $m = 0$ and b) d^2 is set to be very large, so that the prior distribution is non-informative (DuMouchel, 1994). Such diffuse priors are often termed reference, vague, and non-informative. The specification of a prior with a large variance allows for a range of values that is wide enough to encompass all plausible values of θ (Spiegelhalter, Abrams, & Myles, 2004). The use of a vague prior is common

Bayesian practice because it allows for greater objectivity regarding the overall treatment effect (Sutton, 2001) and aids in estimation of the location parameter in the absence of other external information regarding μ (DuMouchel, 1994).

An alternative to the normal distribution for the random effects.

A t-distribution with heavier tails.

In hierarchical models it is frequently assumed that the random effect follows a normal distribution (Lee & Thompson, 2007). This normalcy assumption for the RE is a strong assumption about the random effects, which is typically made without supporting evidence in favor of the assumption (Higgins, et al., 2009). Such an assumption for the RE may be restrictive and invalid particularly when there are a number of outlying data points. Instead of using a normal distribution to model the RE, a more flexible distribution such as the *t*-distribution with 4 degrees of freedom may also be used to model the RE (Higgins, et al., 2009) with the use of Bayesian modeling. The *t*-distribution with 4 degrees of freedom is somewhat similar in shape to the normal distribution, but it has heavier tails and reduced concentration around the mean of the distribution. The use of a *t*-distribution for the random effects may reduce the influence of outlying studies and aid in the estimation of predictive distributions (Lee & Thompson, 2007).

Posterior Probability Distributions.

Posterior probability distributions are estimated via the application of Bayes' Theorem for the parameters (i.e., $\tau^2, \mu,$ and β) given the data from studies $y_i \ldots y_k$. The posterior probability distribution represents the conditional distribution of the unknown quantity of interest, given the data. The posterior distribution is obtained by multiplying the prior probability density function by the likelihood function that represents information about the unknown quantity provided by the current data. The posterior density function is used in a Bayesian analysis for all inferences made regarding the unknown quantities of interest (Sutton et al. 2000).

It is important to note that Bayesian meta-analytic results are especially dependent upon the posterior distribution of the random effect (DuMouchel, 1994). Posterior distributions for the model parameters are best understood and displayed with trace plots that graph the posterior expectation of μ and θ_i, given the posterior distribution of τ (DuMouchel, 1994). The posterior distribution of τ is often skewed and because of the skewness, the median of the distribution is commonly used for point estimation instead of the mean (Higgins et al., 2009).

Markov Chain Monte Carlo Methods (MCMC)

Computation of the posterior probability distributions requires integral calculus (i.e., calculation of the area under the curve of f(x)). Such integration can be exceptionally difficult and complex particularly when additional unknown parameters, termed nuisance parameters, are present (Spiegelhalter et al. 2004), thus requiring the integrals to be evaluated over several dimensions. In such situations, posterior distributions are best calculated with computer-based simulation methods such as Monte Carlo methods that evaluate these complex integrations via simulation rather than algebraic analysis (Spiegelhalter, et al. 2004). Gibbs sampling is a type of MCMC method that successively samples variables from the posterior conditional distributions of each parameter (Sutton & Abrams, 2001; WinBUGS User Manual, 2003). With this method the unknown quantities are given initial values and successive samples are obtained from the conditional distribution of each variable, given the current sampled value of the other variables, with the premise that sampling will eventually occur from the correct posterior distribution of the unknown parameters (Smith, Spiegelhalter, & Thomas, 1995).

Credible Intervals

In Bayesian theory, intervals containing 95% probability are termed *credible* or *posterior* intervals. Bayesian 95% credibility intervals can be distinguished from the traditional 95% Neyman-Pearson confidence interval in several important ways (Spiegelhalter et al., 2004). The Bayesian 95% probability interval is interpreted as the 95% probability that the true underlying θ

lies in the 95% Bayesian credible or posterior interval, whereas the traditional 95% confidence interval is theorized to represent a long repeated series of confidence intervals in which 95% of these intervals should contain the true underlying parameter value (Spiegelhalter et al., 2004). Furthermore, Bayesian credibility intervals can be narrower than traditional confidence intervals as a result of the addition of prior information into the conceptual framework of the meta-analytic model (Spiegelhalter et al., 2004).

Probabilities for the parameters.

The Bayesian framework offers the advantage of determining the probability that a parameter is less or greater than a specific value (e.g., the probability that $\mu, \beta,$ or $\theta_{new} > 0$) with the use of posterior distributions for the parameters. In a Bayesian analysis, the probabilities are estimated as the proportion of MCMC iterations in which the parameter is greater than a pre-specified value (Higgins et al., 2009). Higgins et al. (2009) support the computation of posterior probabilities as a good alternative to classical meta-analysis hypothesis testing.

Predictive Distributions.

Bayesian methods offer the unique advantage of being able to predict a treatment effect for a new trial, θ_{new}, by using a full Bayesian predictive distribution,

$$p(\theta_{new}|data) = \int p\,(\theta_{new}|\mu,\tau)p(\mu,\tau|data)d\mu d\tau. \quad (30)$$

MCMC methods are used to simulate values of θ_{new} at each iteration from a fully predictive distribution (Spiegelhalter et al., 2004). The posterior predictive distributions reflect both sampling uncertainty and parametric uncertainty (Lynch & Western, 2004). It can be argued that Bayesian predictive distribution for θ_{new} offer a more appropriate summary of the treatment effect rather than the conclusions made regarding the weighted mean effect size, because the Bayesian full predictive distribution considers the uncertainty both in μ and τ (Spiegelhalter et al, 2004). Higgins et al. (2009) endorse the use of predictive distributions because such

distributions offer the most sensible way to summarize the results from a heterogeneous group of meta-analytic studies by giving consideration to the whole distribution of the effects.

Fully Bayesian Estimates of study-specific effects.

A unique and noteworthy advantage of Bayesian meta-analytic models is that they provide fully Bayesian shrinkage estimates of study-specific effects. Fully Bayesian shrinkage estimates, θ_i^*, assume that there is full uncertainty in τ:

$$\theta_i^* = E[\theta_i \mid y] = \int \left\{ \mu^*(\tau) + [y_i - \mu^* \tau] \frac{\tau^2}{\tau^2 + s_i^2} \right\} \pi(\tau|y)\, d\tau. \quad (31)$$

DuMouchel and Normand (2000) describe a fully Bayesian shrinkage estimators as a compromise of the observed effect size, Y_i , and the prior mean where the shrinkage factor or weight for the i^{th} study is represented by:

$$W_i(\tau) = \frac{s_i^2}{s_i^2 + \tau^2} \quad (32)$$

Like empirical-Bayes estimates, fully Bayesian shrinkage estimates represent a weighted average of Y_i and μ, except fully Bayesian methods consider the uncertainty of the estimate of τ. DuMouchel and Normand (2000) suggest the use of trace plots as a means of visualizing how fully Bayesian shrinkage estimates vary as a function of τ. The majority of published meta-analyses have not employed trace plots for depicting the dependency of meta-analytic results on varying values of τ.

Bayesian Meta-analytic Model Checking

Sensitivity Analyses of Results to Different Prior Distributions.

Prior distributions can influence the overall stability of meta-analytic results particularly when the prior is placed on a scale parameter in a meta-analysis of a small number of studies (Lambert, Sutton, Burton, Abrams, & Jones, 2005) or when studies are included that have extreme effect size values (Smith, Spiegelhalter, & Thomas, 1994). For this reason, it is strongly recommended that sensitivity analyses be conducted to compare the posterior

distribution of results across the different collection of priors that represent varying degrees of beliefs about the values of τ. If the inferences made from the meta-analysis do not vary among the different prior distributions, then it can be concluded that the Bayesian inferences are robust to prior specification (Greenhouse & Iyengar, 2009).

A Bayesian Measure of Model Fit.

The Deviance Information Criteria (DIC) was proposed by Spiegelhalter, Best, Carlin, and van Der Linde (2002) as a pragmatic and relatively simple measure for assessing the complexity and fit of a Bayesian model and for comparison among models. The DIC equation, as defined by Spiegelhalter et al. (2002), is

$$\text{DIC} = \overline{D} + P_D, \tag{33}$$

where $\overline{D}$ is the posterior mean deviance that is used as a Bayesian measure of model fit to which a penalty term for complexity, p_D, is added, where p_D represents twice the effective number of model parameters. The Bayesian model with the smallest DIC value is typically considered to be the model that would best predict a replicate data set of the same structure as that of the observed data (Spiegelhalter, Thomas, Best, & Lunn 2003). Spiegelhalter et al. (2003) advise that the DIC may not be an appropriate measure of Bayesian model adequacy when the posterior mean of the deviance has extreme skewness or bimodality.

Cross-Validation and Detection of Outlying Data Points.

DuMouchel & Normand (2000) recommend the use of cross-validation for checking Bayesian model assumptions, model fit, and the presence of outliers. Bayesian leave-one-out cross-validation is a useful technique for illustrating the posterior predictive distribution with all of the data except for the value of the leave-one-out Y_i. In this procedure, the cross-validated predictive probability of the leave-one-out Y_i describes the probability value of the predictive cumulative distribution evaluated at the leave-one-out Y_i (DuMouchel & Normand, 2000). Cross-validation of the residuals can be conducted so that the residuals from the predictive

distribution of each data point are obtained by analyzing the data with that data point deleted from the meta-analysis. A Bayesian cross-validated *Q*-*Q* plot of the normalized residuals can be used for a visual inspection to determine if there is evidence against the assumption of the normality of the random effects (DuMouchel & Normand, 2000).

In order to detect if the largest absolute residual data point is exceptionally different from the other values in the meta-analysis, the Bonferroni statistic of the probability of the event, can be calculated for the largest cross-validated residual (DuMouchel & Normand, 2000). If outlying studies are identified, DuMouchel and Normand (2000) recommend the sensitivity analyses strategies of removing the outlying study and/or increasing the s_i of that study by a factor of two to determine the impact on overall meta-analytic inferences.

Advantages of Bayesian Methods for Research Synthesis

Bayesian research synthesis methods offer many desirable modeling properties over more traditional meta-analytic methods particularly in the typical case of a meta-analysis of a small number of studies. Schmid (2001) supports the use of Bayesian models because Bayesian models provide a statistically informative summary of the parameters, incorporate all sources of variation into one model, and do not require normal distributions. Furthermore, Sutton & Abrams (2001) and Sutton et al. (2000) recognize the following advantages of Bayesian methods, because these methods offer: a) full modeling of parameter uncertainty, b) inclusion of the totality of evidence by allowing the consideration of other pertinent evidence (i.e. non-randomized evidence or expert opinion) that may otherwise be excluded by traditional methods, and c) flexibility and extendibility with more complex data. Additional advantages offered by hierarchical Bayesian modeling that Sutton et al. (2000) and Spiegelhalter et al. (2004) mention include the ability of Bayesian models to offer:

1. *Unified modeling approaches:* This unified modeling approach resolves the controversy regarding FE or RE model choice because the between-study variance can be explicitly modeled with Bayesian methods.
2. *Strength from other studies:* Bayesian methods provide updated and improved individual study-specific estimates that borrow strength from each other, leading to shrinkage of the study-specific estimate toward the overall mean along with a reduced confidence intervals for the study-specific estimates.
3. *Direct probability statements:* Probability statements about quantities of interest can be directly expressed (i.e., the posterior probability that the true treatment effect is less than 0).
4. *Predictions for subsequent trials:* A treatment effect for a new trial, θ^{new}, can be predicted with the use of a full Bayesian predictive distributions.
5. *Meta-regressions to model study-level differences:* The relationship between study-level covariates and effect-size estimates can easily be explored with a Bayesian hierarchical linear model.

Disadvantages of Bayesian methods

Sutton et al. (2000) list the following disadvantages of applying Bayesian methods to meta-analysis: a) the principle of scientific objectivity may be considered to be undermined with the specification of informative prior distributions for μ and τ^2; b) different prior distributions may yield different conclusions with a small *k*, d) sensitivity analyses of the use of different priors is always required with a Bayesian analysis; d) estimation of the parameters of the posterior distribution is computationally complex; and e) there will be small differences in parameter estimates with traditional, EB, and FB (non-informative prior) methods *when* the number of studies in the meta-analysis is large. Furthermore, the application of Bayesian hierarchical methods requires an adequate understanding of meta-analytic modeling and sophisticated

computer programming, and the successful application of these methods may be exceedingly difficult for the typical non-statistician researchers who are interested in conducting a systematic review.

Chapter III: Method

This research project involved the re-analysis of five previously published meta-analyses using hierarchical Bayesian linear models and classical meta-analytic models in order to determine whether: the use of Bayesian methods produced significantly different meta-analytic conclusions from those made from the classical fixed-effects (FE), random-effects (RE), and mixed-effects model and to demonstrate how the estimation of meta-analytic parameters can be improved with Bayesian methods. The Bayesian approach, results, advantages and disadvantages, and ability to address the relevant meta-analytic hypotheses of Higgins et al. (2009) were compared to the classical meta-analytic models. The results of the different classical meta-analytic models (i.e., classical FE, RE, and mixed-effects models) were compared to the fully Bayesian models in terms of parameter estimates, standard errors, confidence/credibility intervals, predictive intervals, and the ability of the models to address the following five objectives proposed by Higgins et al. (2009):

1. Quantification of heterogeneity,
2. Estimation of underlying mean effect,
3. Estimation of study-specific effects,
4. Prediction of an effect in a new study, and
5. A test of whether an effect exists in a new study or in any study and whether the effect has a consistent direction.

In addition, the efficacy of classical and Bayesian models were examined for addressing a proposed, sixth important meta-analytic objective: illustrating how meta-analytic inferences change depending upon uncertainty in the estimate of heterogeneity. In order to meet this sixth objective, a classification system similar to the guidelines proposed by Rothstein, Sutton, and Borenstein (2007) for describing the impact of publication bias was used. Here, the impact of how meta-analytic results may change depending upon the uncertainty in the heterogeneity was classified with the use of qualitative indicators similar to those used by Rothstein et al. (2007), where the impact is described among τ^2 sensitivity analyses as: (a) “minimal”, when the basic meta-analytic fitted model and estimates are similar; (b) “modest”

when the fitted model remains the same but the estimates change moderately; and (c) “severe” when both the fitted model and estimates differ substantially.

A primary goal of this research project was to determine if the relevant meta-analytic objectives could validly be met within a practical, classical RE framework or if the Bayesian methods, which are computationally more complex and labor intensive, are the most appropriate framework to use for meta-analyses for these selected data sets. The results are presented in a way that places particular emphasis upon a comparison of the classical random/mixed-effects model with REML estimation to the Bayesian hierarchical linear models.

By re-analyzing meta-analyses using a Bayesian framework, this project sought to show that Bayesian methods offer a superior ability to address relevant RE meta-analytic hypotheses by allowing for: (a) the flexible modeling of between-study variation and the random effect; (b) the inclusion of study-level covariates into the model to account for some variation; (c) the benefits of fully Bayesian, θ_i^*, study-specific shrinkage estimates; (d) the computation of the posterior probabilities of parameter values; and (e)the computation of predictive distributions for a new study.

The meta-analytic data sets selected for the Bayesian re-analyses met the following criteria: they (a) included a sufficiently small number of studies and effect size estimates, defined by $5 \leq k \leq 55$; (b) used either an FE or simple RE model; (c) had a large amount of heterogeneity, defined as I^2 values greater than 70%; (d) had one or more potential study-level covariates (i.e., moderator variables) that could be considered for inclusion in a mixed-effects model; and (e) systematically reviewed a topic relevant to education, psychology, or healthcare interventions. Furthermore, the published meta-analyses must have reported the following necessary data requirements: individual study effect sizes, Y_i; the variance of the individual study effect size, s_i^2, or 95% confidence intervals about the average mean effect sizes; and information regarding potential study moderators, so that the meta-analytic computations could easily be replicated without having to retrieve the information from the primary studies within the

original meta-analysis. The effect sizes that were used for the illustrative examples were identical to those from the original meta-analytic data. For example, artifact corrections were not made to the meta-analytic data unless artifact corrections were made to the original meta-analytic data.

Description of Classical Analyses

The primary focus of the classical re-analyses was to investigate the ability of the classical random/mixed-effects model with REML estimation to address the relevant meta-analytic objectives proposed by Higgins et al. (2009), because it was hypothesized that this model may be the most appropriate meta-analytic model to use within a classical framework. Thus, the results are displayed in a way that highlights the estimates from the classical mixed-effects REML model as compared to those from hierarchical Bayesian linear models.

For the classical meta-analyses the data sets were analyzed using the statistical software program R (version 2.10) with the *metafor* (Viechtbauer, 2009) package and the Classical Meta-Analysis 2 program (Borenstein, Hedges, Higgins, & Rothstein, 2004). The R code for the classical meta-analyses can be found in the Appendix. Initially, the meta-analytic results that were obtained via the model used in the original meta-analysis were replicated first in R and in Comprehensive Meta-analysis 2 (CMA 2) in order to ensure that the meta-analytic results obtained here were identical to those from the original meta-analyses.

For the purposes of sensitivity analyses, FE and D-L model were also fit for the simple (i.e., no covariates) models with the use of the *metafor* package in R. For the mixed-effect models, REML estimation methods were used both with and without the Knapp and Hartung adjustment (2003) made to the standard errors and test statistics of the regression coefficients. Furthermore, an additional sensitivity analysis was conducted with the REML mixed-effect models in order to determine how sensitive the models were to different values of τ^2 by fixing

the value of τ^2 to both the lowest ($REML_{Low}$) and highest bound ($REML_{High}$) of its 95% CI from the Viechtbauer (2007) *Q*-profile method.

Backward stepwise regression model selection procedures were conducted in order to identify the significant study-level covariates for inclusion in the final mixed-model. With this approach the mixed-effects models originally included all of the potential covariates (a full model), and then the analyses proceeded with backward deletion in which the covariates were sequentially removed from the model one by one. The removal process involved sequential elimination of the least significant covariate (the covariate with the highest *p* value) after which the mixed-effects analyses were rerun in *metafor* using REML until a final model that included only the significant study-level covariate(s) was obtained. For this research there are some important caveats to consider regarding the use of the backward elimination for model selection and the interpretation of the significance of the covariates in meta-regression. These include: (a) the increased probability of making type I or type II errors, (b) the lack of consideration of interaction terms as potential predictors, (c) the small number of meta-analytic studies (*k*) to estimate all of the model parameters, and (d) the multicolinearity that may exist among two or more of the study-level covariates. As proposed by Higgins and Thompson (2004), permutation tests for assessing the statistical significance of the covariates were were also conducted in R with the *metafor* package in order to mitigate the false-positive rates that may occur in meta-regression. The *p* values for the the permutation tests were compared to those of the Knapp and Hartung (2003) adjustment.

Calculation of the power of the tests for moderators of effect sizes.

When conducting meta-regression, it is important to note that there may not be sufficient power to detect the presence of the significant moderator variables that predict effect sizes, even if such “true” effects do exist. Hedges and Pigott (2004) recommend that statistical power for the test for moderator variables be computed before conducting tests of moderator variables. If the computed power is low, then Hedges and Pigott recommend that researchers use one of

the following techniques to aid in the interpretation of the moderator analyses: report the power of the moderator tests explicitly or avoid conducting that moderator test. This project took the first approach in which the power of the regression coefficients for the mixed-effects model was computed in R using a two-sided test (α = .05) given by

$$1 - \Phi\left(c\alpha_{/2} - \beta_j \Big/ \sqrt{\sigma^*_{ii}}\right) + \Phi\left(-c\alpha_{/2} - \beta_j \Big/ \sqrt{\sigma^*_{ii}}\right) \quad (34)$$

(Hedges and Pigott, 2004). Here, Φ represents the cumulative distribution function of the normal distribution, $c\alpha_{/2}$= 1.96, β_j represents the difference considered to be of minimal substantive importance (set to 0.333), and $\sqrt{\sigma^*_{ii}}$ is the standard error for the $\beta_j{}^*$ as computed from the diagonal of the covariance matrix.

Measures of model fit.

Within the classical framework several classical models were fit to the same data set and model selection indices were used to compare among the models. The following model fit statistics were computed in R with the *metafor* package in order to compare the different classical models and assess the adequacy of the meta-analytic models: (a) log-likelihood, (b) deviance, (c) Akaike information criterion (AIC), and (d) Bayesian information criterion (BIC). The deviance is equal to -2 times the log-likelihood, which is -2 times the logarithm of the probability of the data, given the model parameters that are estimated (Gelman & Hill, 2007). The AIC is $2k - 2\ln(L)$, where L is the maximized likelihood and k represents the number of model parameters. The BIC is $-2\ln(L) + k\ln(n)$, where n represents the number of studies. For the classical meta-analyses the models with the lowest AIC and BIC values were considered to be the models that best supported the data structure.

Assessment of publication bias.

It is important to assess for the presence of publication bias because meta-analytic results and conclusions need to be interpreted with caution when bias exists (Rothstein, Sutton,

& Borenstein, 2005). Meta-analytic bias may occur when studies with non-significant results remain unpublished and, hence, the meta-analysis suffers from a "file drawer problem" (Rosenthal, 1979) in which the non-significant studies remain relegated to a researcher's file drawer. Within a classical framework, publication bias was examined with the *metafor* package in R with the use of funnel plots, the Egger regression test for funnel plot asymmetry (Egger, Davey Smith, Schneider, & Minder, 1987), the Orwin (1983), Fail-safe *N* method, and the Duval and Tweedie (2000) trim and fill method. Funnel plots were graphed and evaluated for symmetry for the fitted meta-analytic models in which the funnel plot displays the residual values plotted against their corresponding standard errors. Egger's regression test for funnel plot asymmetry was used with se_i as a predictor in order to determine if there was a relationship between the observed effect sizes, Y_i, and their corresponding standard errors, se_i (Viechtbauer, 2010). Two of the methods for examining publication bias, Orwin Fail-safe *N* values and the trim and fill method, could only be used with the simple RE models with the *metafor* package. Orwin Fail-safe *N* values were computed with a simple (i.e., with no covariates) unweighted model. Orwin Fail-safe *N* represents the number of missing studies averaging null results, which, if added to the observed meta-analytic data set, would reduce the unweighted average effect size by half**.** The Duval and Tweedie (2000) trim and fill method was used with the simple RE model as a technique for inferring the number of studies that may be missing from a meta-analysis on one side of the funnel plot.

Inspection of the data for outlying data points.

The meta-analyses were examined for the presence of outliers using the final model with REML by inspecting: (a) the mixed-effects forest plots, (b) the normal Q-Q plots, and (c) the leave-one out case diagnostics with the use of the *influence.rma* function in *metafor*. The *influence* function computes the leave-one-out case diagnostics that indicate the influence on model fit of deleting one case at a time from the model (Viechtbauer, 2009). For each meta-analysis the following leave-one-out case diagnostic measures were examined with the i^{th} study

was deleted from the model: (a) externally standardized residuals, (b) covariance ratios for values below 1.00, (c) DFBeta values indicating the extent in standard deviation units by which the estimated model coefficients change, (d) τ^2 values, and (e) the *Q* test statistics for the residual heterogeneity, for when the i^{th} study is omitted from the model. The externally standardized residuals were computed in *metafor* by fitting the model with the i^{th} study removed from the data set by first calculating the predicted value for the i^{th} study, then calculating the difference between the predicted and observed value for the i^{th} study, and finally standardizing the residual (Viechtbauer, 2009). Viechtbauer (2009) defines the covariance ratio as the determinant of the variance-covariance matrix of the parameter estimates with the i^{th} study deleted divided by the determinant of the variance-covariance matrix of the parameter estimates for the entire data set.

In addition, a dummy-coded variable for the outlier study was created and the significance of the regression coefficient for the dummy-coded outlier study was examined after controlling for all other study-level covariates. If an outlying data point was identified, a sensitivity analysis was conducted by leaving that data point out of that analysis to determine if the meta-analytic inferences changed with the exclusion of the outlier.

Assessment of between-study heterogeneity.

The estimated values of the heterogeneity (or residual variance) and the confidence intervals for the heterogeneity were obtained for both the simple and mixed-effects models iteratively, with REML using the Viechtbauer (2007) *Q*-profile method.

Values for I^2, a descriptive statistic that represents the proportion of total variation in effect size estimates that is due to heterogeneity rather than chance (Higgins & Thompson, 2002), were calculated for the simple RE models. Values for the *Q* heterogeneity test statistic, which is compared to a chi-square distribution, were reported for the random/mixed effects models.

Estimators of study-specific effects.

Empirical Bayes estimates of study-specific effects (described by Viechtbauer, 2009 as best linear unbiased predictors), $\theta_i^*(\tau)$, were computed in R with the *metafor* package by extracting the study-specific estimate through the combination of the fitted values based on the fixed effects in the model and the estimated contributions of the random effect (Viechtbauer, 2009). Viechtbauer (2009) refers to these estimates as best linear unbiased predictors (BLUPS), although technically these estimates are not unbiased because in general shrinkage estimates are always biased due to the shrunken nature of the estimate.

Predictive intervals for a new study.

Prediction intervals for an effect in a new study, $\hat{\beta} \pm t_{k-2}^{\alpha}\sqrt{\left\{\tau^2 + \widehat{SE}\left(\hat{\beta}\right)^2\right\}}$, for the estimates of regression coefficients were computed in R for each REML model but not for the FE and D-L models, because these estimation methods do not make a distributional assumption for the random effect (Higgins et al., 2009) and it would therefore not be appropriate to compute a prediction interval from such models.

Description of Bayesian analyses

The fully Bayesian Windows-based program, WinBUGS (Spiegelhalter et al., 2003), which uses Gibbs sampling (a type of MCMC simulation), was used to implement the Bayesian analyses. The R2WinBUGS (Sturtz, Ligges, & Gelman, 2005) package was used to run WinBUGS from R. Inserting the WinBUGS package into R offered the programming advantages of: (a) running WinBUGS in batch mode using scripts, (b) performing data manipulation in R prior to calling WinBUGS, (c) importing the WinBUGS results back into R, (d) updating WinBUGS models (Sturtz, Ligges, & Gelman, 2005), and (e) facilitating specification of the multiple models. Covariates were selected in WinBUGS using a backward model selection procedure with the commonly used baseline uniform prior on τ (Spiegelhalter, et al., 2004) and with $\delta_i \sim Normal$. The posterior probabilities that the covariates were greater than zero were

computed, and study-level covariates were sequentially removed from the model according to the size of the posterior probability values. The study-level covariates were kept in the model if there was a .025 or .975 probability that its effect was greater (or less) than zero.

The statistical software program S-Plus 2000 with the Hierarchical Bayes Linear Model (*hblm*) function (DuMouchel, 1995) was used for cross-validation and construction of elegant trace plots. After fitting the hierarchical Bayesian linear model with a DuMouchel prior on τ, outliers were identified by examining the cross-validated residuals and the normal probability (Q-Q) plots of the residuals. In cross-validation the residual for a study's effect size represents a measure of how much a study's effect size differs from its Bayesian predictive distribution with that effect size data point deleted (DuMouchel & Normand, 2000). Predictive probabilities, which represent the value of the cumulative distribution evaluated for each study, were calculated in S-Plus with the *hblm* function (DuMouchel and Normand, 2000). A predictive probability for a study was considered to be more extreme than would be expected, if it was less than 2.5%, or greater than 97.5% of its cumulative distribution. The Bonferroni bound for the probability of the study with the largest absolute residual was also calculated in S-Plus with the *hblm* function in order to determine if that study differed significantly from the other studies. If an outlying study was detected with Bayesian cross-validation, then a sensitivity analysis was conducted both with and without the outlying data point in order to determine if inferences changed when the outlying data point was excluded.

Trace plots were constructed in S-Plus with the *hblm* function (DuMouchel, 1995), using a DuMouchel prior distribution on τ in order to depict the impact that different values of τ have on the meta-analytic results for the average effect size and the study-specific effect estimates. Bayesian meta-analytic trace plots, as illustrated by DuMouchel (1994), use histograms to display the posterior distribution of τ, with the height of the histogram depicting the posterior probability (left-axis) of one of nine specified values of τ (x-axis). The y-axis displays the

conditional mean (e.g., the posterior expectation of the intercept and study-specific estimates given τ), and the inlaid curve traces these values over the posterior distribution of τ.

Bayesian framework: Prior distributions for fixed effects: $\boldsymbol{\mu}$ and $\boldsymbol{\beta}$.

WinBUGS requires the specification of prior distributions in terms of their means and precisions (1/variance). WinBUGS will not allow strictly non-informative priors, as these can result in improper posterior distributions. For these analyses, nearly non-informative, normal priors that are uniform over the likely values of the parameters were chosen: $\mu \sim Normal\ (0, .001)$ and $\beta \sim Normal\ (0, .001)$, where 0 represents the mean of the distribution and .001 represents the precision (1/variance), which is equivalent to the variance equaling 1000. By using nearly uniform distributions, the posterior distributions will have virtually the same shape as the likelihood function (Spiegelhalter et al., 2004).

Bayesian framework: Prior distributions for $\boldsymbol{\tau}$ or $\boldsymbol{\tau^2}$.

A robust Bayesian approach ensures that the prior distributions represent a range of specifications and considers the sensitivity of the results to the different prior distributions (Greenhouse & Iyengar, 2009). In order to follow a robust approach, the following seven prior distributions were specified in WinBUGS (note that in WinBUGS distributions need to be specified in terms of their means and precisions) in order to describe the distributional form for the heterogeneity variance:

1. A uniform distribution on τ was specified as a baseline prior, as this prior distribution is the preferred prior distribution when there is reasonable information from the data (Spiegelhalter et al., 2004). Higgins et al. (2009) also support the use of the uniform on τ prior as the most appropriate choice for a minimally informative prior. Here, $\tau \sim Uniform\ (0, 5)$ was specified as the baseline prior for the Bayesian analyses.

2. An inverse gamma distribution was specified for τ^2 (or in WinBUGS, a gamma distribution on the precision) similar to the default prior discussed by Spiegelhalter et al. (2004), allowing preference to be given to a smaller variance, if the likelihood supports such low values of τ. In WinBUGS, this prior was specified as $\frac{1}{\tau^2} \sim Gamma\,(.01,.01)$.

3. A DuMouchel (DuMouchel, 1994) distribution was specified, in which the prior is $\pi(\tau) \sim \frac{s_0}{(s_0+\tau)^2}$, where s_0^2 represents the harmonic mean of the *K* sampling variances.

4. A DuMouchel distribution on τ was specified where s_0 was replaced by $\frac{s_0}{3}$ so that preference is given to stronger shrinkage and smaller values of τ: $\pi\,(\tau) = \frac{{s_0}/{3}}{\left({s_0}/{3}+\tau\right)^2}$.

5. A DuMouchel distribution on τ where s_0 was replaced by $3s_0$ so that preference is given to weaker shrinkage and larger values of τ: $\pi\,(\tau) = \frac{3s_0}{(3s_0+\tau)^2}$.

6. A half-normal distribution for τ^2, as recommended by Pauler and Wakefield (2000) was specified. This distribution has a mode at 0, is restricted to only positive values, has an upper 95% point at τ_μ, and steadily declines in values of τ: $\tau \sim HN\left[\left({\tau_\mu}/{1.96}\right)^2\right]$. (35)

Here, , a $\tau_\mu$value of 6.2 was chosen for the upper 95% point for τ (τ_μ). When τ_μ= 6.2, τ^2 is modeled from the absolute value of a normal distribution with a mean of 0 and a variance of 10.

7. A uniform distribution on τ^2 on the interval from 0 to 1000 was specified, where preference is given to larger values of τ: $\tau^2 \sim Uniform\ (0, 1000)$.

Bayesian framework: A *t*-distribution for the random effect.

In addition to the standard assumption that the random effect comes from a normal distribution, the Bayesian analyses for each of the aforementioned prior distributions for the heterogeneity variance were also conducted using the more flexible assumption that the random effect comes from a *t*-distribution with 4 degrees of freedom ($\delta_i \sim t_4$). Seltzer (1993) recommended the use of this approach in order to investigate the sensitivity of meta-analytic inferences to possible outliers, because modeling from a heavier-tailed t_4 distribution allows for a downweighting of the outliers.

Monitoring Bayesian model convergence.

Each of the Bayesian hierarchical linear models was monitored for convergence in WinBUGS by using the following recommended techniques: (a) running three Markov chains to obtain the final set of posterior estimates , (b) examining the dynamic trace plots of value of the unknown parameter against the iteration number to look for evidence of when the simulation stabilized, (c) evaluating the full-history plots for stabilization and overlap of the chains, (d) evaluating the smoothed kernel density plot of each variable, (e) monitoring the autocorrelation plot of each variable for every sample out to lag 50, and (f) evaluating the proximity to one of the Brooks-Gelman-Rubin diagnostic statistic (R hat) for each variable, by comparing the ratio of the variability within- and between-chains (WinBUGS User Manual, 2003).

Samples from the posterior distribution were obtained by running the model through 25,000 further sampling iterations so that the initial 25,000 iterations were discarded, as these initial draws were considered to be calculated from an unstable burn-in period. Each of the parameter values were evaluated in order to ensure that the Monte Carlo error value for each parameter did not exceed the recommended value of 5% or less of the sample standard

deviation (WinBUGS User Manual, 2003). The DIC measure was used to compare the different Bayesian models for model fit. The Bayesian model with the smallest or most negative DIC value was considered to be the model that would best predict a replicate data set of the same structure as that of the observed data (WinBUGS User Manual, 2003). The practical rule "that works reasonably well" for comparing among DIC values, proposed by Spiegelhalter et al. (2002), was applied for comparing the DIC values from the Bayesian models. Spiegelhalter's guidelines for comparing among DIC values is similar to the proposed guidelines of Burnham and Anderson (1998) for the AIC (Burnham & Anderson, 1998, as cited in Spiegelhalter et al., 2003) where the models that have AIC values within 1 – 2 units of each other deserve the best consideration and models deviating by more than 3–7 units garner less support for statistical model fit.

Development of a Bayesian Diagnostic Checklist.

A checklist was developed to guide each of the Bayesian analyses in order to ensure that a robust approach was followed and reduce the probability of error given the computational intensity of the Bayesian models. The checklist that was developed is composed of portions of the Bayesian quality assurance checklist proposed by Spiegelhalter et al. (2004) and also includes the aforementioned items that needed to be monitored in order to ensure Bayesian model convergence. Spiegelhalter et al. (2004) recommend the use of a checklist for reporting results due to both the complexity of applying Bayesian methods and the lack of established standards guidelines for publishing Bayesian meta-analytic results.

The posterior distributions of the monitored parameters were represented with summary statistics of the mean; standard deviation; median; R hat value; 95% credibility interval for $\mu, \beta, \theta_i, \theta_{new}, \tau$; and the probability that the mean values of μ, β, θ_{new}, $\beta_0 + \beta_1$, $\beta_0 + \beta_2$, etc. were greater (or less) than zero depending upon the direction of a beneficial effect. The illustrative reanalysis examples also report Bayesian study-specific estimates and demonstrate how, using all available evidence, individual estimates borrow strength from each other and shrink to a best

estimate of individual study effects. Graphs were developed in R in order to compare the widths of the 14 Bayesian 95% credibility intervals for $\theta_{new}, \beta_0,$ and τ to the mixed-effects REML 95% confidence intervals.

Data Overview

Data Set 1: Teacher expectancy effect data set.

Raudenbush and Bryk (2002) provide the data from a meta-analysis by Raudenbush (1984) of the effect of teacher expectancy on the intelligence quotient scores of school-aged children. The data for the Teacher Expectancy Effect (TEE) meta-analysis were available in the form of standardized mean differences, standard errors, and the coded value for the covariate variable, *weeks of prior contact* (Raudenbush & Bryk, 2002). The effect size estimate for each of the 19 studies represents the reported standardized mean difference in student intelligence quotient scores between the experimental group for whom teachers were given high expectations and the control group for whom teachers were given no expectations (Raudenbush & Bryk, 2002). The study-level covariate *weeks* was coded as X_i = 0, 1, 2, or 3, where the numeric value for X_i of 0, 1, and 2 indicates the number of weeks of prior contact between teachers and students before experimental manipulation. When X_i=3, *weeks* represents 3 or more weeks of contact prior to implementation of the experiment. This data set is displayed in Table 2 and will hereafter be referred to as the Teacher Expectancy Effect (TEE) data set. With a mixed-effects model and a maximum likelihood estimate of τ, Raudenbush and Bryk (2002) computed a significant predicted effect size in a study where X_i = 0, $\hat{\beta}_0$ = 0.407, *se* = 0.087, a significant covariate, weeks, $\hat{\beta}_1$ = -0.157, *se* = 0.036, and an estimate of .000 for the residual variation after controlling for weeks of prior contract.

Table 2

Summary Data for Teacher Expectancy Meta-analysis

Study Name	Effect size Estimate, Y_i	Standard Error(se_i) of Y_i	Weeks of Prior Contact, X_i
Rosenthal et al.	0.03	0.125	2
Conn et al.	0.12	0.147	3
Jose & Cody	-0.14	0.167	3
Pellegrini & Hicks a	1.18	0.373	0
Pellegrini & Hicks b	0.26	0.369	0
Evans & Rosenthal	-0.06	0.103	3
Fielder et al.	-0.02	0.103	3
Claiborn	-0.32	0.22	3
Kester	0.27	0.164	0
Maxwell	0.80	0.251	1
Carter	0.54	0.302	0
Flowers	0.18	0.223	0
Keshock	-0.02	0.289	1
Henrickson	0.23	0.29	2
Fine	-0.18	0.159	3
Greiger	-0.06	0.167	3
Rosenthal	0.3	0.139	1
Fleming & Anttonen	0.07	0.094	2
Ginsburg	-0.07	0.174	3

Note. Adapted from “Analyzing effect sizes: Random effect models” by S. Raudenbush, 2009. In Cooper, L. Hedges & J. Valentine (Eds.). *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 295-315). New York, NY: Russell Sage Foundation.

Data Set 2: Effects of academic detracking.

Rui (2009) performed a systematic review of the evidence regarding the academic and non-academic effects of the practice of detracking students. The term *tracking* is defined as grouping students into differentiated classrooms or academic programs based upon scholastic capabilities, also known as homogenous grouping (Rui, 2009). Detracking occurs when students are heterogeneously grouped into classes of mixed ability – it is a reversal of the policy of tracking students. Rui's meta-analytic data set is comprised of 10 studies representing 22 different effect sizes (i.e., within each study, the students were subgrouped by students' initial ability and grade level). This data set is displayed in Table 3 and will be referred to as the Detracking meta-analysis.

Rui performed an overall meta-analysis that examined the academic effects of detracking students using both an FE ($\hat{d}$ = 0.087, *k* = 22, N = 15,577, *p* < .001) and an RE ($\hat{d}$= 0.202, k = 22, N = 15,577, *p* < 0.01) model. The effect size estimate represents the standardized mean difference in achievement scores between detracked and tracked students, with a positive effect size favoring the detracked students. Rui also conducted separate subset FE and RE meta-analyses of the studies based upon type of student: (a) all-, medium-, and high-achieving students (FE $\hat{d}$ = 0.075, *k* = 14, *p* < .001; RE $\hat{d}$= 0.170, *k* = 14, *p* = .125) and (b) low-achieving students (FE $\hat{d}$ = 0.113, *k* = 8, *p* < .001; RE $\hat{d}$ = 0.283, *k* = 8, *p* < .005). Rui performed separate subset meta-analyses for studies that used experimental designs for (a) all-, medium-, and high-achieving students (FE $\bar{\hat{d}}$ = -0.300, *k* = 7, p < .001) and (b) low-achieving students (FE $\hat{d}$ = 0.627, *k* = 4, p < .001; RE $\hat{d}$ = 0.640, *k* = 4, *p* < .005). Rui conducted a sensitivity analysis by excluding the study by Kissoon-Singh that consisted of two extremely large effect sizes ($\hat{d}$ = 1.772 for high-ability students and $\hat{d}$ = 3.543 for average-ability students) and involved computer-based instruction. Rui's meta-analysis with these data points excluded revealed that there were no effects of detracking for high- and average-ability students (FE $\hat{d}$=

−0.005, *p*=.837, 95% CI −0.053, 0.043, *k* = 8). Rui also coded the following characteristics of the studies: type of subject (math, social science, English, or reading) and grade-level; however, he did not conduct subset meta-analyses of either of these potential moderator variables.

Table 3

Summary Data for the Effects of Academic Detracking Meta-analysis

Author	Y_i	se_i	Subj[a]	Grade[b]	Quasi[c]	Obs.[d]	Mid[e]	High[f]	All[g]
Cartwright	0.19	0.251	0	1	0	0	0	0	0
Cartwright	1.296	0.282	0	2	0	0	0	0	0
Cartwright	0.414	0.286	0	3	0	0	0	0	0
Marascuilo	-0.052	0.174	1	9	0	0	0	1	0
Marascuilo	0.687	0.227	1	9	0	0	1	0	0
Marascuilo	0.478	0.166	1	9	0	0	0	0	0
Slavin a	-0.575	0.13	0	4-6	0	0	0	0	1
Slavin b	-0.74	0.126	0	3-5	0	0	0	0	1
Thacker	0.141	0.278	1	6	1	0	0	1	0
Thacker	0.301	0.304	1	6	1	0	0	0	0
Thacker	0.393	0.268	1	6	1	0	1	0	0
Argys	-0.44	0.061	0	10	0	1	0	1	0
Argys	0.25	0.063	0	10	0	1	0	0	0
Kissoon	1.772	0.307	0	7	0	0	0	1	0
Kissoon	3.543	0.409	0	7	0	0	1	0	0
Hawkins	0.171	0.032	1	8	0	1	0	0	1
Hallinan	-0.702	0.13	1	9	0	1	0	1	0
Hallinan	-0.115	0.084	1	9	0	1	0	0	0
Mulkey	0.14	0.037	0	12	0	1	0	1	0
Mulkey	0.065	0.037	0	12	0	1	0	0	0
Burris	0.283	0.06	0	10	1	0	0	0	1
Burris	0.224	0.081	0	10	1	0	0	1	0

Note. [a] 0 = Math/Science; 1 = Reading/English/Social Science. [b]0 = Grade level.
[c]0 = Experimental or Observational; 1 = Quasi-experimental. [d]0 = Experimental or Quasi-experimental; 1 = Observational. [e]0 = High-, low-, or all-ability; 1=Mid-ability. [f]0 = Mid-,low-, or all-ability; 1 = High-ability. [g]0 = Mid, low, or high-ability; 1 = All-ability.
Note. Adapted from “Four decades of research on the effects of detracking reform: Where do we stand? a systematic review of the evidence”, by N. Rui, 2009, *Journal of Evidence-Based Medicine*, *2*(3), p. 168-170.

Data Set 3: Zinc for the common cold.

A Cochrane Collaboration systematic review by Singh and Das (2011) investigated the role of zinc on common cold symptoms. They included randomized, double-blind, placebo-controlled trials (RCT) that used zinc for at least five consecutive days to treat the common cold. Singh and Das (2011) used FE meta-analytic models when the heterogeneity was considered to be low (I^2 < 50%) and RE models when the heterogeneity was considered to be moderate to high (I^2> 50%). For the primary outcome measures, their simple RE meta-analysis with D-L estimation revealed that zinc resulted in significant reduction in the duration ($\hat{d}$ = -0.97; 95% CI -1.56, -0.38; p =.001, k = 6, I^2 = 93%, τ^2= .48) and severity of common cold symptoms ($\hat{d}$ = -0.39; 95% CI -0.77, -0.02; p= 0.04, k=5, I^2 = 75%, τ^2 = .13).

All of the studies included in the original Zinc for the Common Cold meta-analysis used multiple outcome measures (e.g., duration of symptoms, k=6; severity of symptoms k=5, time to resolution of cough, k=4; and time to resolution of nasal congestion, k=5) on the same subjects. In order to account for the statistical dependency among the multiple outcome measures, the most representative effect size, duration of cold symptoms, was chosen for the unit of analysis for the purpose of reanalysis for this research project.

Singh and Das (2011) did not conduct subgroup analyses of study-level characteristics because they did not have a sufficient number of included studies in order to perform such an analysis. However, they identified the following study-level characteristics that may contribute to possible sources of heterogeneity: (a) dosage/formulation of zinc, (b) age of participants, and (c) mean duration of cold symptoms prior to the administration of zinc. The data set displaying the effect sizes, standard errors, and study-level characteristics is displayed in Table 4.

Table 4

Summary Data from Zinc for the Common Cold Meta-analysis

Study Name	Outcome	Y_i	se_i	Syrup[a]	Adult[b]
Kurugol a	Duration of cold	-0.80	0.148	1	0
Kurugol b	Duration of cold	-0.50	0.189	1	0
Macknin	Duration of cold	0.00	0.128	0	0
Petrus	Duration of cold	-0.32	0.199	0	1
Prasad a	Duration of cold	-2.08	0.367	0	1
Prasad b	Duration of cold	-2.66	0.393	0	1

Note. Adapted from "Zinc for the Common Cold," by Singh, M. & Das, R., 2011, 2*, Cochrane Database of Systematic Reviews*.

[a]The outcome represented is the mean duration of symptoms for each study. [b]1 denotes the use of a syrup and 0 denotes the use of a zinc lozenge. [c]1 denotes a randomized control trial of adults (ages 18+) and 0 denotes a RCT of children between ages 2-16.

Data Set 4: Effects of Chondroitin for Hip or Knee Osteoarthritis

Reichenbach, Sterchi, Scherer, Trelle, Burgi, Burgi et al. (2007) conducted an RE meta-analysis to determine the efficacy of chrondroitin in treating the symptoms of pain in people with hip and/or knee osteoarthritis. This data set will be referred to as the Chondroitin meta-analysis. When Reichenbach et al. included all of the trials (*k*=20) in an RE meta-analysis using D-L estimation, they obtained results of $\hat{d}$ = -0.75, 95% CI (-0.99, -0.50) and $I^2 = 92\%$, indicating that patients in the chondroitin group experienced a significant reduction in pain as compared to the placebo group. However, because the I^2 value was very large (indicating significant between-trial variability), Reichenbach et al. made the decision to restrict the meta-analysis to include only the three trials that had large sample sizes (these three trials represented 40% of the patients) and an intention-to-treat analysis (ITT). Their subset RE meta-analysis of the *k* = 3 trials revealed an average effect size of -0.03, 95% CI (-0.13, 0.07), and $I^2 = 0\%$, concluding that large-scale and methodologically sound trials indicate that there is minimal to no benefit of chondroitin in reducing pain. Reichenbach et al. also conducted stratified univariate standard RE meta-analyses for the following study-characteristics and reported the associated *p* values for interaction: a) allocation concealment (*p*=.05), b) placebo controlled (*p*=.63), c) patient blinding (*p*=.22), d) adequacy of analyses with intention-to-treat principle (*p*=.017), e) trial size dichotomously coded as greater than or less than 200 (*p*=.022), f) length of follow-up dichotomously coded as greater than or less than 6 months (*p*=.152), g) funding (*p* =.186), h) route of administration (*p*=.062), and i) unequal use of analgesic co-intervention among experimental and control group(*p*=.043). This data set is displayed in Table 5 and it will be referred to as the Chondroitin data set and meta-analysis.

Table 5

Summary Data from Chrondroitin Meta-analysis

Study	Y_i	se_i	Conceal[a]	Placebo[b]	Blind[c]	ITT[d]
Kerzberg 1987	-1.01	0.474	1	0	1	1
Rovetta 1991	-2.14	0.337	1	0	0	1
Conrozier 1992	-1.93	0.27	1	0	0	1
Lhirondel 1992	-0.53	0.179	1	0	0	1
Mazieres 1992	-0.64	0.194	1	0	0	1
Morreale 1996	-1.81	0.179	1	0	0	1
Bourgeois 1998	-0.87	0.184	1	0	0	1
Busci 1998	-0.94	0.219	1	0	1	1
Conrozier 1998	-0.57	0.199	1	0	1	1
Uebelhart 1998	-1.17	0.296	1	0	0	1
Alekseeva 1999	-0.57	0.204	1	1	1	1
Malaise 1999	-0.42	0.189	1	0	0	1
Pavelka 1999	-1.23	0.204	1	0	0	1
Mazieres 2001	-0.23	0.179	1	0	0	1
Nasonova 2001	-0.86	0.107	1	1	1	1
Soroka 2002	-0.34	0.199	1	1	1	1
Michel 2005	-0.14	0.112	1	0	0	0
Clegg 2006	0.01	0.082	0	0	0	0
Kahan 2006	-0.02	0.082	0	0	1	0
Mazieres 2006	-0.3	0.112	1	0	1	1

Note. Adapted from "Meta-analysis: Chondroitin for the Knee or Hip," by Reichenbach et al. 2007, *Annals of Internal Medicine,* p. 582-585*.*

[a] 0 = Adequate concealment. 1 = Unclear or inadequate. [b]0 = Placebo controlled trial, 1 = Not placebo controlled. [c] 0= Adequate blinding, 1 =No blinding or unclear. [d]0 = Data analyzed in accordance with intention-to-treat principle, 1 =Unclear or no ITT analysis.

(continued) Summary Data from Chondroitin Meta-analysis

Study	≥200[e]	*N*	Route[f]	Analgesic[g]	≥ 6 months[h]	Weeks
Kerzberg 1987	1	17	1	1	1	6
Rovetta 1991	1	40	1	1	0	51
Conrozier 1992	1	56	0	1	1	24
Lhirondel 1992	1	129	0	1	1	25
Mazieres 1992	1	120	0	1	1	21
Morreale 1996	1	146	0	1	1	25
Bourgeois 1998	1	127	0	1	1	13
Busci 1998	1	85	0	1	0	26
Conrozier 1998	1	104	0	1	0	52
Uebelhart 1998	1	46	0	1	0	52
Alekseeva 1999	1	100	0	1	0	39
Malaise 1999	1	120	0	1	0	52
Pavelka 1999	1	105	0	1	1	13
Mazieres 2001	1	132	0	0	0	27
Nasonova 2001	0	555	0	0	0	26
Soroka 2002	1	100	0	1	0	52
Michel 2005	0	300	0	0	0	103
Clegg 2006	0	631	0	0	1	24
Kahan 2006	0	622	0	1	0	132
Mazieres 2006	0	311	0	0	0	34

Note. Adapted from "Meta-analysis: Chondroitin for the Knee or Hip," by Reichenbach et al. 2007, *Annals of Internal Medicine,* p. 582-585.

[e]0 = Trial size greater than or equal to 200, 1 = Trial size less than 200. [f]0 = Oral supplement 1 = Intramuscular injection. [g]0 = Similar analgesic drug use among treatment and control groups, 1= Control group reported greater analgesic use than treatment group or unclear. [h]0 = Time of pain assessment greater than or equal to 6 months.

Data Set 5: Effects of Occupational Stress Management Interventions

Richardson and Rothstein (2008) conducted a simple RE meta-analysis that examined the effectiveness of workplace stress management interventions (SMI) from 36 experimental studies. There were *k* = 55 different treatment-control contrasted interventions that were represented within these 36 studies (e.g., 16 of the 36 included studies compared multiple SMI to a common control group). There was some statistical dependency among these effect sizes estimates as: (a) thirty-five of the thirty-six studies use multiple outcome measures on the same subjects (e.g., multiple-endpoint studies) and (b) some of the studies included multiple-treatment designs in which multiple variants of a treatment were compared to the same control group. In the original meta-analysis, in order to account for the statistical dependency among the multiple outcome measures, the average effect size and standard error of the multiple outcomes was computed in CMA2, and the composite summary effect was used as the unit for analysis. This is one of the approaches recommended by Lipsey and Wilson (2001) for meta-analyzing data from studies that include multiple dependent effect sizes, although Lipsey and Wilson recognize that such an approach has the disadvantage of omitting potentially meaningful information. This approach for considering the correlation (e.g., dependency) that exists among the outcome measures essentially sets the correlation between the measures to 1.00 (i.e., a perfect correlation) and thereby risks overestimating the standard error and underestimating the precision of the composite effect (Borenstein et al., 2009). Alternative approaches would be either to use one representative effect size per study or to estimate the covariance between the dependent outcome measures, given that such a covariance matrix is available (Lipsey & Wilson, 2001). The original SMI meta-analysis did not make any adjustments for the multiple-treatment design studies, and so for purposes of this comparative research project, these adjustments were not made in this reanalysis.

Their simple RE meta-analysis with D-L estimation revealed a significant average effect size in favor of SMI ($\hat{d}$= 0.526, 95% CI 0.364, 0.687, *k* = 55, I^2 = 73%) after excluding one

outlying study. Richardson and Rothstein (2008) conducted the following moderator analyses of variables at the intervention-level in order to determine if variation in the effect sizes could be explained by the following intervention-level moderators: type of SMI, number of components included in the SMI, and duration of treatment. The subgroup moderator analysis of type of SMI revealed the following average effect sizes for each of the SMI categories using an RE model with D-L estimation: (a) cognitive– behavioral ($\hat{d}$ = 1.164, 95% CI 0.456, 1.871, *k* = 7), (b) relaxation ($\hat{d}$ = 0.497, 95% CI 0.309, 0.685, *k* = 17), (c) organizational ($\hat{d}$ = 0.144, 95% CI -0.123, 0.411, *k* = 5) , (d) multimodal ($\hat{d}$ = 0.239, 95% CI 0.092, 0.386, *k* = 19), and (e) alternative ($\hat{d}$= 0.909, 95% CI 0.318, 1.499, *k* = 7). The moderator analysis of the number of SMI components suggested that single-component interventions produced larger effects than multiple component interventions using a RE model with D-L estimation: (a) one component ($\hat{d}$ = 0.643, 95% CI 0.309, 0.977, *k*=20), (b) two components ($\hat{d}$ = .607, 95% CI 0.346, 0.868, *k*=18), (c) three components ($\hat{d}$ = -0.104, 95% CI -0.627, 0.418, *k*=2), and (d) four or more components ($\hat{d}$ = .271, 95%CI 0.102, 0.440, *k* =15). Richardson and Rothstein (2008) performed a subgroup analysis of treatment duration by categorizing the length of duration into four categories and by classifying studies by duration of treatment and SMI type because treatment length might have been confounded by the type of SMI (see Richardson and Rothstein Table 4 for details).

Richardson and Rothstein (2008) also conducted subgroup moderator analyses at the sample-level by investigating the effect of the type of industry sector using an RE model with D-L estimation: (a) office workers ($\hat{d}$ = 0.680, *k* = 19), (b) health care workers ($\hat{d}$= 0.492, *k*=8), and (c) education workers ($\hat{d}$=1.2555, *k*=7). The industry workers were also further classified by SMI type (see Richardson and Rothstein Table 7 for details). At the outcome-level, they also conducted a subgroup analysis by disaggregating the data by type of outcome variable (e.g., psychological, physiological, and organizational) and SMI type. They performed outlier

analyses by examining the widths of the confidence intervals displayed in the forest plots at the combined level and the subgroup level, and they identified one study (Taylor, 1991) as an outlier.

The effect size data and confidence intervals from the included studies were not reported in Richardson and Rothstein's (2008) article, but they graciously agreed to provide full access to their CMA2 data file for the purposes of this research project. Richardson and Rothstein's data from the 36 included studies representing 55 treatment-control interventions is displayed in Table 6 and will be referred to as the SMI data set.

Table 6

Summary Data from SMI data set

Study	Y_i	se_i	Type of SMI	Weeks	Industry	Parts
Aderman	0.899	0.354	Relaxation	12	Mixed	2
Aderman	0.25	0.347	Other	12	Mixed	1
Alford	0.429	0.26	Other	0.5	Child Services	1
Bertoch	0.524	0.372	Multimodal	12	Education	4
Bond	0.323	0.307	Cognitive	14	Office	2
Bond	0.162	0.316	Organizational	14	Office	2
Bruning	0.523	0.369	Multimodal	8	Office	4
Bruning	0.38	0.366	Relaxation	8	Office	1
Bruning	0.818	0.378	Other	8	Office	1
Carson	-0.044	0.276	Organizational	5	Health Care	1
Cecil	0.13	0.342	Cognitive	6	Education	1
Cecil	0.056	0.335	Organizational	6	Education	1
Chen	1.731	0.168	Other	1	Office	1
Collins	1.646	0.565	Multimodal	5	Office	4
Collins	0.827	0.53	Multimodal	5	Office	2
de Jong	0.094	0.218	Multimodal	8	Mixed	4
de Jong	0.345	0.221	Multimodal	8	Mixed	4
Fava	0.718	0.342	Multimodal	28	Manual/Technical	4
Fiedler	0.059	0.263	Relaxation	9	Manual/Technical	2
Ganster	0.149	0.242	Multimodal	8	Office	2
Gildea	0.354	0.503	Multimodal	missing	Child Services	4
Gildea	-0.101	0.503	Relaxation	missing	Child Services	2
Higgins	0.711	0.352	Relaxation	6	Office	2
Higgins	0.314	0.336	Multimodal	6	Office	4
Hoke	0.048	0.202	Multimodal	12	Mixed	5
Jackson	0.433	0.252	Organizational	24	Health Care	1
Kolbell	0.082	0.402	Relaxation	4	Child Services	1
Kolbell	-0.097	0.403	Organizational	4	Child Services	1
Lee	0.988	0.283	Cognitive	2	Health Care	2
Maddi	2.374	0.462	Cognitive	10	Office	1
Maddi	-0.062	0.497	Relaxation	10	Office	2
Murphy	0.566	0.447	Other - EMG	1.5	Manual/Technical	1
Murphy	0.306	0.47	Relaxation	1.5	Manual/Technical	1
Peters	0.643	0.222	Relaxation	8	Office	2
Peters	0.329	0.238	Relaxation	8	Office	1
Peters, 99	-0.028	0.32	Multimodal	10	Manual/Technical	4
Pruitt	0.335	0.252	Multimodal	missing	Manual/Technical	4
Shapiro	0.128	0.395	Relaxation	8	Health Care	2
Sharp	1.993	0.327	Cognitive	4	Education	1
Sharp	2.037	0.323	Other	4	Education	1
Shimazu	0.174	0.14	Cognitive	11	Office	2
Stanton	1.177	0.396	Relaxation	missing	Office	1
Taylor[a]	3.988	.430	Alternative	6	Health Care	1
Taylor[a]	2.015	.318	Alternative	6	Health Care	1
Thomason	0.159	0.389	Multimodal	6	Manual/Technical	4
Thomason	-0.261	0.396	Multimodal	6	Manual/Technical	3
Thomason	0.337	0.389	Other	6	Manual/Technical	1

Tsai	0.468	0.173	Relaxation	5	Health Care	2
Tunnecliffe	3.196	0.807	Cognitive	5	Education	2
Tunnecliffe	1.523	0.607	Relaxation	5	Education	2
Vaughn	1.574	0.542	Relaxation	4	Office	1
von Baeyer	1.307	0.591	Multimodal	1	Health Care	4
Wirth	-0.05	0.309	Multimodal	4	Mixed	3
Wirth	0.026	0.361	Multimodal	4	Mixed	3
Yung	0.248	0.31	Relaxation	4	Health Care	1
Yung	0.873	0.313	Relaxation	4	Health Care	2
Zolnierczyk	0.043	0.217	Multimodal	10	Office	4

[a]The study by Taylor was excluded from the original meta-analysis.

Chapter IV: Results

Data Set 1: Teacher Expectancy Meta-analysis

Classical Meta-Analysis Results for TEE

The results for the classical simple models obtained by Raudenbush (2009) were successfully replicated in *metafor* with the use of FE, D-L, and REML estimation, as displayed in Table 7. The proportion of variability that is due to heterogeneity, I^2, was estimated from the simple RE model as 41.9% with REML estimation and 49.8% using the standard D-L method of estimation. The *Q*-test for heterogeneity in the simple RE models was significant at *p*=.007, indicating the presence of significant variation between effect sizes. As Raudenbush (2009) indicated in his analysis and as shown in Table 7, it is evident that the FE estimate for $\hat{\beta}_0$ may be inaccurate and its standard error too small, because the estimate does not reflect the extra heterogeneity that exists among the effect sizes.

The mixed-effect models with REML in *metafor* produced results that were identical to those computed by Raudenbush (2009). Comparison of the model fit statistics revealed that the mixed-effect model was the preferred classical model for this data, as it had the lowest AIC and BIC values. The estimates of the intercept and covariate were highly significant ($\hat{\beta}_0 = -0.407, p = .000$; $\hat{\beta}_1$= -0.157, *p*<.001), and the estimate of the residual variance in effect sizes was .000, except for the model used in the sensitivity analysis where τ^2 was set equal to the upper limit of its 95% CI from the *Q*-profile method. Here, when τ^2 was specified to its upper 95% CI boundary, the estimate of the regression coefficients and particularly their standard errors increased substantially although the impact is 'minimal' as the basic fitted model and meta-analytic conclusions did not change. The Higgins et al. (2009) approximate prediction interval for an effect in a new study ranged from 0.223 to 0.591 when $X_i = 0$, and this prediction interval increased markedly to -0.552, 1.514 when τ^2 was set equal to its upper 95% boundary.

Assessment of Publication Bias.

There was no evidence of outlying data points that may have influenced the data, as determined by inspection of the forest plots (Figure 1), *Q*-*Q* Plot (Figure 3), and leave-one-out case diagnostics (Table 8). In terms of assessing the presence of publication bias for the mixed-effects model, the funnel plot shows slight asymmetry with some studies possibly missing in the lower left quadrant; however, the Egger regression test for funnel plot symmetry using a mixed-effect meta-regression model was not significant (p=.491). For the simple RE model, the trim and fill method estimated that three studies may be missing from the left side of the mean effect size; and, if these missing studies were added to the data set, the resulting mean effect size would decrease from .084 to .028. The Orwin Fail-safe *N* value from a simple unweighted model was 19, indicating that, if 19 studies averaging null results were added to this data set, the unweighted average effect size of a simple RE model would be reduced from 0.164 by half, to a target effect size value of 0.082. However, it is important to apply caution when interpreting the results from both the Orwin Fail-safe *N* and trim and fill methods, because these methods were only applied to the simple RE model and therefore may not be valid as there was strong supporting evidence in favor of fitting a mixed-effect model for this data set.

Table 7

TEE: Classical Meta-analysis Results

Estimation Method	FE	D-L	REML	REML	REML K&H[a]	$REML_L$[b] τ^2=.006	$REML_H$[c] τ^2=.198
Model	Simple	Simple	Simple	Mixed	Mixed	Mixed	Mixed
$\hat{B}_0$	.060	.089	.084	.407	.407	0.418	0.481
s.e.	.037	.056	.052	.097	.086	0.095	0.204
95% CI low	-.112	-.020	-.018	.237	.226	0.231	0.081
95% CI high	.132	.199	.185	.578	.589	0.604	0.881
p value	.098	.109	.105	<.001	<.001	<.001	.019
$\hat{B}_1$ Weeks				-.157	-0.157	-0.162	-0.187
s.e.				.036	.035	0.040	0.092
95% CI low				-.228	-0.232	-.239	-.368
95% CI high				-.087	-0.083	-.084	-.007
p value				<.001	<.001	<.001	.042
τ		.161	.137	.001	.001	0.078	0.445
95% CI low		.077	.077	0	0	0	0
95% CI high		.445	.445	.268	.268	0.268	0.2680
τ^2		.026	.019	.000	.000	0.006	0.198
s.e.			.016	.007	.007	0.010	0.082
95% CI low		.006,	.006	.000	.000	0	0
95%CI high		.198	.198	.072	.072	0.072	0.072
Q	.	35.8	35.8	16.6	16.6	16.6,	16.6
p value		.007	.007	.484	.484	.484	.484
PI $\hat{B}_0$ Lower[d]			-.23,	.223,	.226	0.159,	-0.552,
PI $\hat{B}_0$ Upper[e]			.39	.591	.589	0.676	1.514
PI $\hat{B}_1$Lower				-.232	-.231	-0.345	-1.146
PI $\hat{B}_1$Upper				-.083	-.083	0.022	0.771
I^2	.	49.8%	41.9%				
Q_{Mods}				19.3	F=19.8	16.6	4.133
p value				<.001	<.001	<.001	.042
Log-like	3.460	-3.28	-5.21	1.36	1.36	0.574	-8.816
Deviance	6.921	6.56	10.42	-2.73	-2.73	-1.147	17.632
AIC	8.921	10.56	14.42	3.27	3.27	4.853	23.632
BIC	9.870	12.45	16.20	5.77	5.77	7.354	26.131
Min $\theta_i^*(\tau)$				-.065	-.065	-0.095	-0.273
Max $\theta_i^*(\tau)$				.407	.408	0.449	0.892

Note. [a]K&H denotes Knapp and Hartung (2003) adjustment. [b]$REML_L = \tau^2$ was set equal to the lower bound of its 95% CI from the simple REML model. [c]$REML_H = \tau^2$ was set equal to the upper bound of its 95% CI from the simple REML model. Simple = no covariates included. Mixed = covariates from the REML fitted model included. [d,e] denotes the upper and lower boundaries of Higgins et al. (2009) approximate prediction interval for a new study. Permutation test *p* value for the mixed-effects model $\hat{B}_1$ was <.001.

Figure 1

TEE Meta-analysis: Forest Plot Mixed-effects Model (REML)

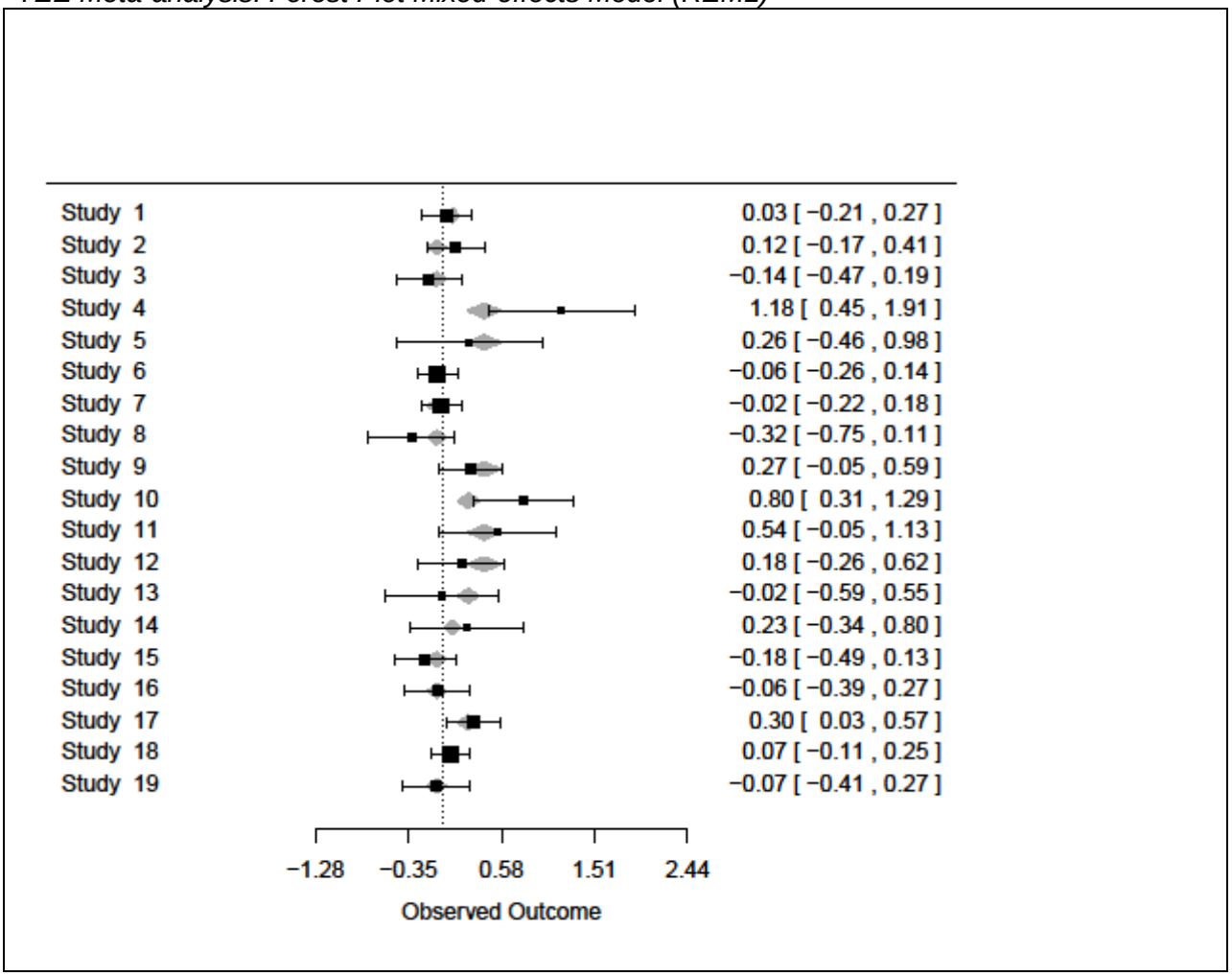


Figure 2

TEE Meta-analysis: Funnel Plot

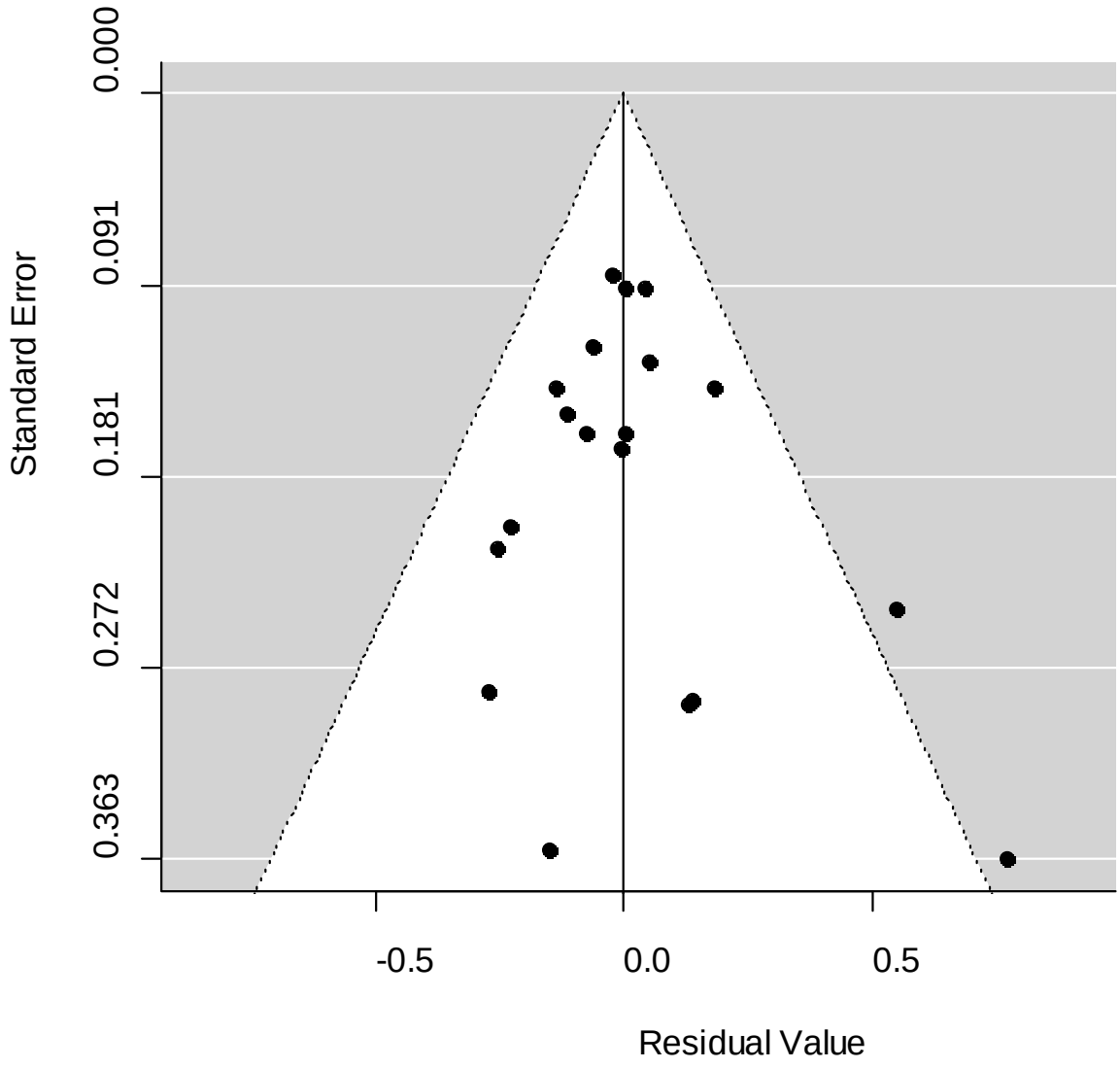


Table 8

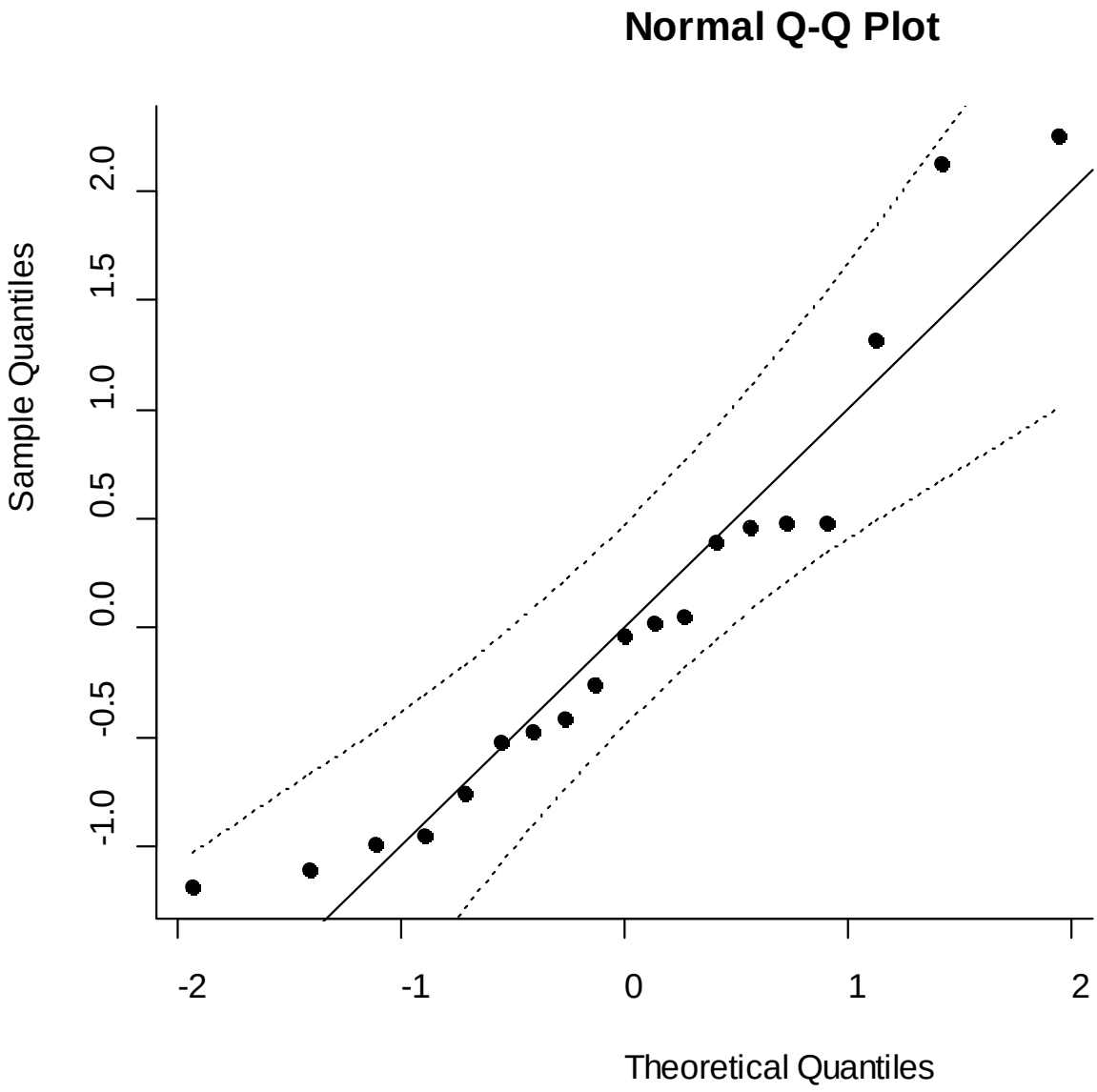
TEE: Leave-One-Out Case Diagnostics

Study	*r*student[a]	cov.r[b]	τ^2.del[c]	*Q*.del[d]	B_0[e]	β_1[e]
1	-0.525	1.097	0.000	16.292	-0.097	0.032
2	1.323	1.110	0.000	14.818	-0.100	0.270
3	-0.470	1.083	0.000	16.347	0.031	-0.083
4	2.131	1.057	0.000	12.028	0.511	-0.464
5	-0.411	1.059	0.000	16.399	-0.100	0.091
6	0.050	1.253	0.000	16.566	-0.006	0.015
7	0.484	1.253	0.000	16.334	-0.056	0.150
8	-1.188	1.046	0.000	15.157	0.058	-0.157
9	-0.987	1.393	0.000	15.593	-0.618	0.562
10	2.249	1.053	0.000	11.509	0.501	-0.397
11	0.459	1.090	0.000	16.357	0.138	-0.126
12	-1.107	1.180	0.000	15.343	-0.469	0.426
13	-0.953	1.040	0.000	15.661	-0.183	0.145
14	0.477	1.017	0.000	16.340	0.036	-0.012
15	-0.759	1.092	0.000	15.992	0.053	-0.142
16	0.028	1.084	0.000	16.567	-0.002	0.005
17	0.394	1.199	0.000	16.413	0.169	-0.134
18	-0.263	1.186	0.000	16.499	-0.067	0.023
19	-0.032	1.077	0.000	16.567	0.002	-0.005

Note. [a]denotes the *r*student (externally standardized residual) value with the i^{th} study excluded. [b]denotes the covariance ratio. [c]denotes the estimate value of τ^2 based on the data set with the i^{th} study excluded. [d]denotes the *Q* statistic for the test of heterogeneity based on the data set with the i^{th} study removed. [e]denotes by how many standard deviation units the regression coefficients change when the i^{th} study is excluded (Viechtbauer, 2009).

Figure 3

TEE Meta-analysis: Normal Q-Q Plot



Bayesian Meta-Analysis Results for TEE

A sample of the Bayesian quality assurance checklist that was used with the uniform prior distribution on τ (0,5) and $\delta_i \sim t_4$ to fit the models, monitor the parameters, and check the items related to model convergence is displayed in Figure 4. As shown in Table 9 and in the Appendix, the Bayesian model estimates were very similar to those from the REML classical mixed-effects model. The Bayesian results for the 14 different meta-analytic models (e.g., the seven different prior distributions on the heterogeneity variance for which $\delta_i \sim$ normal distribution or $\delta_i \sim t_4$ are displayed in the Appendix). All of the 14 different Bayesian models produced similar results with negative DIC values, and the Bayesian credibility intervals were smaller for almost all of the parameters when the $\delta_i \sim t_4$. Bayesian cross-validation in S-plus with the *hblm* function did not indicate any outlying data points (see Figure 6 and Table 10). Figure 5 displays a trace plot that depicts how the meta-analytic results are dependent upon the posterior probability values of τ for the fitted model. For example, as the value of τ approaches .045 or less, as depicted in the histogram, the traced curves for the estimates show that the intercept (labeled A) and study-specific estimates (labeled B-T) shrink toward their fixed-effect estimates for $X_i\beta$, and as the values of τ approach 0.711 or greater, the study-specific estimates are close to their observed values so that there is little shrinkage or “borrowing of strength” from the other studies.

Figure 4

TEE Meta-analysis: Bayesian Quality Assurance Meta-analysis Checklist

☒ Cross-validation Outlier identified? ☒ No ☐ Yes

Fitted model: $X_i\beta = \hat{\beta}_0 + \hat{\beta}_{i1}$(Weeks)

Prior Distribution for τ:

☒ Uni. on τ ☐ Uni. on τ^2 ☐ Half-normal ☐ Inverse Gamma ☐ DuM. s0 ☐ DuM. s0/3 ☐ DuM. 3s0

Form for the random-effects distribution: Prior Distribution for **β**:

☐ Normal distribution ☒ t_4-distribution ☒ Normal (0,1000)

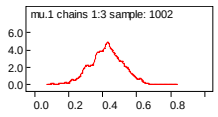
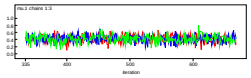
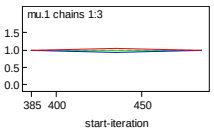
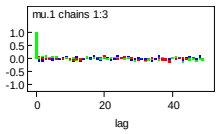
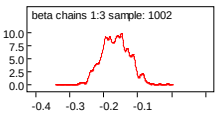
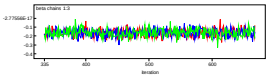
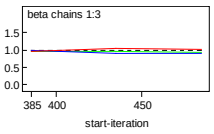
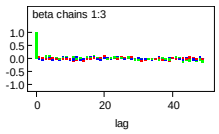
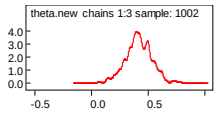
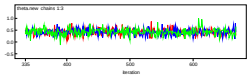
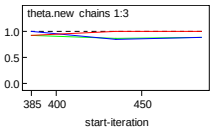
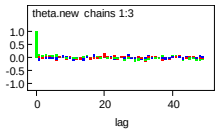
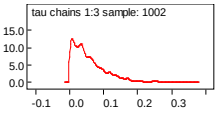
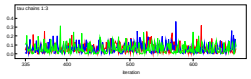
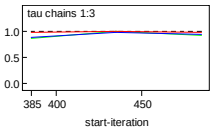
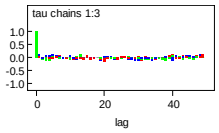
Model specification:

☒ Run 3 chains ☒ 50,000 iterations ☒ Discard half ☒ DIC = -6.228

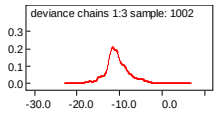
Do meta-analytic inferences change from original fitted hblm: ☒ No ☐ Yes If yes, impact: ___

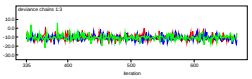
Node	Mean	95% Cr Int	Set param.	MC error< 5%	Rhat prox. to 1
β_0	0.416	0.211, 0.609	☒	☒	☒
β_1	-0.161	-0.240, -0.083	☒	☒	☒
τ	0.043	0.002, 0.015	☒	☒	☒
$\theta_{new\beta0}$	0.415	0.188, 0.639	☒	☒	☒
$\theta_{new\beta1}$	-0.162	0.073, -0.123	☒	☒	☒
Deviance	-10.682	-16.249, -3.939	☒	☒	☒
$p\ \beta_0 < 0$	.000	.000, .000	☒	☒	☒
$p\ \beta_1 > 0$	.001	.000, .000	☒	☒	☒
$p\ \theta_{new\beta0} < 0$	.001	.000, .000	☒	☒	☒
$P\ \theta_{new\beta1} > 0$	.017	.000, .000	☒	☒	☒
θ_i^*	θ_8^* min = -.088, θ_4^* max = .449		☒	☒	☒

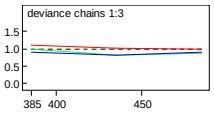
Items pertaining to relevant parameter diagnostic checks:

	Kernel Density	History	BGR	Autocorrelation
☒ β_0				
☒ β_1				
☒ θ_{new}				
☒ τ				

☒ Dev.


deviance chains 1:3 sample: 1002
0.3
0.2
0.1
0.0
-30.0
-20.0
-10.0
0.0




deviance chains 1:3
1.5
1.0
0.5
0.0
385
400
450
start-iteration

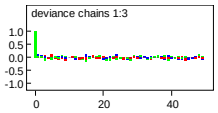

deviance chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

Table 9

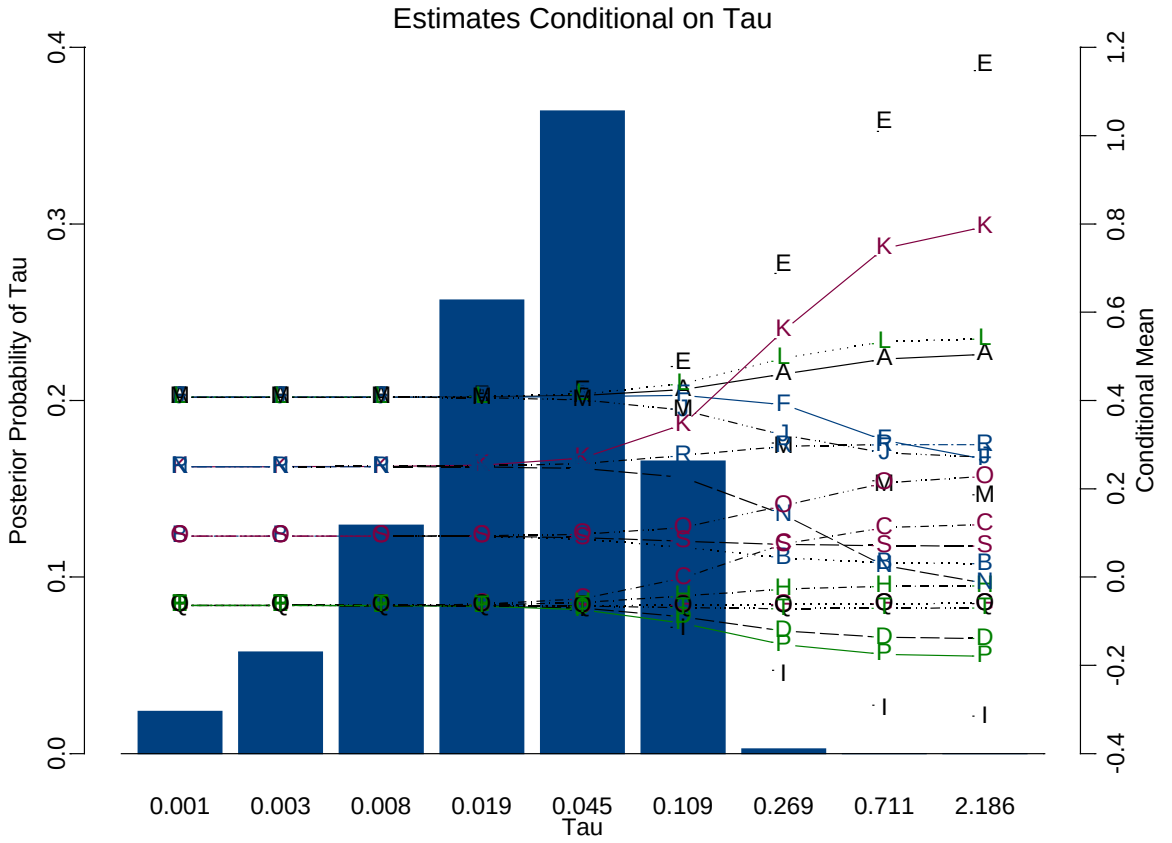
TEE Meta-analysis: Bayesian Results

Parameter	Posterior mean	*s.d.*	Prob > 0	95% Cr Int[a]
$\hat{\beta}_0$	0.416	0.098	.999	0.211, 0.609
$\hat{\beta}_1$	-0.161	0.040	.001	-0.240, -0.083
$\theta_{new.\beta 0}$	0.415	0.112	.999	0.188, 0.639
$\theta_{new.\beta 1}$	-0.158	0.083	.001	-0.327, 0.011
τ	0.043	.036		0.015, 0.142

Note. Prior for τ ~ uniform (0,5). Prior for **β** ~ (0,1000). $\delta_i \sim t_4$. DIC = -6.23.
[a]95% Credibility Interval.

Figure 5

TEE: Trace Plot of Tau for Model Intercept and Study-specific Estimates



A=(I B=Rs C=Cn D=J&C E=P&Ha F=P&Hb G=E&R H=Fea I=Cl J=K&L K=Mx L=Cr M=Fl N=Ks O=Hn P=Fn Q=Gr R=R&J S=F&A T=Gn

Figure 6

TEE: Bayesian Q-Q Plot of Fitted Model

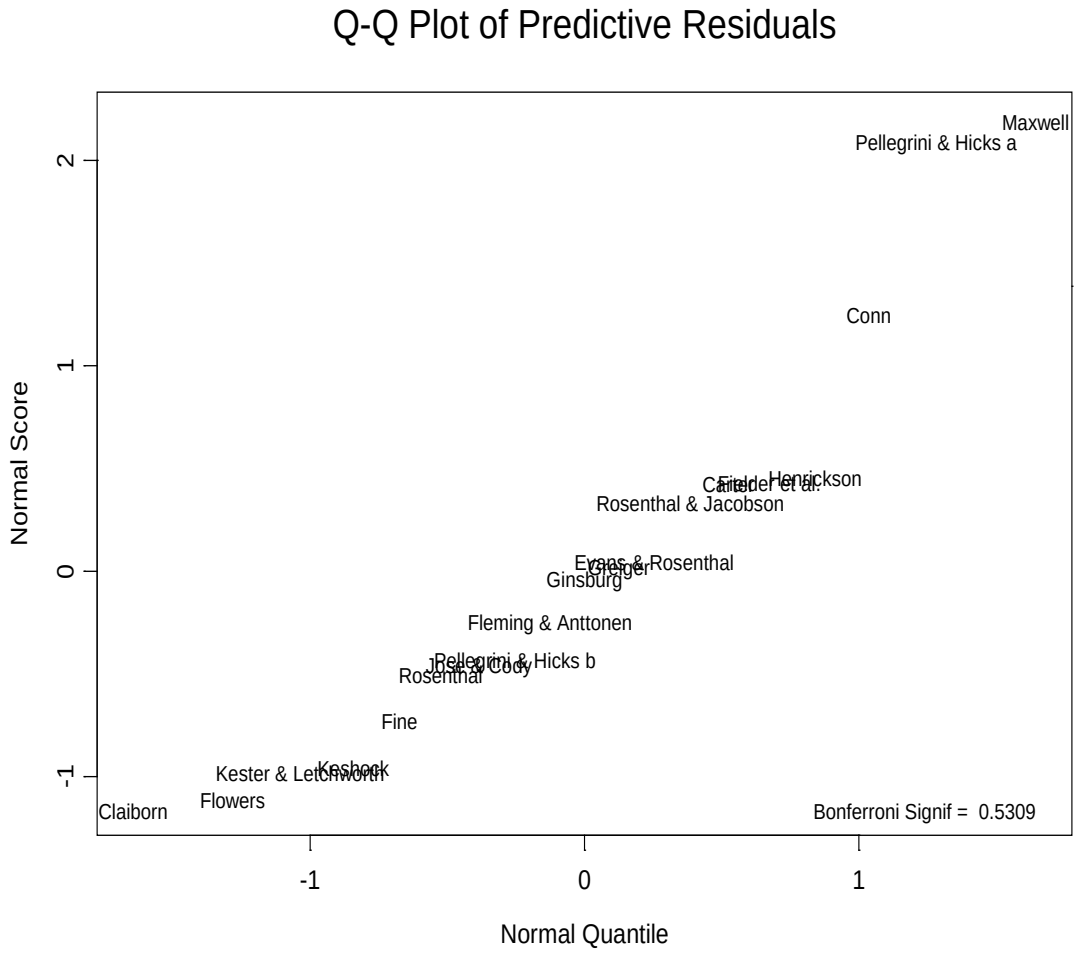


Table 10

TEE Meta-analysis: Bayesian Cross-validation Results for Fitted Model

Study	Y_i	se_i	Prior Mean	Prior *s.d.*	Pred. Prob[a]	τ^2	Post. Mean[b]	Post. *s.d.*	Prob > 0
Rosenthal	0.03	0.125	0.100	0.074	0.312	0.004	0.086	0.056	0.937
Conn	0.12	0.147	-0.087	0.076	0.896	0.003	-0.048	0.068	0.206
Jose	-0.14	0.167	-0.059	0.079	0.330	0.003	-0.072	0.066	0.121
Pellegrini a	1.18	0.373	0.364	0.107	0.982	0.003	0.427	0.108	1.000
Pellegrini b	0.26	0.369	0.422	0.110	0.337	0.003	0.409	0.103	1.000
Evans	-0.06	0.103	-0.068	0.085	0.523	0.004	-0.065	0.058	0.126
Fielder	-0.02	0.103	-0.077	0.084	0.671	0.004	-0.058	0.059	0.150
Claiborn	-0.32	0.22	-0.053	0.075	0.125	0.003	-0.079	0.071	0.108
Kester	0.27	0.164	0.467	0.122	0.166	0.003	0.400	0.096	1.000
Maxwell	0.8	0.251	0.222	0.078	0.986	0.002	0.275	0.084	1.000
Carter	0.54	0.302	0.400	0.111	0.669	0.003	0.415	0.103	1.000
Flowers	0.18	0.223	0.456	0.115	0.135	0.003	0.400	0.100	1.000
Keshock	-0.02	0.289	0.264	0.083	0.172	0.003	0.244	0.078	0.995
Henrickson	0.23	0.29	0.091	0.069	0.680	0.003	0.098	0.065	0.951
Fine	-0.18	0.159	-0.055	0.078	0.238	0.003	-0.075	0.066	0.109
Greiger	-0.06	0.167	-0.067	0.080	0.514	0.003	-0.065	0.065	0.140
Rosenthal	0.3	0.139	0.244	0.088	0.636	0.003	0.257	0.070	1.000
Fleming	0.07	0.094	0.097	0.078	0.407	0.004	0.089	0.051	0.957
Ginsburg	-0.07	0.174	-0.066	0.079	0.491	0.003	-0.066	0.066	0.139

Note. [a]Predictive probability. [b]Posterior mean.

Comparison of Classical and Bayesian Results for TEE

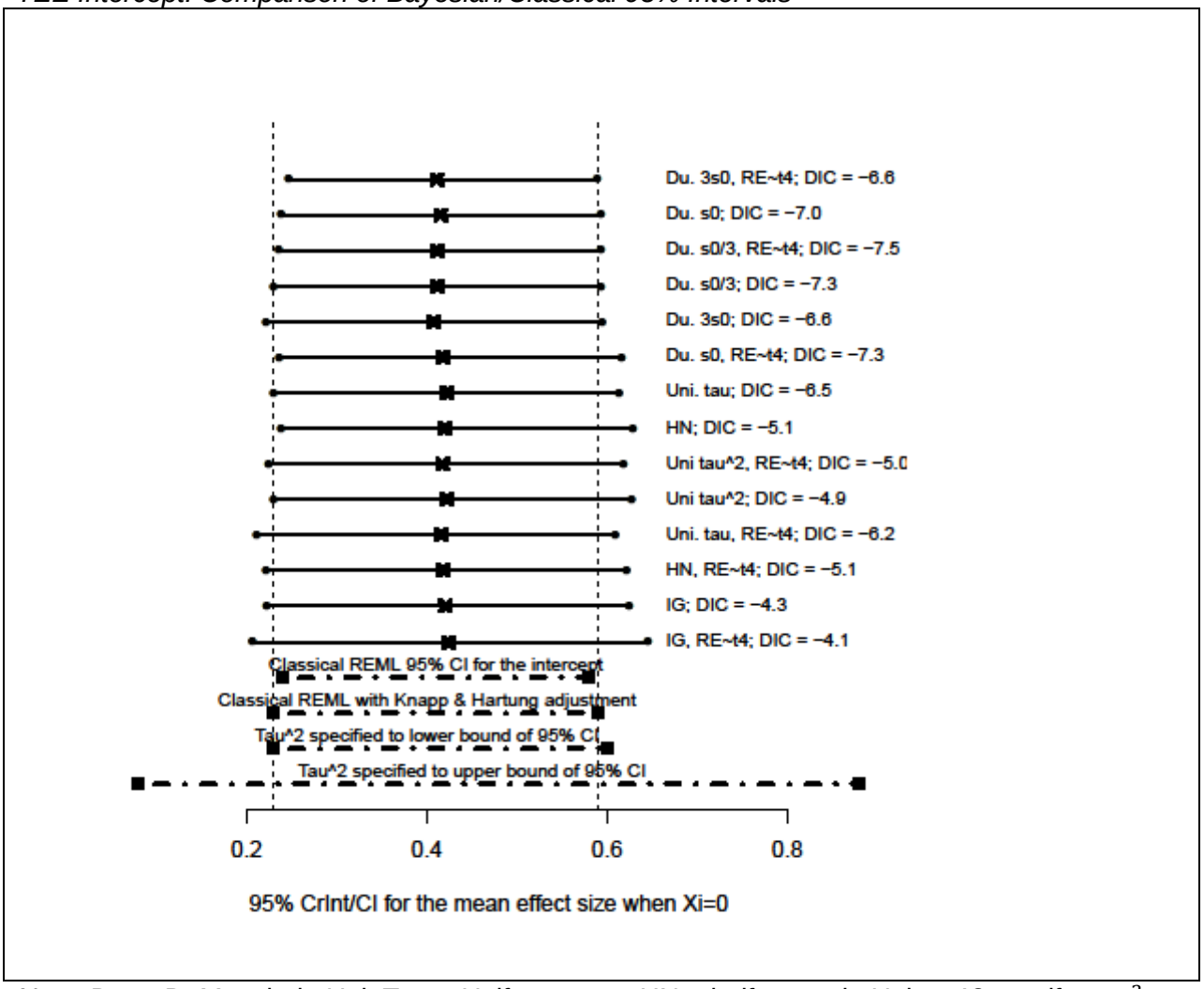
Overall, the meta-analytic inferences do not change among the classical mixed-effects with REML estimation for τ^2 and Bayesian models for the TEE dataset. Figures 7-9 display the widths of the Bayesian 95% credibility intervals compared to the classical 95% CI widths with REML estimation for: (a) mean effect size, (b) the predicted effect in a new study, and (c) τ, the standard deviation of the residual variance. In terms of the ability to estimate the model coefficients, the Bayesian parameter estimates for $\hat{\beta}_0$ and $\hat{\beta}_1$ are very similar to those from the classical analyses; however, the Bayesian 95% credibility intervals (Cr Int) for these regression coefficients are wider than those from the classical analysis, because the Bayesian approach allows for full consideration of the uncertainty in the estimates of these parameters. The 95% CI estimate for the model intercept (and its prediction interval) from the classical (REML) mixed-effects model with the Knapp and Hartung (2003) adjustment made to the standard errors is the closest in similarity to the Bayesian models. The DuMouchel $s_0/3$ prior distribution, which shows preference for low values of τ and stronger shrinkage, produces the 95% CrInt that most closely resemble those estimated from REML for the regression coefficients.

Consideration of the ability of the classical models to predict an effect in a new study requires the recognition that the classical approximate prediction interval width for an effect in a new study is smaller than all of the Bayesian credibility intervals. The classical interval for the predicted effect in a new study ranges from 4% (DuMouchel $s_0/3$) to 67% (inverse gamma) smaller than the Bayesian credibility intervals, indicating that for this data structure the classical approximate prediction interval may underestimate the true interval for an effect in a new study. In terms of the ability to quantify the residual variance, the mean Bayesian parameter estimates for τ^2 are almost identical to the mean estimate from the classical mixed model with REML estimation. However, the width of the classical 95% CI for τ^2 as computed by the classical REML Q-profile method is larger than all of the Bayesian 95% Credibility Intervals for the residual variance as displayed in Figure 9.

With regard to the ability of the models to estimate study-specific effects, the various values of the Bayesian θ_i^* as depicted in Figure 8 provide better estimates, because the Bayesian θ_i^* consider the full uncertainty in all the auxiliary parameters as compared to the classical study-specific estimates. Figure 10 depicts the study-specific shrinkage estimates for the REML model compared to those from the fully Bayesian model.

Figure 7

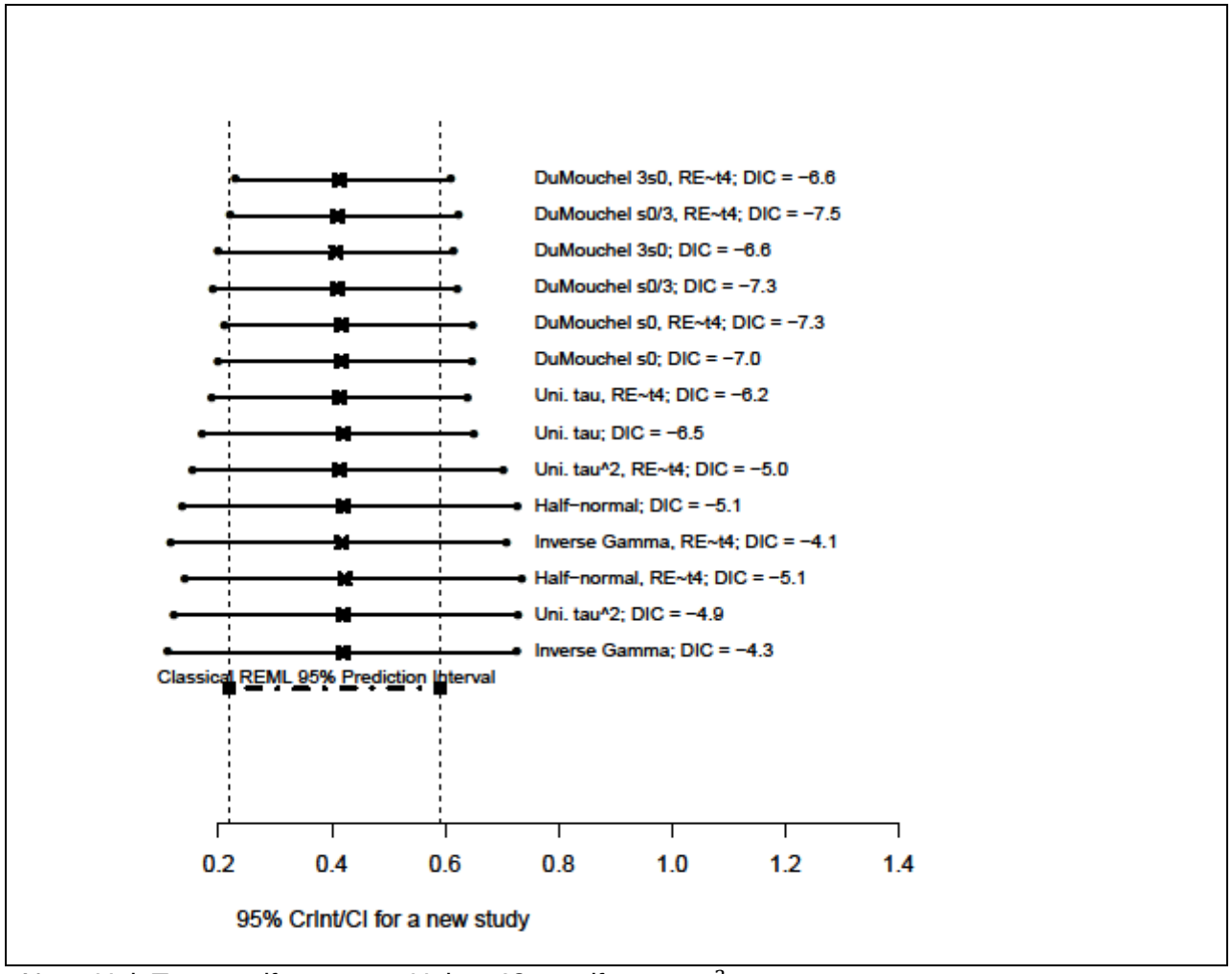
TEE Intercept: Comparison of Bayesian/Classical 95% Intervals



Note. Du. = DuMouchel. Uni. Tau = Uniform on τ. HN = half-normal. Uni tau^2 = uniform τ^2. IG = inverse gamma. Tau^2 = τ^2.

Figure 8

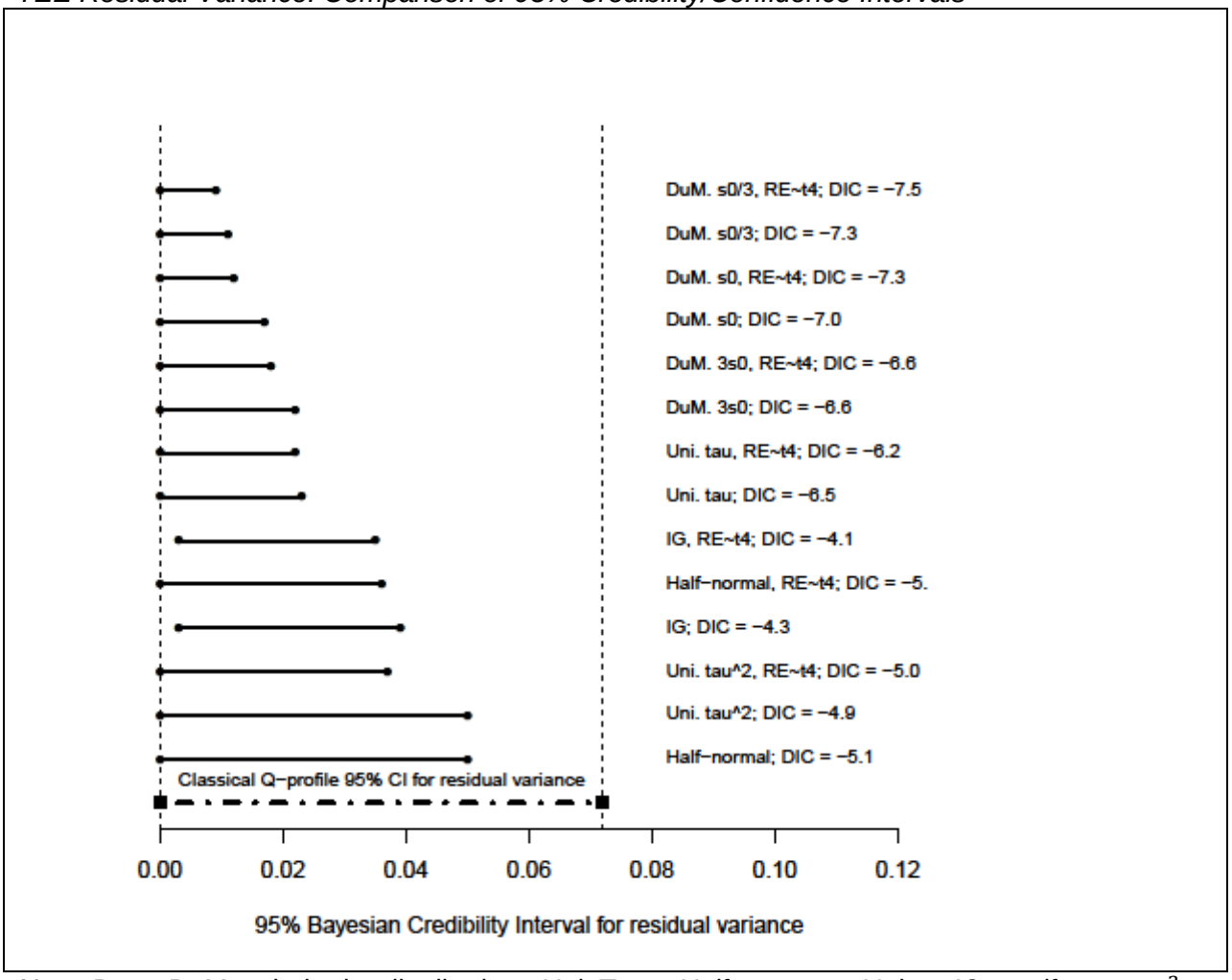
TEE Prediction Interval: Comparison of Bayesian/Classical 95% Intervals



Note. Uni. Tau = uniform on τ. Uni tau^2 = uniform on τ^2.

Figure 9

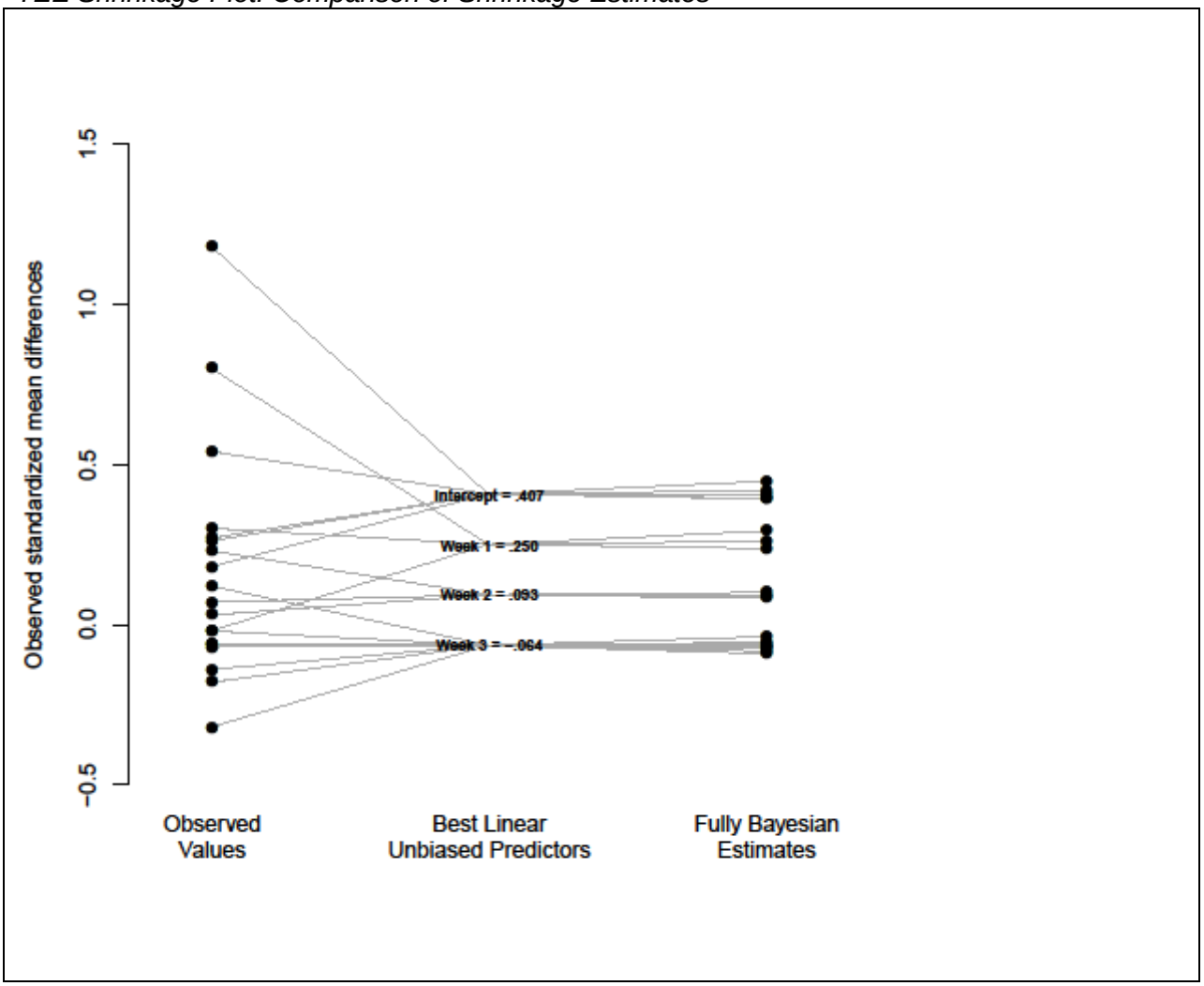
TEE Residual Variance: Comparison of 95% Credibility/Confidence Intervals



Note. Du. = DuMouchel prior distribution. Uni. Tau = Uniform on τ. Uni tau^2 = uniform on τ^2. IG = inverse gamma.

Figure 10

TEE Shrinkage Plot: Comparison of Shrinkage Estimates



Data Set 2: Effects of Academic Detracking

Classical Meta-Analysis Results for Detracking

The meta-analysis results of Rui (2009) were successfully replicated with the *metafor* package in R 2.11.1 with the use of a simple FE model ($\hat{d}$ = 0.087, *k* = 22, *p* < .001) and a simple RE model with D-L estimation ($\hat{d}$ = 0.202, k = 22, *p* =.009, τ = 0.317 , I^2 = 94.04%), with both models revealing a significant effect in favor of detracking. However, the simple RE model using REML estimation revealed a non-significant effect of detracking ($\hat{d}$ = 0.312, *k* = 22, *p* = .080, τ = 0.809, I^2 = 99.04%). The *Q*-test for heterogeneity in the simple RE models was significant at *p* < .001, indicating the presence of significant variation between effect sizes, suggesting that the estimates from a random/mixed-effects model would have greater credibility. It is important to note that in this data set there is nesting of the effect sizes (*k*=22) within the 10 studies. For example, some of the studies that report multiple effect sizes may have assigned different grade levels to detracking, thus resulting in the nesting of individuals within grades/classrooms and within studies. Hedges (2009) discusses how meta-analytic models that do not take into account the clustering at the different levels may lead to substantial bias in the variance of the effect size estimates. In order to make inferences from the identical data set that Rui used, consideration of the nesting of the effect sizes was omitted from the reanalysis, and, therefore the estimates of the standard errors and effect sizes may not be entirely valid or appropriate.

Coding of moderator variables for the mixed-effects model

For the reanalysis the type of study design was made into a dummy coded variable that included the study-specific information for the three separate categories of study design: (a) experimental, (b) quasi-experimental, and (c) observational. Rui's coding of student ability type was dummy coded in order to reflect the four categories of students: (a) low-ability, (b) medium-ability, (c) high-ability, and (d) all-ability. Rui coded and provided information for two additional

potential covariates in his systematic review: (a) grade level and (b) academic subject, although he did not perform moderator analyses of these variables. These variables were coded and included as potential covariates in the reanalysis as follows where: (a) $X_i = 0$ for Grade 6 or below and $X_i = 1$ for Grade 7 and above and (b) $X_i = 0$ if the subject was math/science and $X_i = 1$ if the subject was reading/English/Social Science.

Table 11

Detracking: Recoded Moderator Variable Data

Author	Y_i	se_i	Subj[a]	Grade[b]	Quasi[c]	Obs.[d]	Mid[e]	High[f]	All[g]
Cartwright	0.19	0.251	0	0	0	0	0	0	0
Cartwright	1.296	0.282	0	0	0	0	0	0	0
Cartwright	0.414	0.286	0	0	0	0	0	0	0
Marascuilo	-0.052	0.174	1	1	0	0	0	1	0
Marascuilo	0.687	0.227	1	1	0	0	1	0	0
Marascuilo	0.478	0.166	1	1	0	0	0	0	0
Slavin a	-0.575	0.13	0	0	0	0	0	0	1
Slavin b	-0.74	0.126	0	0	0	0	0	0	1
Thacker	0.141	0.278	1	0	1	0	0	1	0
Thacker	0.301	0.304	1	0	1	0	0	0	0
Thacker	0.393	0.268	1	0	1	0	1	0	0
Argys	-0.44	0.061	0	1	0	1	0	1	0
Argys	0.25	0.063	0	1	0	1	0	0	0
Kissoon	1.772	0.307	0	1	0	0	0	1	0
Kissoon	3.543	0.409	0	1	0	0	1	0	0
Hawkins	0.171	0.032	1	1	0	1	0	0	1
Hallinan	-0.702	0.13	1	1	0	1	0	1	0
Hallinan	-0.115	0.084	1	1	0	1	0	0	0
Mulkey	0.14	0.037	0	1	0	1	0	1	0
Mulkey	0.065	0.037	0	1	0	1	0	0	0
Burris	0.283	0.06	0	1	1	0	0	0	1
Burris	0.224	0.081	0	1	1	0	0	1	0

Note. Adapted from "Four decades of research on the effects of detracking reform: Where do we stand? a systematic review of the evidence", by N. Rui, 2009, *Journal of Evidence-Based Medicine*, *2*(3), p. 168-170. [a]0 = Math/Science; 1 = Reading/English/Social Science. [b]0 = Grade 6 or below; 1 = Grade 7 or above. [c]0 = Experimental or Observational; 1 = Quasi-experimental. [d]0 = Experimental or Quasi-experimental; 1 = Observational. [e]0 = High-, low-, or all-ability; 1= Mid-ability. [f]0 = Mid-, low-, or all-ability; 1 = High-ability. [g]0 = Mid, low, or high-ability; 1 = All-ability.

Mixed-Effect Model with REML, k = 22 effect sizes

In order to investigate the effects of the moderator variables, instead of conducting separate subset meta-analyses of the data as Rui (2009) did, all of the potential moderator variables were entered into a hierarchical linear model (hlm), and then the moderators were removed sequentially using backward-stepwise model selection procedure with REML estimation. This procedure resulted in a final mixed-effect model with estimates of $\hat{\beta}_0 = 0.141$, p = .394, 95% CI (-0.183, 0.466); $\hat{\beta}_1$(Medium-Ability) = 1.288, p = .006, 95% CI (0.365, 2.21); and τ = 0.701. In this mixed-effects model, the estimated amount of residual variance is now 0.491, suggesting that 25% of the total heterogeneity from the simple RE model can be accounted for by including the study-level covariate, medium-ability. As shown in Table 14, the mixed-effect model with REML had the best model fit statistics (i.e., lowest AIC and BIC), representing improvement over the simple RE (D-L) model that was used by Rui in his original meta-analysis.

Let $X_i\beta = \hat{\beta}_0 + \hat{\beta}_{i1}(Mid)_i$, where the estimated intercept, $\hat{\beta}_0 = 0.141$, represents the average mean effect of detracking in terms of standardized mean differences for students who are classified as high-, all-, or low-achieving. In this hlm the intercept is not significant, indicating that for high-, all-, or low-achieving students there is no appreciable effect of detracking. The linear combination estimate for the mid-achieving students is 1.430 (*s.e.* = .441, p < .001), indicating that there is a large effect of detracking for the mid-achieving students.

In terms of publication bias, the funnel plot for the mixed-effects model shows asymmetry, and the regression test for funnel plot asymmetry using se_i as a predictor was significant at p = .005. The Orwin Fail-safe N value for a simple RE model was 22, indicating that 22 studies with an average of a null result would have to be added to this data set in order to reduce the simple RE unweighted average effect size of .351 in half to a target effect size value of .176. For a simple RE model the trim and fill method estimated that 7 studies were

missing on the right side (i.e., in favor of detracking) of the funnel plot, and, if these studies were imputed, the meta-analytic inference would change, and there would be an appreciable effect of detracking with estimates for the simple RE model with REML of: $\hat{\beta}_0 = 0.649$, $p < .001$, 95% CI (0.300, 0.999), and τ= .943. The practical impact of the publication bias can be termed modest to severe using the criteria proposed by Rothstein, et al., 2005, as the significance of the effect of detracking changes (i.e., the *p* value changes from .08 to <.001) for the simple RE model when the missing studies are imputed.

Regarding the sensitivity of the different mixed-effects REML classical meta-analytic models, the coefficients for mid-ability is no longer significant (p=.096) when $\tau^2$was specified to its upper bound of its 95% CI resulting in 'modest to severe' impact as the fitted model changed. When the Knapp and Hartung (2003) adjustment is applied to the standard errors and test statistics for the regression coefficients, the standard errors are larger and the testing of the significance of the regression coefficients is more conservative, although the meta-analytic inferences do not change. For the FE model the estimate for $\hat{\beta}_0$ is likely to be inaccurate and its standard error too small, given the large amount of heterogeneity that exists among the effect sizes. Furthermore, the D-L (method of moments) simple RE model produced estimates for $se_\mu$and τ that were half the size of the estimates from the simple REML RE model, resulting in differing interpretations of the overall weighted mean effect of detracking and its precision (i.e., the D-L RE model shows evidence in favor of detracking, p = .009, while the REML simple RE model has a p = .079, reflecting that there is not a significant effect of detracking).

An outlier sensitivity analysis was conducted on the fitted mixed-effects (REML) model, revealing that the study by Kissoon-Singh (1996), which used computer-based instruction methods, showed evidence of being an outlier, as shown by visual inspection of the forest plot given by Figure 11 and the *Q*-*Q* Plot given by Figure 12. The Kissoon-Singh study contributed two effect sizes to the data (i.e., an effect size of d = 1.772 for high-ability students and d = 3.543 for mid-ability students). As displayed in Table 13, the leave-one-out case diagnostics for

the Kissoon-Singh study had *r*-student values of 2.446 and 3.840 and covariance values of 0.650 and 0.503. When the Kissoon-Singh study is excluded from the model, the estimates for $\hat{\beta}_0$ change by 0.573 and -.138 and the estimate for $\hat{\beta}_1$ change by -0.138 and 2.348 in terms of standard deviation units, respectively. These leave-one-out case diagnostics suggest that, when the Kissoon-Singh study is excluded from the model, the resulting model may have better model fit and greater precision and the coefficient for $\hat{\beta}_1$ will no longer be significant. Furthermore, as shown in Table 12, even after controlling for all of the potential covariates, when the Kissoon-Singh study was dummy variable coded and included in the meta-analytic model, this study still differs significantly from the other studies. For these reasons the decision was made to perform two re-analyses of the Detracking data with: (a) the Kissoon-Singh included and (b) it excluded.

Figure 11

Detracking: Forest Plot for Mixed-effects Model (k = 22, REML)

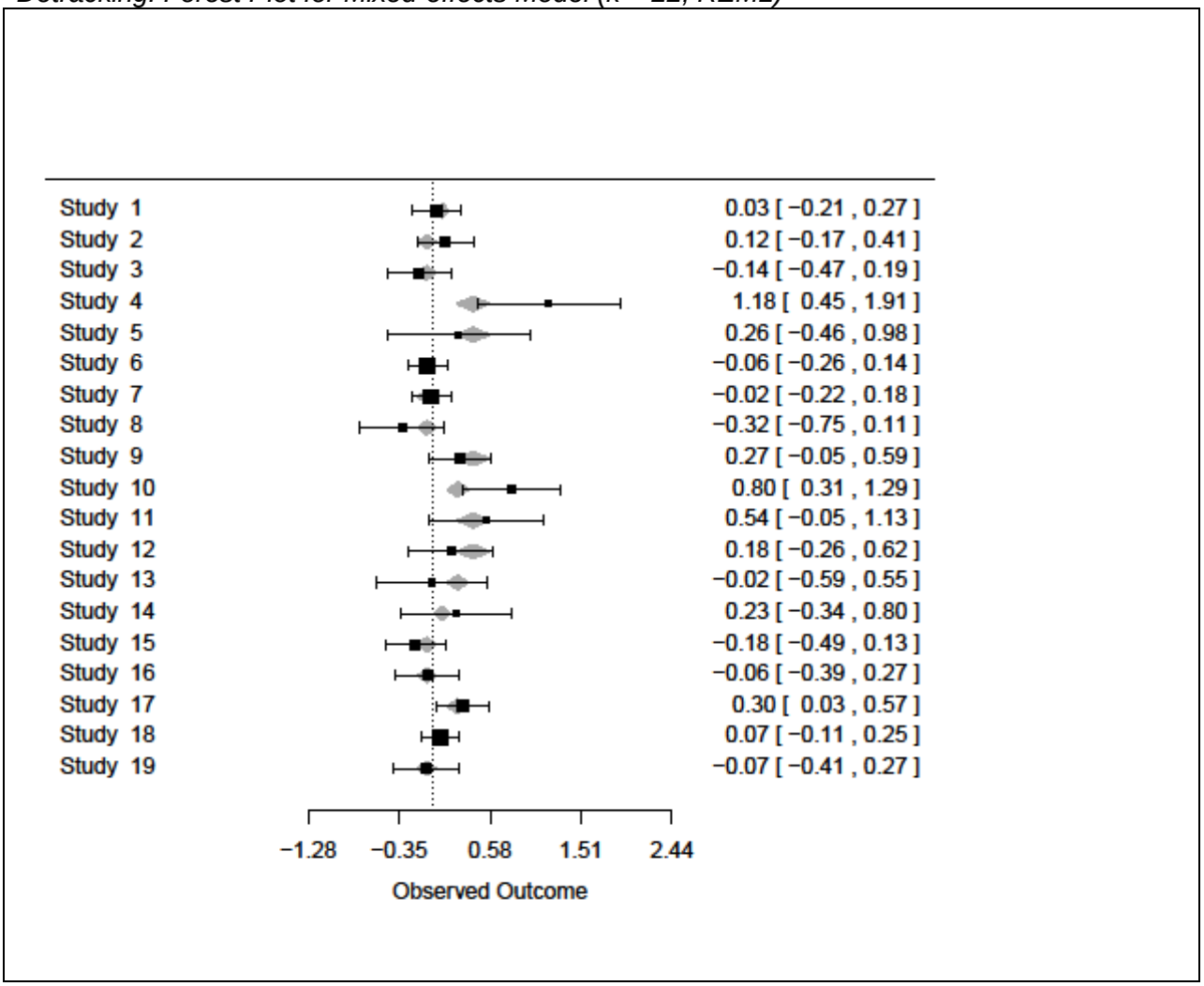


Figure 12

Detracking: Normal Q-Q Plot (k = 22, all studies)

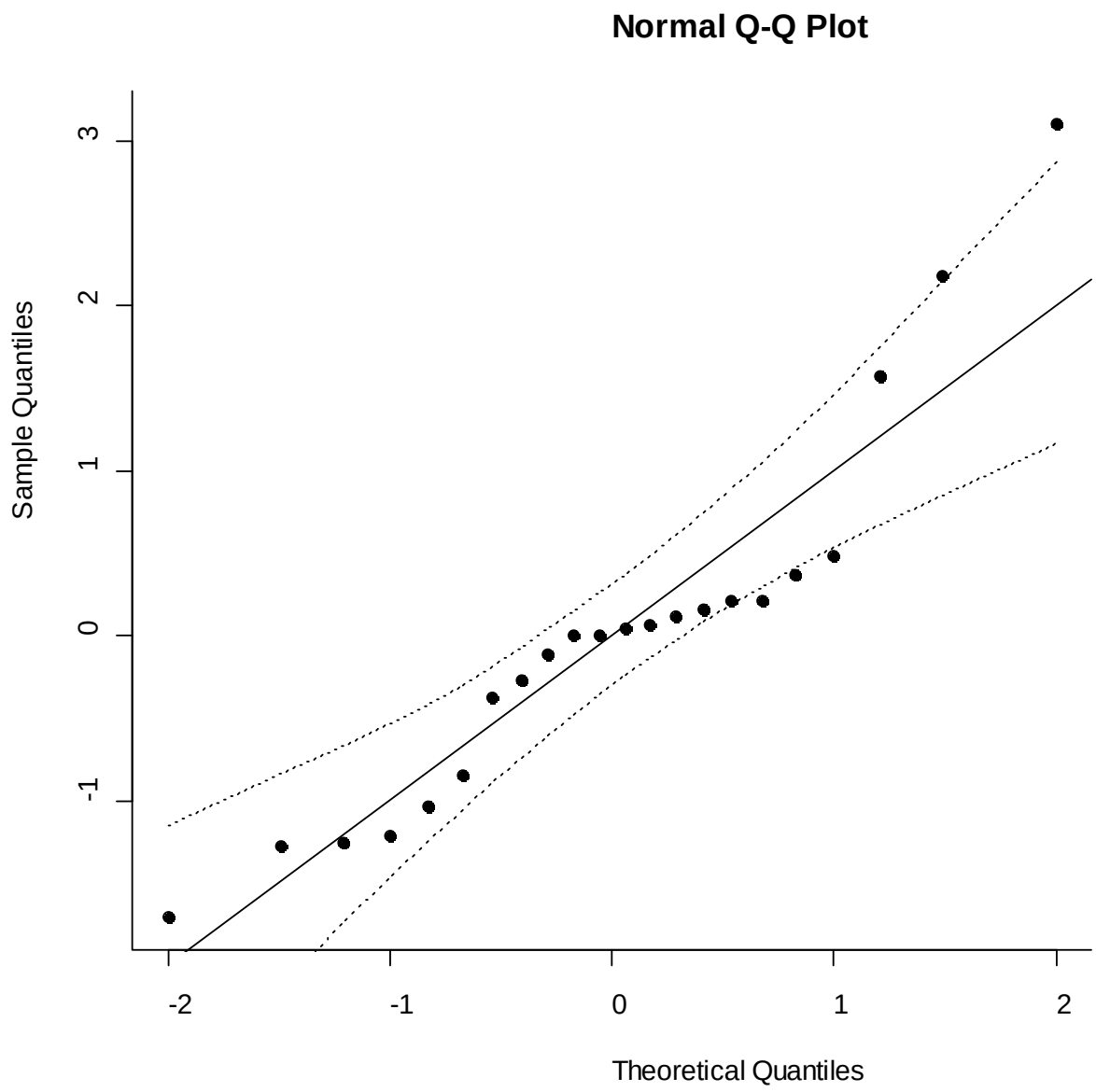


Figure 13

Detracking: Funnel Plot for Mixed-effects Model (k=22)

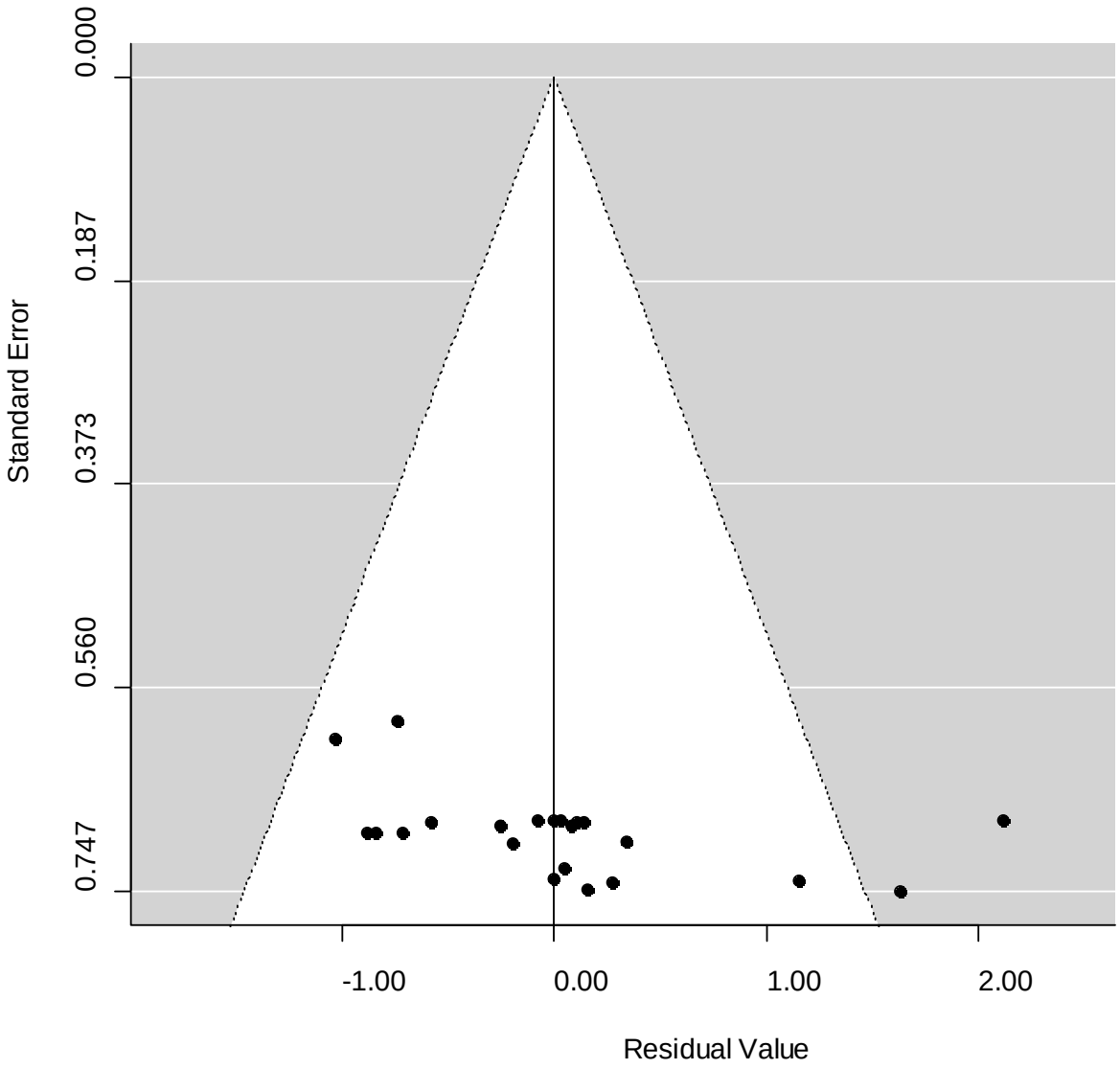


Table 12

Detracking: Coefficients for the Mixed-effects Model (k = 22)

	Mean	*se*	*p* value
Intercept	0.303	0.215	.160
Subject	-0.086	0.204	.672
Grade 7 or above	0.302	0.254	.226
Mid-Ability	0.066	0.392	.867
High-Ability	-0.541	0.239	.024
All-Ability	-0.603	0.252	.017
Quasi-experimental	0.220	0.256	.390
Observational	-0.337	0.265	.203
Kissoon-Singh (High)	1.704	0.561	.002
Kissoon-Singh (Mid)	2.868	0.700	.000

AIC = 46.00, BIC= 51.33, τ = 0.379.

Table 13

Detracking: Leave-one-out Case Diagnostics for Fitted Model (k=22)

Study	*r*student[a]	cov.r[b]	τ^2.del[c]	*Q*.del[d]	B_0[e]	β_1[e]
Cartwright	0.064	1.177	0.523	317.634	0.008	-0.014
Cartwright	1.640	0.875	0.443	299.110	0.380	-0.112
Cartwright	0.360	1.153	0.517	316.448	0.075	-0.036
Marascuilo	-0.269	1.181	0.523	317.275	-0.070	0.014
Marascuilo	-1.267	1.513	0.483	313.656	0.002	-0.885
Marascuilo	0.468	1.156	0.516	311.965	0.105	-0.046
Slavin	-1.034	1.056	0.490	292.233	-0.247	0.087
Slavin	-1.297	0.984	0.471	275.019	-0.306	0.115
Thacker	-0.002	1.174	0.522	317.783	-0.007	-0.008
Thacker	0.208	1.162	0.520	317.294	0.040	-0.024
Thacker	-1.763	1.304	0.449	309.435	0.011	-1.189
Argys	-0.842	1.103	0.502	240.408	-0.206	0.068
Argys	0.152	1.195	0.525	309.834	0.029	-0.022
Kissoon	2.446	0.650	0.375	287.294	0.573	-0.138
Kissoon	3.840	0.503	0.275	272.858	0.109	2.348
Hawkins	0.040	1.201	0.526	306.454	0.002	-0.013
Hallinan	-1.234	1.002	0.476	281.301	-0.292	0.108
Hallinan	-0.364	1.181	0.522	312.371	-0.095	0.022
Mulkey	-0.004	1.201	0.526	314.339	-0.009	-0.009
Mulkey	-0.110	1.200	0.526	317.690	-0.034	0.000
Burris	0.199	1.192	0.524	305.220	0.041	-0.026
Burris	0.115	1.196	0.525	314.439	0.020	-0.019

Note. [a]denotes the *r*student (externally standardized residual) value with the i^{th} study excluded. [b]denotes the covariance ratio. [c]denotes the estimate value of tau-squared based on the data set with the i^{th} study excluded. [d]denotes the *Q* statistic for the test of heterogeneity based on the data set with the i^{th} study removed. [e]denotes by how many standard deviation units the regression coefficients change when the i^{th} study is excluded (Viechtbauer, 2009).

Table 14

Detracking: Classical Meta-analysis Results (k=22)

Estimation Method	FE	D-L	REML	REML	$REML_{KH}$[a]	$REML_L$[b] τ^2 =.389	$REML_H$[c] τ^2=1.64
Model	Simple	Simple	Simple	Mixed	Mixed	Mixed	Mixed
$\hat{B}_0$	0.087	0.202	0.311	0.141	0.141	0.136	0.156
s.e.	0.016	0.077	0.178	0.166	0.173	0.149	0.293
p value	<.001	0.009	0.079	.394	.815	.358	.595
95% CI low	0.056	0.051	-0.037	-0.183	-0.220	-0.155	-0.419
95%CI High	0.119	0.353	0.660	0.466	0.503	0.428	0.731
$\hat{B}_1$ Mid-ability				1.288	1.288	1.271	1.345
s.e.				0.471	0.493	0.427	0.808
p value				.006	.017	.003	.096
95% CI low				0.365	0.259	0.434	-0.238
95% CI high				2.212	2.317	2.108	2.928
τ		0.316	0.809	0.701	0.701	0.624	1.266
95% CI low		0.624	0.624	0.533	0.533	0.533	0.624
95% CI high		1.266	1.266	1.112	1.112	1.112	1.266
τ^2		0.100	0.655	0.491	0.491	0.389	1.604
s.e.			0.214	0.166	0.166	0.134	0.507
95% CI low		0.389	0.389	0.284	0.284	0.284	0.389
95%CI high		1.604	1.604	1.236	1.236	1.236	1.604
Q	352.18	352.18	352.18	352.18		352.18	352.18
p value	< .001	< .001	< .001	< .001		< .001	< .001
PI $\hat{B}_0$ Lower[d]			-1.414	-1.360	-1.364	-1.201	-2.369
PI $\hat{B}_0$ Upper[e]			2.040	1.643	1.650	1.474	3.036
PI $\hat{B}_1$Lower				-0.473	-0.499	-0.305	-1.788
PI $\hat{B}_1$Upper				3.049	3.075	2.848	4.479
I^2		94.04%	99.04%				
Q_{Mods}				7.475	6.82(*F*)	8.862	2.772,
p value				.006	.017	.003	.096
Log-like	-151.96	-42.909	-28.552	-24.985	-24.985	-25.186	-31.211
Deviance	303.92	85.819	57.105	49.970	49.970	50.373	62.422
AIC	305.92	89.819	61.105	55.970	55.970	56.373	66.422
BIC	307.01	92.001	63.194	58.958	58.958	59.360	68.511
Min, max $\theta_i^*(\tau)$			-0.715 2.885	-0.712 3.006	-0.712 3.006	-0.706 2.901	-0.731 3.380

Note. FE= Fixed-effect model. D-L = DerSimonian & Laird estimation. REML = Restricted Maximum Likelihood Estimation. [a]Knapp and Hartung (2003) adjustment. [b]$REML_L = \tau^2$ was set equal to the lower bound of its 95% CI from the simple REML model. [c]$REML_H = \tau^2$ was set equal to the upper bound of its 95% CI from the simple REML model. [d,e] Upper and lower bound of the Higgins et al. (2009) approximate prediction interval. The permutation test *p* value for the REML mixed-effects model $\hat{B}_1$ was <.032.

Mixed-Effect Model with REML, k = 20 effect sizes

As shown in Table 15, after exclusion of the Kissoon-Singh (1996) study, the final model resulted in a significant intercept and two significant moderators (high-ability and all-ability). Let $X_i\beta = \hat{\beta}_0 + \hat{\beta}_{i1}(High)_i + \hat{\beta}_{i2}(All)_i$, where the estimated intercept, $\hat{\beta}_0 = 0.359$, represents the average mean effect of detracking in terms of standardized mean differences for students who are low- or mid-achieving. Now, by including these two study-level moderators in the model, the estimate for the residual variance is 0.143, resulting in an explanation of 23% of the heterogeneity variance from the simple RE model (where τ^2 = 0.186). In this hierarchical linear model the intercept is significant at *p* = .008, indicating that the low- and mid-achieving detracked students performed better than students who were not detracked. The linear combination estimate for the high-achieving students is -0.124, *se* = 0.164, *p* = .452, and for the all-achieving students it is -0.195, *se* = 0.164, and *p* = .317, indicating that there is not an appreciable effect of detracking for students who are in the high- or all-achieving groups.

The sensitivity analysis of the different mixed-effects REML classical meta-analytic models, shows that the study-level coefficients are no longer significant when $\tau^2$was set to the upper bound of its 95% CI ($REML_H$) resulting in a 'modest to severe' impact on meta-analytic conclusions as the fitted model changed. The model fit statistics (i.e., AIC and BIC) were at their smallest values with the mixed-effects REML and $REML_H$ models. When the Knapp and Hartung (2003) adjustment was applied to the standard errors and test statistics for the model coefficients, all coefficients remain significant and the overall meta-analytic inferences did not change, although the corresponding *p* values were not quite as small.

For this data set the Orwin Fail-safe *N* value for the simple RE model was 20, which indicates that 20 studies with an average of a null result would be required in order to reduce the average unweighted effect size (simple model) from .121 by half, to a target effect size value of .060. In terms of publication bias, the funnel plot for the mixed-effects model shows improved symmetry, and the regression test for funnel plot asymmetry using se_i as a predictor was non-

significant (*p*=.412). For a simple RE model the trim and fill method estimated that 4 studies were missing on the left side of the funnel plot (i.e., studies in favor of tracking) and, if these studies were imputed, the estimates for the simple R model with REML would be $\hat{\beta}_0 = -0.041$, *p* = .712, and 95% CI (-0.260, 0.177). Here, with the *k*=20 data set, the impact of publication bias is termed minimal, as both versions of the simple RE meta-analysis yield similar results and the same meta-analytic conclusions.

Table 17 displays the statistical power of the covariates (moderators) to be able to detect differences that may exist among the effect sizes due to the study-level characteristics. The statistical power is low (below .423) for all of the potential moderator variables. It is important to note that the low power of these covariates may lead to misleading moderator tests, because the tests may not be sufficiently powerful enough to detect real effects, even if such real effects were present (Hedges and Pigott, 2004).

Table 15

Detracking: Classical Meta-analysis Results (k=20)

Estimation Method	FE	D-L	REML	REML	$REML_{KH}$	$REML_L$ τ^2= .069	$REML_H$ τ^2=.453
Model	Simple	Simple	Simple	Mixed	Mixed	Mixed	Mixed
$\hat{B}_0$	0.077	0.062	0.089	0.358	0.358	0.331	0.383
s.e.	0.016	0.069	0.104	0.136	0.134	0.103	0.223
p value	<.001	.369	.389	.008	.016	.001	.087
95%CI low	0.046	-0.073	-0.114	0.092	0.075	0.128	-0.055
95% CI High	0.109	0.196	0.292	0.625	0.642	0.533	0.820
$\hat{B}_1$ High-				-0.482	-0.482	-0.457	-0.502
s.e.				0.214	0.211	0.158	0.359
p value				.024	.035	.004	.162
95% CI low				-.901,	-0.926	-0.767	-1.205
95% CI high				-.064	-0.038	-0.147	0.202
$\hat{B}_2$ All-				-0.554	-0.554	-0.507	-0.591
s.e.				0.238	0.235	0.173	0.407
p value				.020	.031	.003	.146
95% CI low				-1.021	-1.049	-0.847	-1.389
95% CI high				-.088	-0.058	-0.167	0.206
τ		0.265	0.431	0.379	0.379	0.263	0.673
95% CI low		.312	.312	.262	.262	0.262	0.262
95% CI high		.673	.673	.588	.588	0.588	0.588
τ^2		0.070	0.186	.143	.143	0.069	0.453
s.e.			0.069	0.058	0.058	0.031	0.166
95% CI low		.098,	.098,	.069	.069	0.069	0.069
95%CI high		.453	0.453	.346	.346	0.346	0.346
Q		250.29	250.29	233.2	233.2	233.2	233.2
p value		*p*<.001	*p*<.001	*p*<.001	*p*<.001	*p*<.001	*p*<.001
PI $\hat{B}_0$ Lower			-.843,	-.486,	-.484	-0.262,	-1.107
PI $\hat{B}_0$ Upper			1.021	1.204	1.202	.923	1.873
PI $\hat{B}_1$Lower				-1.395,	-1.29,	-1.101,	-2.104
PI $\hat{B}_1$Upper				.430	.428	.187	1.101
PI $\hat{B}_2$Lower				-1.494,	-1.489,	-1.168,	-2.244
PI $\hat{B}_2$Upper				.385	.382	.154	1.061
I^2		92.41%	96.99%				
Q_{Mods}				7.724	3.957(*F*)	12.177	3.957
p value				.021	.039	.002	.039
Log-like	-101.255	-16.044	-14.219	-11.809	-11.809	-13.706	-11.804
Deviance	202.510	32.088	28.439	23.619	23.619	27.411	23.608
AIC	204.510	36.088	32.439	31.619	31.619	35.411	31.608
BIC	205.505	38.079	34.328	34.952	34.952	38.744	34.941
Min, max $\theta_i^*(\tau)$			-0.843, 1.021	-.686, .962	-0.685 0.960	-0.635 0.779	-0.722, 1.160

Note. FE= Fixed-effect model. D-L = DerSimonian & Laird estimation. REML = Restricted Maximum Likelihood Estimation. $REML_L$ = τ^2 was set equal to the lower bound of its 95% CI from the simple REML model. $REML_H$ = τ^2 was set equal to the upper bound of its 95% CI from the simple REML model. The permutation test *p* values for $\hat{B}_1$ and $\hat{B}_2$ were .038 and .030.

Figure 14

Detracking: Forest Plot for Mixed-effects Model (k=20), REML

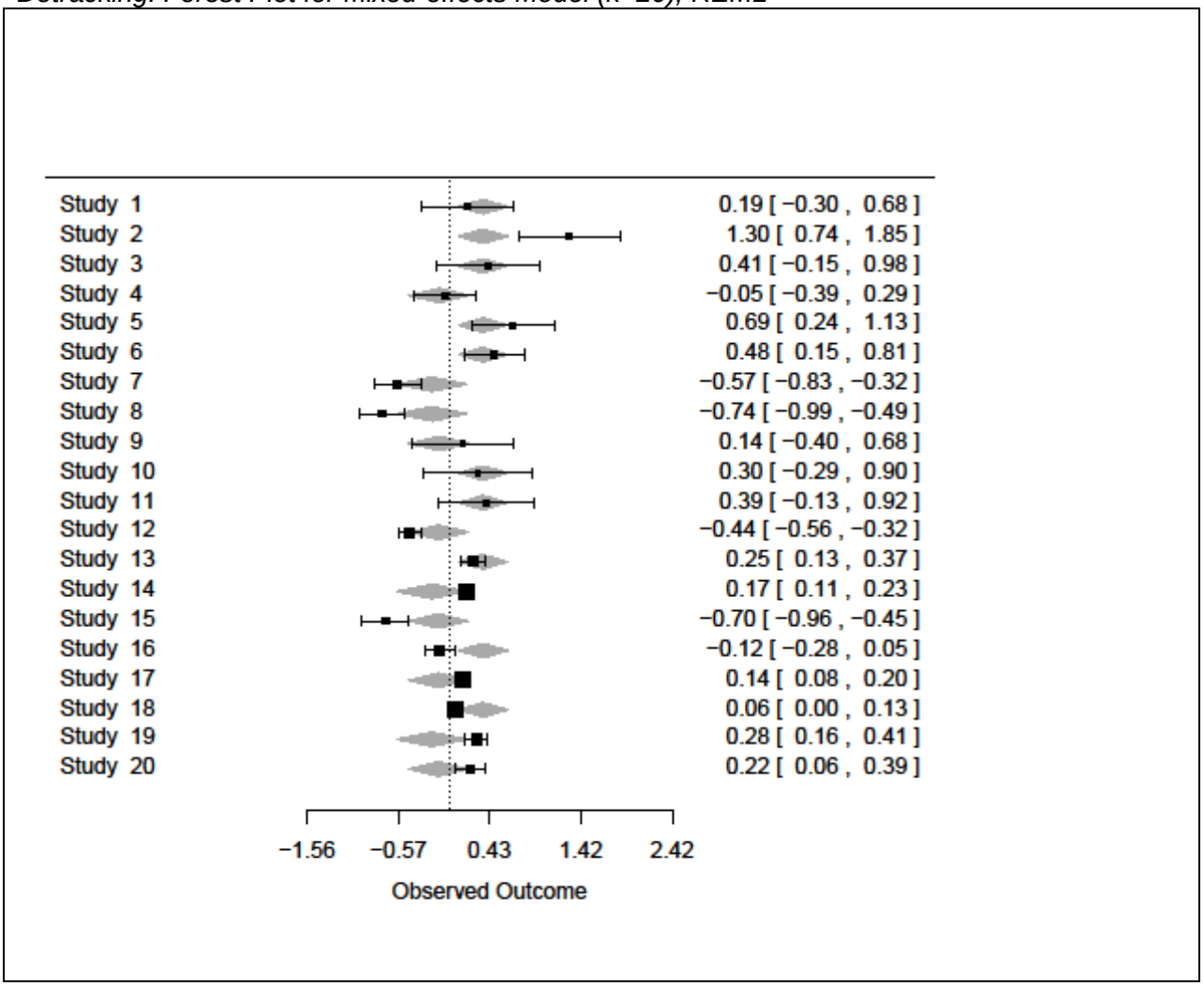


Figure 15

Detracking: Normal Q-Q Plot (k = 20), Mixed-effects, REML

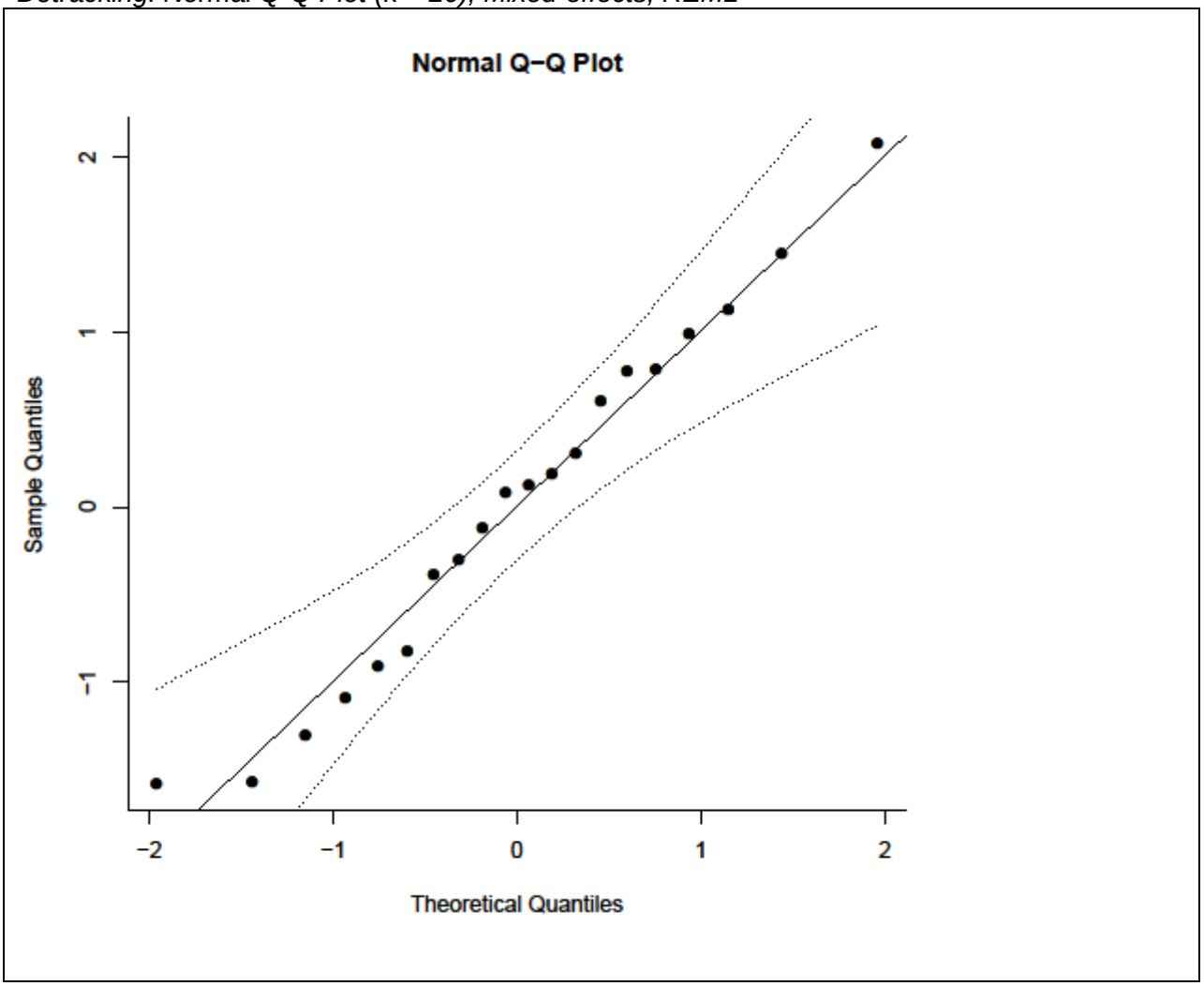


Figure 16

Detracking: Funnel Plot for Mixed-Effects Model (k=20)

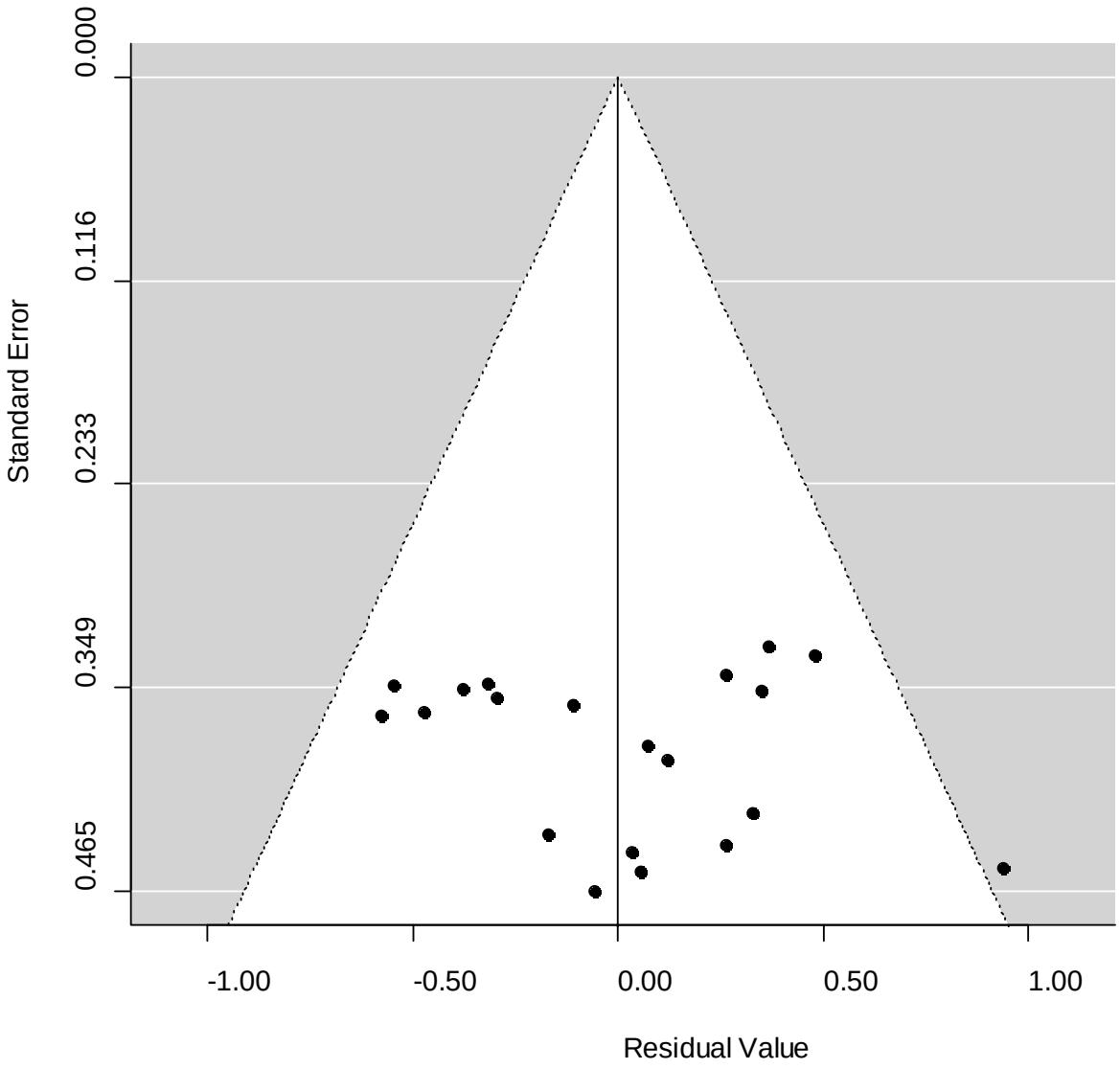


Table 16

Detracking: Correlation among Study-level Covariates (all studies, k=22)

	Subject	G7 or above	Mid-ability	High-ability	All-ability	Quasi.	Obs.	Y_i
Subject	1							
Grade	.052	1.000						
Mid	.208	.025	1.000					
High	.027	.314	-.271	1.000				
All	-.153	-.134	-.187	-.322	1.000			
Quasi	.211	-.266	.101	.095	.026	1.000		
Observational	.027	.516	-.271	.162	-.069	-.370	1.000	
Y_i	-.192	.146	.528	-.150	-.298	-.050	-.336	1

Note. Quasi = quasi-experimental studies. Obs. = observational studies.

Table 17

Detracking: Power for Two-sided Test of the Regression Coefficients, k=20

	Power for Individual Regression Coefficients Backward Selection Regression Models					
Coefficient	Full Model	Model 1	Model 2	Model 3	Model 4	Final Model
Mid	.135					
Subject	.367	.423				
Quasi	.252	.268	.285			
Grade	.255	.270	.292	.290		
Observational	.238	.259	.277	.327	.420	
High	.282	.321	.345	.354	.347	.340
All	.259	.289	.318	.315	.294	.284
Intercept	.337	.365	.431	.484	.589	.680

Note. α = .05 with a clinically meaningful effect set equal to 0.33 standard deviations.

Bayesian Meta-Analysis Results for Detracking

Bayesian Model, k=22

In WinBUGS with all of the data points included, a backward stepwise model selection procedure with a uniform prior distribution for τ (0,5) and $\delta_i \sim Normal$ and $\delta_i \sim t_4$ distribution was used to fit the data. A sample of the Bayesian checklist used for the Bayesian model and the results for the uniform prior for τ (0,5) and $\delta_i \sim t_4$ are shown in Table 18 and Figure 17. Here, there is a 59% chance that the effect of detracking students in a new study will result in a treatment effect that is greater than zero (i.e., a beneficial effect for detracked students) for students classified as low-, high-, or all-achieving. For students classified as mid-achieving, the predicted effect size of detracking for mid-achieving students is 0.946 with a 91.1% probability that this effect is greater than zero. However, in a sensitivity analysis for the seven different prior distributions for τ, whenever $\delta_i \sim Normal$, the fitted model coefficient estimates, 95% credibility intervals, and DIC measures of model fit were larger, as shown in the tables in the Appendix. For example, with a uniform prior for τ and the random effect modeled from a normal distribution, the estimates are $\hat{\beta}_0 = 0.142$, $\hat{\beta}_1$=1.276, τ = .753, and DIC = -3.4. Modeling of the $\delta_i \sim t_4$ resulted in a 'modest to severe' impact on meta-analytic conclusions because the study-level covariate, $\hat{\beta}_1$, was no longer included in the fitted model, when using the criteria for inclusion of <.025 probability the study-level covariate, $\hat{\beta}_1$<0 or <.975 probabilty the study-level covariate, $\hat{\beta}_1$>0.

Bayesian cross-validation was conducted in S-Plus 2000 on the fitted model using a DuMouchel prior with all of the *k* = 22 studies. As depicted in Table 19 and Figure 18, Bayesian cross-validation revealed that both of the effect size data points from the Kissoon-Singh (1996) study were likely to be outliers, as their predictive probabilities were .999 and 1.000 and the Bonferroni significance for the more extreme residual was .012. For the Bayesian analyses the

Detracking data set was also analyzed in two manners: a) with the effect sizes from the Kissoon-Singh study included and b) with the effect sizes from Kissoon-Singh excluded.

Table 18

Detracking: Bayesian results (k=22)

Parameter	Posterior mean	*s.d.*	Prob > 0	95% Cr Int
$\hat{\beta}_0$	0.104	0.138	.800	-0.188, 0.396
$\hat{\beta}_1$	0.859	0.592	.947[a]	-0.153, 2.095
$\hat{\beta}_0 + \hat{\beta}_1$	0.962	0.572	.975	0.003, 2.253
$\theta_{new.\beta 0}$	0.118	0.590	.590	-1.008, 1.241
$\theta_{new.\ \beta 0+\beta 1}$	0.946	0.799	.911	-0.545, 2.738
τ	.535	0.133		0.316, 0.857

Note. Prior for τ ~ uniform (0,5). Prior for **β** ~ (0,1000). $\delta_i \sim t_4$. DIC = -4.9.

[a] When δ_i ~ *Normal*, the probability that $\hat{\beta}_1 > 0$ was .995.

Figure 17

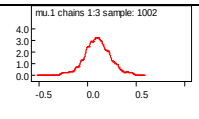
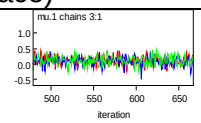
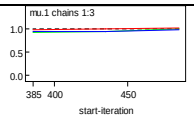
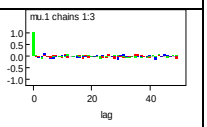
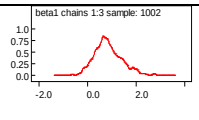
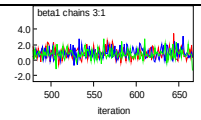
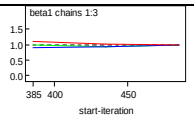
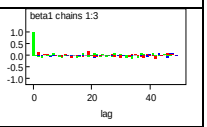
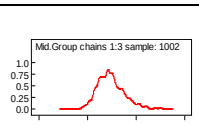
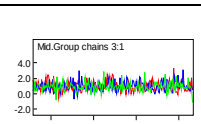
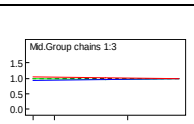
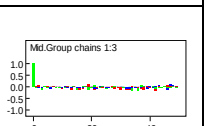
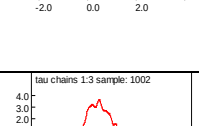
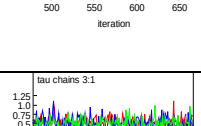
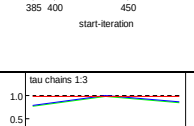
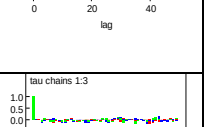
Detracking: Bayesian Quality Assurance Meta-analysis Checklist (k = 22)

Cross-validation ☒ Outlier Detected?: ☐No ☒ Yes
If yes, study = Kissoon Predictive Prob. = 1.000 Bonferroni significance = .012
Fitted model: $X_i\beta = \hat{\beta}_0 + \hat{\beta}_{i1}$(Mid-achieving)
Prior Distribution for τ:
☒ Uniform on τ ☐ Uniform on τ^2 ☐ Half-norm ☐ Inverse Gamma ☐ DuM. s_0 ☐ DuM. $s_0/3$ ☐ DuM. $3s_0$
Random effect modeled from: Prior Distribution for **β**:
☐ Normal distribution ☒ t_4-distribution ☒ Normal (0,1000)
Model specification:
☒ Run 3 chains ☒ 50,000 iterations ☒ Discard first half ☒ DIC = - 4.880
Do meta-analytic inferences change from the fitted hblm?: ☐No ☒Yes.If yes, impact: Modest

Parameters to monitor:

Node	Mean	95% Cr Int	Set parameter	MC error < 5% of s.d.	Rhat proximity to 1
β_0	0.104	(-0.188, 0.396)	☒	☒	☒
β_1	0.859	(-0.153, 2.095)	☒	☒	☒
τ	0.535	(0.316, 0.857)	☒	☒	☒
$\beta_0 + \beta_1$	0.962	(0.003, 2.2530	☒	☒	☒
$\theta_{new\beta0}$	0.118	(-1.008, 1.241)	☒	☒	☒
$\theta_{new\beta0 + \beta1}$	0.946	(-0.545, 2.738)	☒	☒	☒
$\theta_{new\beta1}$	0.869	(-0.616,2.565)	☒	☒	☒
Deviance	-25.56	(-37.16,-11.39)	☒	☒	☒
p $\beta_0 < 0$	.200	(0,1)	☒	☒	☒
p $\beta_1 < 0$	.053	(0,1)	☒	☒	☒
p $\beta_0+\beta_1 < 0$	.025	(0,1)	☒	☒	☒
p $\theta_{new\beta0} < 0$	.410	(0,1)	☒	☒	☒
p $\theta_{new\beta0 + \beta1} < 0$	.089	(0,1)	☒	☒	☒
P $\theta_{new\beta1} < 0$	.134	(0,1)	☒	☒	☒
θ_i^*			☒	☒	☒

Items pertaining to parameter diagnostic checks:

	Kernel Density	History (Dynamic Trace)	BGR	Autocorrelation
☒ β_0				
☒ β_1				
☒ $\beta_0+\beta_1$				
☒ τ				

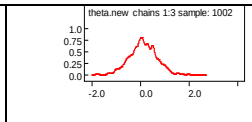
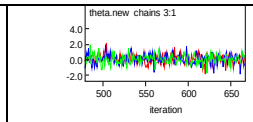
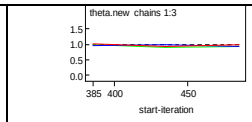
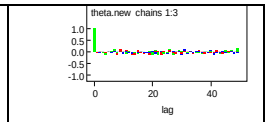
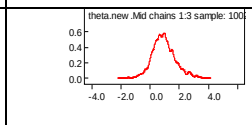
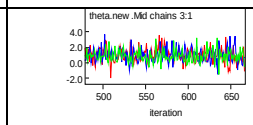
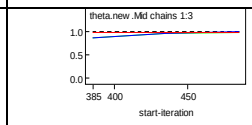
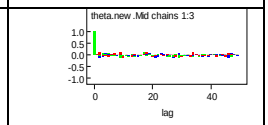
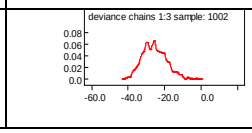
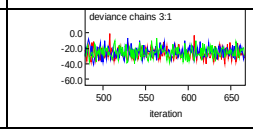
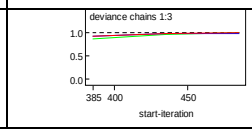
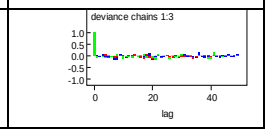
☒θ_{new} for β_0				
☒θ_{new} for $\beta_0 + \beta_1$				
☒ Deviance				

Table 19

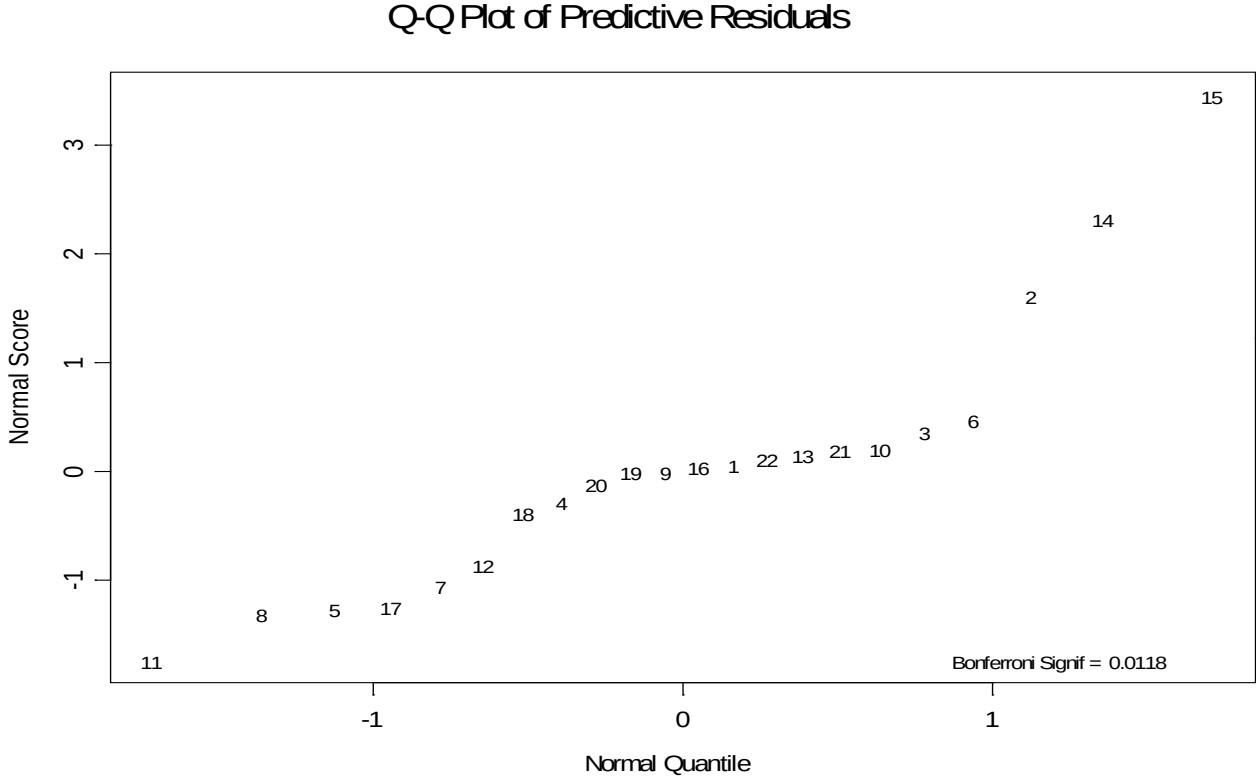
Detracking: Bayesian cross-validation results, k = 22

Study Number	Y_i	se_i	Prior mean	Prior *s.d.*	Pred. prob[a]	τ^2	Post.[b] mean	Post. *s.d.*
1	0.19	0.251	0.138	0.762	0.527	0.548	0.184	0.236
2	1.296	0.282	0.080	0.701	0.947	0.463	1.124	0.267
3	0.414	0.286	0.127	0.757	0.644	0.542	0.372	0.265
4	-0.052	0.174	0.152	0.762	0.393	0.548	-0.040	0.169
5	0.687	0.227	1.836	0.905	0.104	0.506	0.761	0.220
6	0.478	0.166	0.122	0.757	0.685	0.541	0.458	0.162
7	-0.575	0.13	0.181	0.738	0.148	0.514	-0.549	0.128
8	-0.74	0.126	0.189	0.723	0.097	0.493	-0.710	0.124
9	0.141	0.278	0.141	0.761	0.501	0.547	0.140	0.258
10	0.301	0.304	0.133	0.759	0.585	0.545	0.273	0.279
11	0.393	0.268	1.959	0.872	0.041	0.471	0.532	0.259
12	-0.44	0.061	0.174	0.747	0.196	0.527	-0.435	0.061
13	0.25	0.063	0.135	0.764	0.564	0.551	0.249	0.063
14	1.772	0.307	0.057	0.644	0.990	0.392	1.492	0.294
15	3.543	0.409	0.545	0.681	1.000	0.289	2.978	0.402
16	0.171	0.032	0.139	0.765	0.518	0.553	0.171	0.032
17	-0.702	0.13	0.187	0.727	0.108	0.498	-0.672	0.128
18	-0.115	0.084	0.156	0.761	0.355	0.548	-0.111	0.083
19	0.14	0.037	0.141	0.765	0.500	0.553	0.140	0.037
20	0.065	0.037	0.146	0.765	0.457	0.553	0.065	0.037
21	0.283	0.06	0.133	0.763	0.583	0.550	0.282	0.060
22	0.224	0.081	0.136	0.764	0.549	0.551	0.223	0.080

Note. [a]Predictive probability. [b]Posterior.

Figure 18

Detracking: Bayesian Q-Q Plot (k=22)



Bayesian Model, k = 20

With the Kissoon-Singh (1996) data points excluded, a backward stepwise model selection procedure with a uniform prior distribution for τ (0,5) and the $\delta_i \sim Normal$ and $\delta_i \sim t_4$ resulted in different meta-analytic inferences and a different fitted model. Here, with $\delta_i \sim t_4$, as shown in the Bayesian Quality Assurance Checklist and in the Appendix, the estimates were $\hat{\beta}_0 = 0.335$, *prob > 0* = .992; $\hat{\beta}_1$(High-achieving) = -0.439, *prob* < 0 = .972; $\hat{\beta}_2$(All-achieving) = -0.521, *prob* < 0 = .956, τ = 0.361, and DIC = -11.3.

For a new study the predicted effect for the intercept is 0.321 with an 81.6% probability that the effect is greater than zero. For students of high-ability ($\hat{\beta}_0 + \hat{\beta}_1$), the predicted effect for this group in a new study is -0.106 with a 41.6% probability that the effect in a new study is greater than zero. For students of all-ability ($\hat{\beta}_0 + \hat{\beta}_2$), the predicted effect for this group in a new study is -0.202 with a 31.1% probability that the effect in a new study is greater than zero.

The results for the 14 different Bayesian meta-analytic models (e.g., the seven different prior distributions on the heterogeneity variance for which the random effect is modeled from both a normal distribution and a t_4-distribution) are displayed in the Appendix (Bayesian Summary Results). The seven different prior distributions with the δ_i ~ normal distribution produced Bayesian model coefficient estimates that were slightly larger than those estimates where $\delta_i \sim t_4(0, \tau^2)$. For example, when a uniform on tau prior distribution is used with the δ_i ~ normal, the estimates are: $\hat{\beta}_0 = 0.358$, $\hat{\beta}_1$= -0.478, $\hat{\beta}_2$=-0.555, τ = 0.409, and DIC = -10.3.

A trace plot of τ that depicts the dependency of the intercept (curve labeled A) and the study-specific estimates (curves labeled B-U) on the different values of τ is displayed in Figure 20. As shown in the trace plot, the spread of the estimates shrink as τ approaches smaller values, and there is greater spread among the estimates when τ is at larger values. The Cartwright study (represented by label C) experiences the most shrinkage, as it had a large standard error.

Figure 19

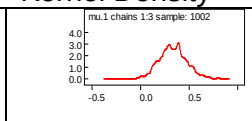
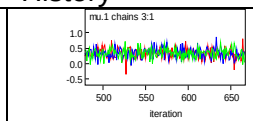
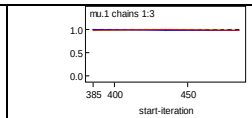
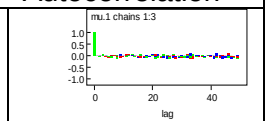
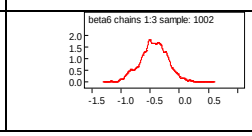
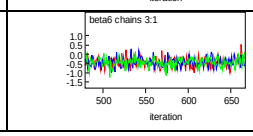
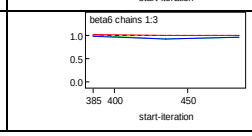
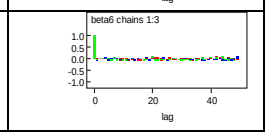
Detracking: Bayesian Quality Assurance Checklist (k = 20)

Cross-validation ☐ Outlier Detected?: ☒No ☐ Yes
If yes, study = Predictive Prob. = Bonferroni significance =
Fitted model: $X_i\beta = \hat{\beta}_0 + \hat{\beta}_{i1}$(High) + $\hat{\beta}_{i2}$ (All)
Prior Distribution for τ:
☒ Uniform on τ ☐ Uniform on τ^2 ☐ Half-norm ☐ Inverse Gamma ☐ DuM. s_0 ☐ DuM. $s_0/3$ ☐ DuM. $3s_0$
Random effect modeled from: Prior Distribution for **β**:
☐ Normal distribution ☒ t_4-distribution ☒ Normal (0,1000)
Model specification:
☒ Run 3 chains ☒ 50,000 iterations ☒ Discard first half ☒ DIC = -11.296
Do meta-analytic inferences change from the fitted hblm? ☐No ☒ Yes If yes, impact: Minimal

Parameters to monitor:

Node	Mean	Credibility Interval	Set parameter	MC error less than 5% s.d.	Rhat prox. to 1
β_0	.335	.064, .636	☒	☒	☒
β_1	-.439	-.907, .019	☒	☒	☒
β_2	-.521	-1.162, .038	☒	☒	☒
τ	.361	.216, .578	☒	☒	☒
$\beta_0 + \beta_1$	-1.04	-.464, .257	☒	☒	☒
$\beta_0 + \beta_2$	-.186	-.740, .321	☒	☒	☒
θ_{new} β_0	.327	-.440, 1.16	☒	☒	☒
θ_{new} β_1	-.441	-1.290, .360	☒	☒	☒
θ_{new} β_2	-.514	-1.486, .441	☒	☒	☒
θ_{new} $\beta_0 + \beta_1$	-.104	-1.009, .757	☒	☒	☒
θ_{new} $\beta_0 + \beta_2$	-.163	-1.098, .701	☒	☒	☒
p $\beta_0 < 0$	.008	0,0	☒	☒	☒
p $\beta_1 < 0$	.972	0,1	☒	☒	☒
p $\beta_2 < 0$	.956	0, 1	☒	☒	☒
p $\beta_0+\beta_1 < 0$	.692	0,1	☒	☒	☒
p $\beta_0+\beta_2 < 0$	.754	0,1	☒	☒	☒
p θ_{new} $\beta_0 < 0$	.197	0,1	☒	☒	☒
p θ_{new} $\beta_0 + \beta_1 < 0$	.610	0,1	☒	☒	☒
p θ_{new} $\beta_1 < 0$	.850	0,1	☒	☒	☒
p θ_{new} $\beta_2 < 0$	.862	0,1	☒	☒	☒
p θ_{new} $\beta_0 + \beta_2 < 0$	.648	0,1	☒	☒	☒
Deviance	-28.586	-38,231, -15.099	☒	☒	☒
θ_i^*	θ_8^* = -0.685 (min), θ_2^* = 1.038 (max)		☒	☒	☒

Items pertaining to parameter diagnostic checks:

	Kernel Density	History	BGR	Autocorrelation
☒ β_0				
☒ β_1				

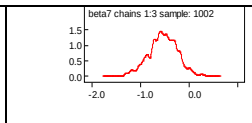
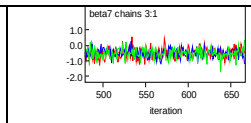
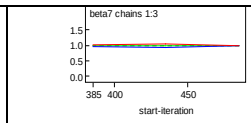
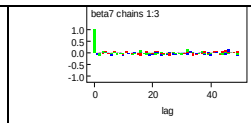
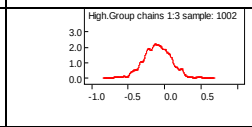
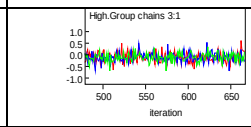
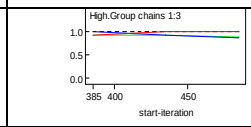
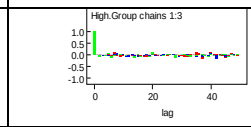
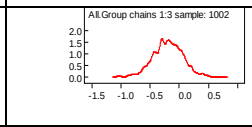
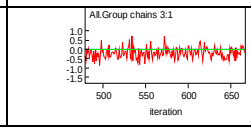
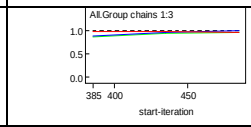
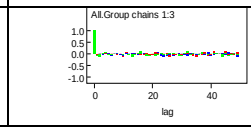
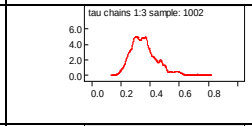
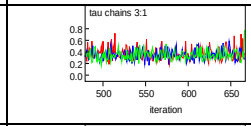
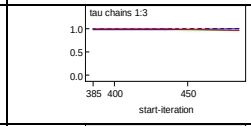
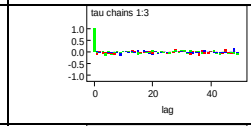
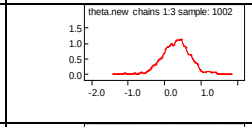
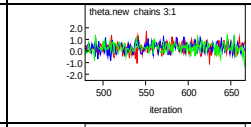
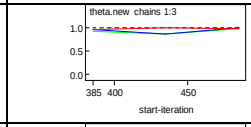
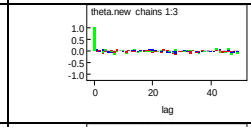
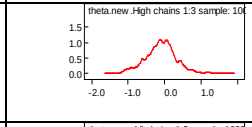
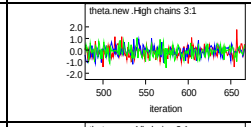
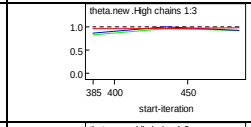
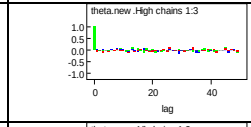
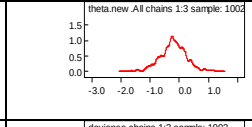
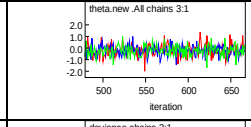
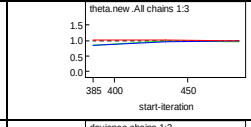
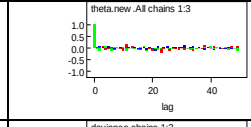
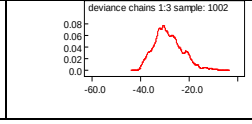
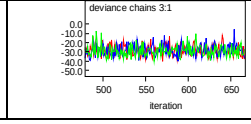
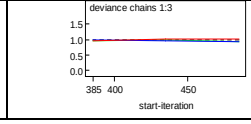
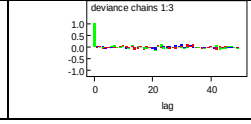
☒ β_2				
☒ $\beta_0+\beta_1$				
☒ $\beta_0+\beta_2$				
☒ τ				
☒θ_{new} for β_0				
☒θ_{new} for $\beta_0 + \beta_1$				
☒θ_{new} for $\beta_0 + \beta_2$				
☒ Deviance				

Table 20

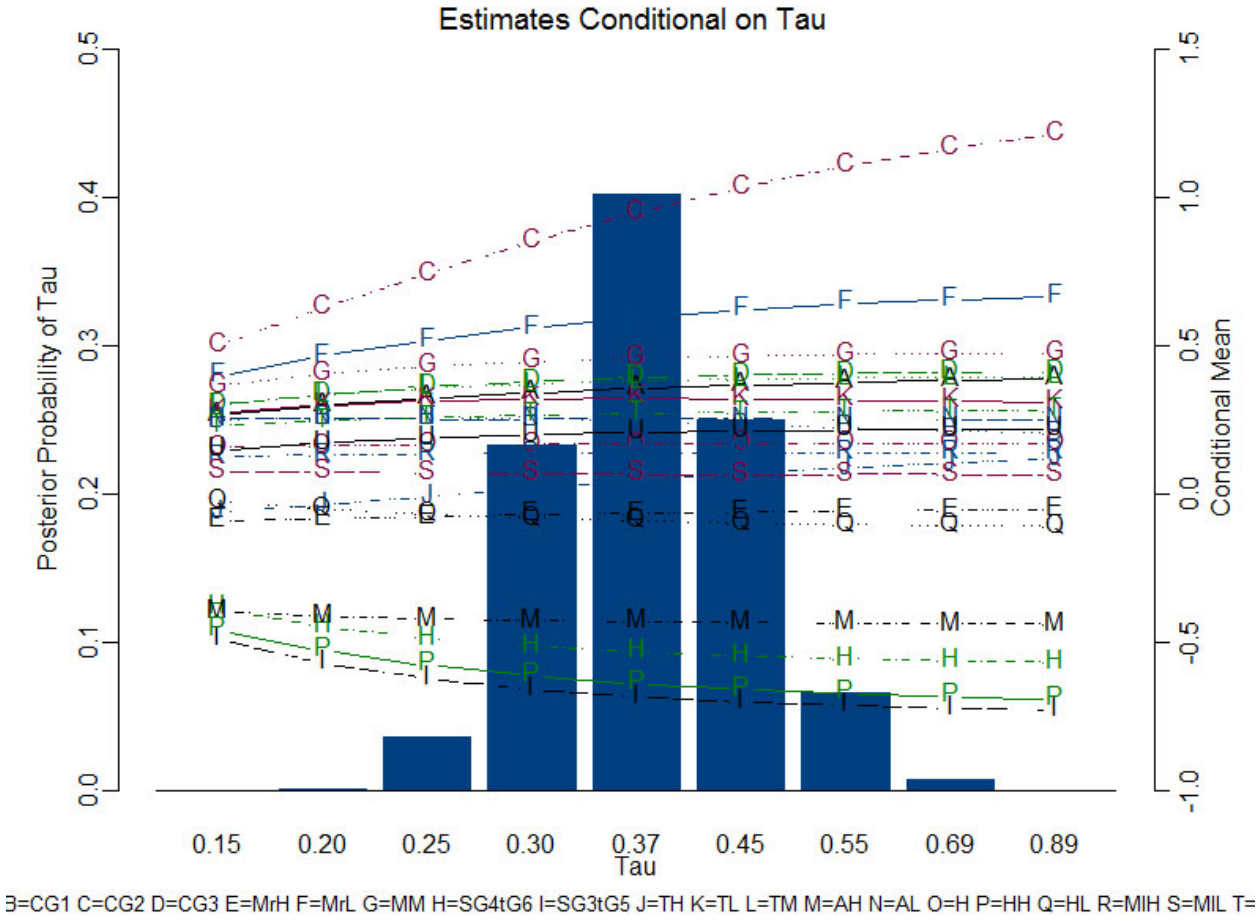
Detracking: Bayesian results (k=20)

Parameter	Posterior mean	*s.d.*	Prob < 0	95% Cr Int[a]
$\hat{\beta}_0$	0.335	0.150	.008	0.051, 0.051
$\hat{\beta}_1$	-0.439	0.242	.972	-0.941, 0.018
$\hat{\beta}_2$	-0.521	0.316	.956	-1.147, 0.097
$\hat{\beta}_0 + \hat{\beta}_1$	-0.104	0.192	.692	-0.513, 0.249
$\hat{\beta}_0 + \hat{\beta}_2$	-0.186	0.274	.754	-0.752, 0.349
$\theta_{new.\beta 0}$	0.327	0.409	.197	-0.489, 1.152
$\theta_{new.\ \beta 0+\beta 1}$	-0.104	0.422	.610	-0.902, 0.725
$\theta_{new.\ \beta 0+\beta 2}$	-0.163	0.470	.648	-1.059, 0.765
τ	0.361	0.092		0.225, 0.584

Note. Prior for τ ~ uniform (0,5). Prior for **β** ~ (0,1000). $\delta_i \sim t_4$. DIC = -11.3. [a]95% Credibility Interval.

Figure 20

Detracking: Trace Plot of Tau for Fitted Model (k=20)



Comparison of Bayesian and Classical Results for Detracking

Comparison of Classical and Bayesian Results (k=22)

Figures 21 - 23 display the widths of the Bayesian 95% credibility intervals compared to the classical 95% CI widths with REML estimation for the fitted models for the: (a) mean effect size, (b) the predicted effect in a new study, and (c) τ^2, the residual variance. With regard to the estimation of the model coefficients, the seven different prior distributions with the $\delta_i \sim$ normal distribution produce Bayesian parameter estimates for $\hat{\beta}_0$ and $\hat{\beta}_1$ and 95% credibility intervals that are very similar to those inferences made from the classical REML analyses for the *k*= 22 data set. However, with $\delta_i \sim t_4(0, \tau^2)$ the Bayesian models produce regression coefficients for $\hat{\beta}_0$ and $\hat{\beta}_1$ that range from 70 – 77% and 62% - 70% smaller as well as smaller standard deviations for $\hat{\beta}_0$ and $\hat{\beta}_1$, resulting in a reduction of the probability that the effect of $\hat{\beta}_1$ is greater than 0 (for the Bayesian $\delta_i \sim t_4$ models, the probability that $\hat{\beta}_1<0$ ranges from .035 - .051, whereas for the Bayesian $\delta_i \sim$ *Normal* models, the probability that $\hat{\beta}_1<0$ ranges from .004 - .009). Here, the discrepancy among the Bayesian fitted meta-analytic models is 'modest to severe' as the fitted model (using a criteria that the probability of $\hat{\beta}_1<0$ is less than .025) and parameter estimates change depending upon the modeling of the δ_i.

In terms of the ability to quantify the residual variance, when $\delta_i \sim$ normal, the median of the Bayesian parameter estimates for τ and the 95% credibility intervals are similar to those from the classical mixed model with REML estimation. As shown in Figure 23, when the $\delta_i \sim t_4$ distribution, all of the medians and 95% credibility intervals of the Bayesian posterior distributions for τ^2 are smaller in comparison to $\delta_i \sim$ normal and the classical REML mixed-effects model. For example, for the prior distributions where $\delta_i \sim t_4$, the values of τ are smaller - ranging from 66% (DuMouchel/3_{s0}) to 79% (half-normal).

In terms of the ability of the meta-analytic models to predict an effect in a new study, the classical approximate prediction interval width for an effect in a new study is similar to the Bayesian 95% credibility intervals, when the δ_i ~ normal. As shown in Figure 22, when the $\delta_i \sim t_4$, the Bayesian parameter estimates and 95% credibility intervals are smaller.

With regard to the ability of the models to estimate study-specific effects, the values of the Bayesian θ_i^* (as depicted in Figure 20 and 24) provide better estimates, because the Bayesian study-specific estimates consider the full uncertainty in all the auxiliary parameters as compared to the classical empirical Bayes study-specific estimates. Figure 24 depicts the study-specific shrinkage estimates for the classical mixed-effects REML model and the fully Bayesian model.

Comparison of Classical and Bayesian Results (k=20)

The data set shows more homogeneity when the outlying study point is excluded from the analysis and there is greater similarity of results among all of the models. Figures 25 – 27 display the widths of the Bayesian 95% credibility intervals compared to the classical 95% CI widths with REML estimation for the fitted models for the: (a) mean effect size, (b) the predicted effect in a new study, and (c) τ^2, the residual variance. With regard to all of these estimates, as depicted in Figures 20-22, the fourteen different Bayesian models produce Bayesian posterior parameter estimates and 95% credibility intervals that are very similar to those inferences made from the classical REML analyses for the *k* = 20 data set.

The meta-analytic inferences from the classical mixed-effects meta-analytic model with REML and the Bayesian hierarchical linear model differ from those inferences made in the original meta-analysis. For example, Rui (2009) used both a simple FE (*d* = 0.087, *p* < .001) and RE (*d*= 0.202, p <.01) model and concluded that, in general, there is a small but positive effect in favor of detracking students. In his meta-analysis Rui divided his data into subgroups that were grouped by the following study-level characteristics: (a) low-achieving

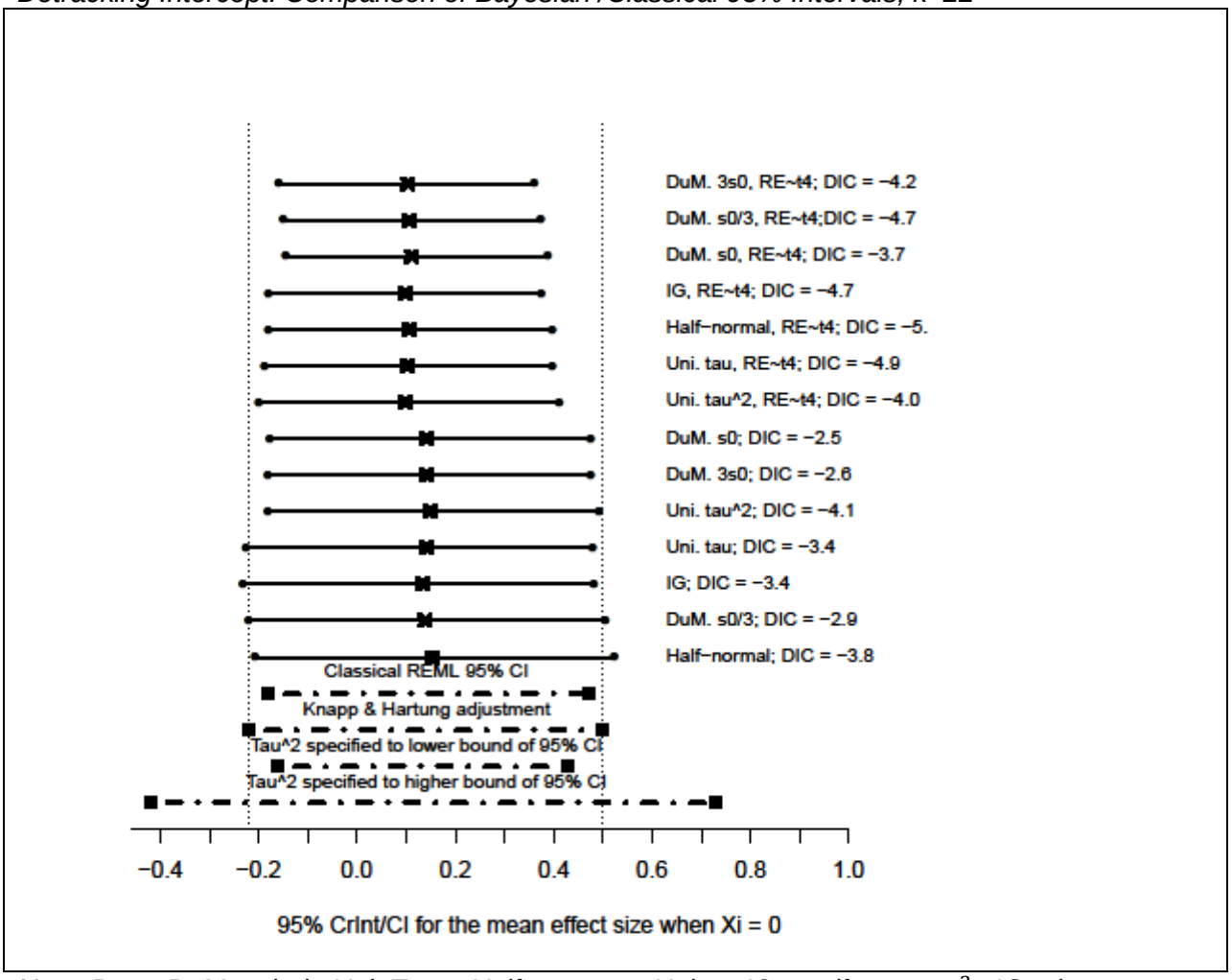
students, (b) high- and all-achieving students, (c) experimental studies, and (d) high-achieving students with the exclusion of the outlier study. Rui's separate subset mini meta-analyses showed that for low-achieving students both the simple FE (d = .113, p < .001) and RE (d = 0.283, p< .005) models showed positive effects of detracking on student achievement. Although Rui used both simple FE and RE models to describe his detracking findings, the inferences made from the RE models are likely to be more valid because the heterogeneity was high (τ^2 = .655, k = 22) and the Q-test for heterogeneity in the simple RE models was significant at p <.001, indicating the presence of significant variation between effect sizes. Although Rui performed separate subset analyses of study-level covariates (type of achieving and type of experimental design) in order to explain some of the heterogeneity, a more efficient, valid, and unified modeling approach for this meta-analytic data set is accomplished with the use of hierarchical linear modeling.

As shown in this reanalysis, the meta-analytic results differ from Rui's (2009) results depending upon: (a) the modeling approach, (b) the estimation method for τ^2, and (c) the inclusion/exclusion of the outlying study. The fitted classical mixed-effects (REML) and Bayesian hierarchical linear models show that there is a non-significant effect of detracking for low-, high-, and all-ability students and that this effect is modified by students classified as mid-achieving (k = 22). With a Bayesian interpretation there is an 80% probability that there is a beneficial effect of detracking that is greater than zero for students classified as low-, high-, or all-achieving and that, for students classified as mid-achieving, there is a 97.5% that this effect is greater than zero. However, these mixed-effects meta-analytic inferences change depending upon whether the outlying study (Kissoon-Singh, 1996) is included or excluded from the data set. When the Kissoon-Singh (1996) study is removed from the data set (k=20), the fitted model results in a significant effect of detracking only for the low- and mid-ability students and non-significant effects for those students classified as high- and all-achieving. Using a Bayesian interpretation (as shown in the Appendix), there is a 99.2% probability that there will be a

beneficial effect of detracking that is greater than zero for low-and mid-achieving students and that there is a 69.2% and 75.4% probability that there will be a deleterious effect of detracking for high- and all-ability students, respectively.

Figure 21

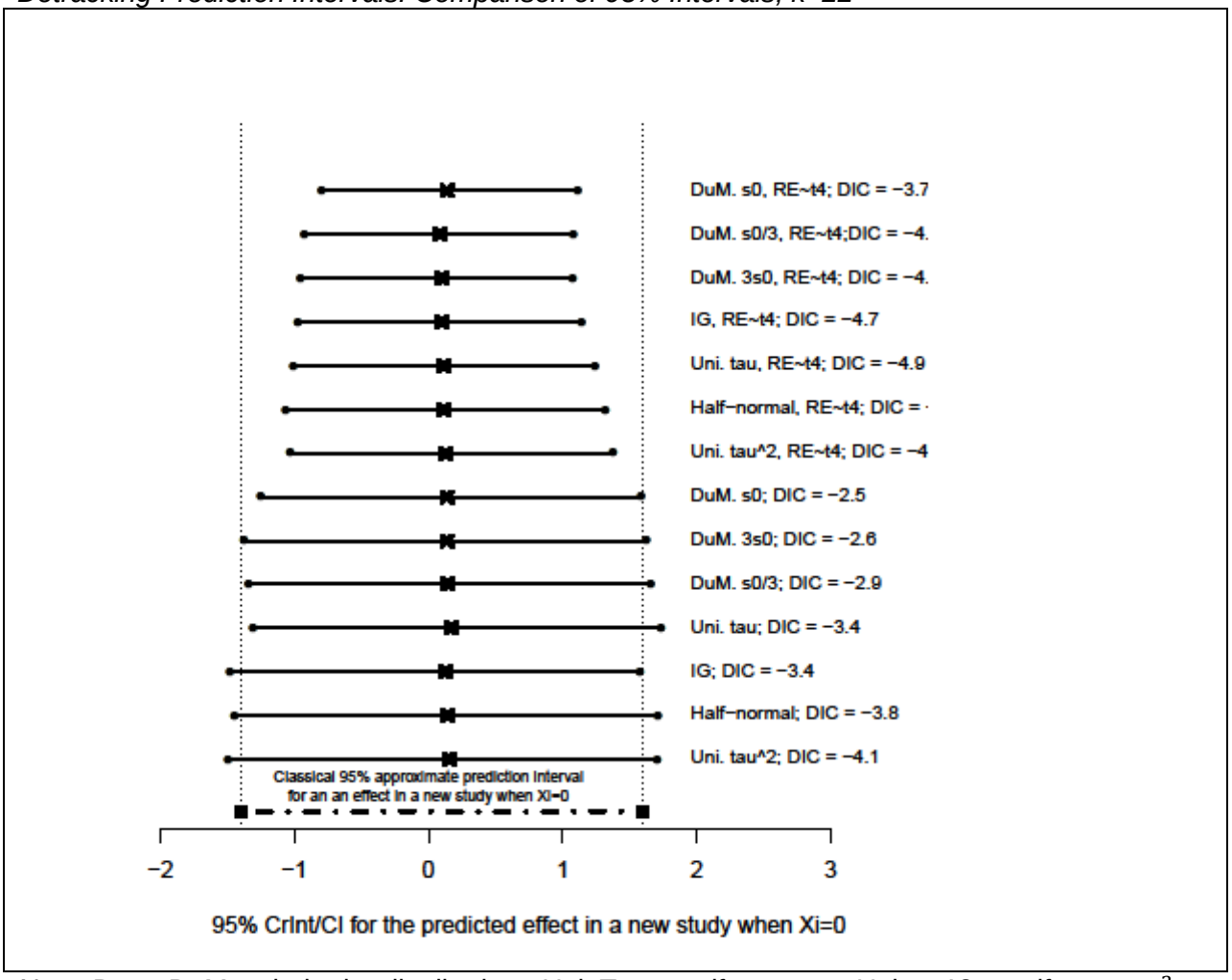
Detracking Intercept: Comparison of Bayesian /Classical 95% Intervals, k=22



Note. Du. = DuMouchel. Uni. Tau = Uniform on τ. Uni tau^2 = uniform on τ^2. IG = inverse gamma. Tau^2 = τ^2.

Figure 22

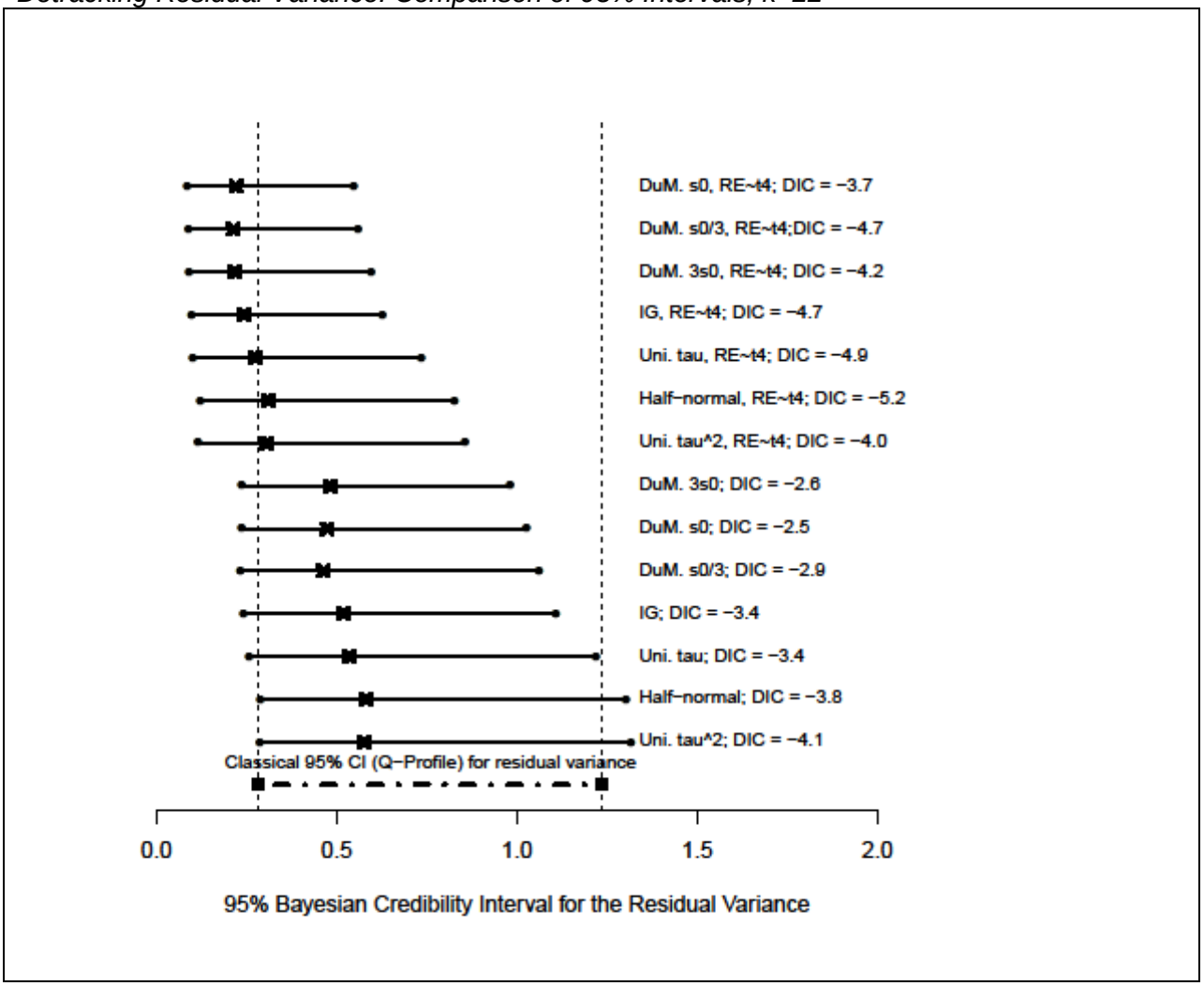
Detracking Prediction Intervals: Comparison of 95% Intervals, k=22



Note. Du. = DuMouchel prior distribution. Uni. Tau = uniform on τ. Uni tau^2 = uniform on τ^2. IG = inverse gamma.

Figure 23

Detracking Residual Variance: Comparison of 95% Intervals, k=22



Note. Du. = DuMouchel prior distribution. Uni. Tau = Uniform on τ. HN = half-normal. Uni tau^2 = uniform on τ^2. IG = inverse gamma.

Figure 24

Detracking Shrinkage Plot: Comparison of Study-Specific Estimates (k=22)

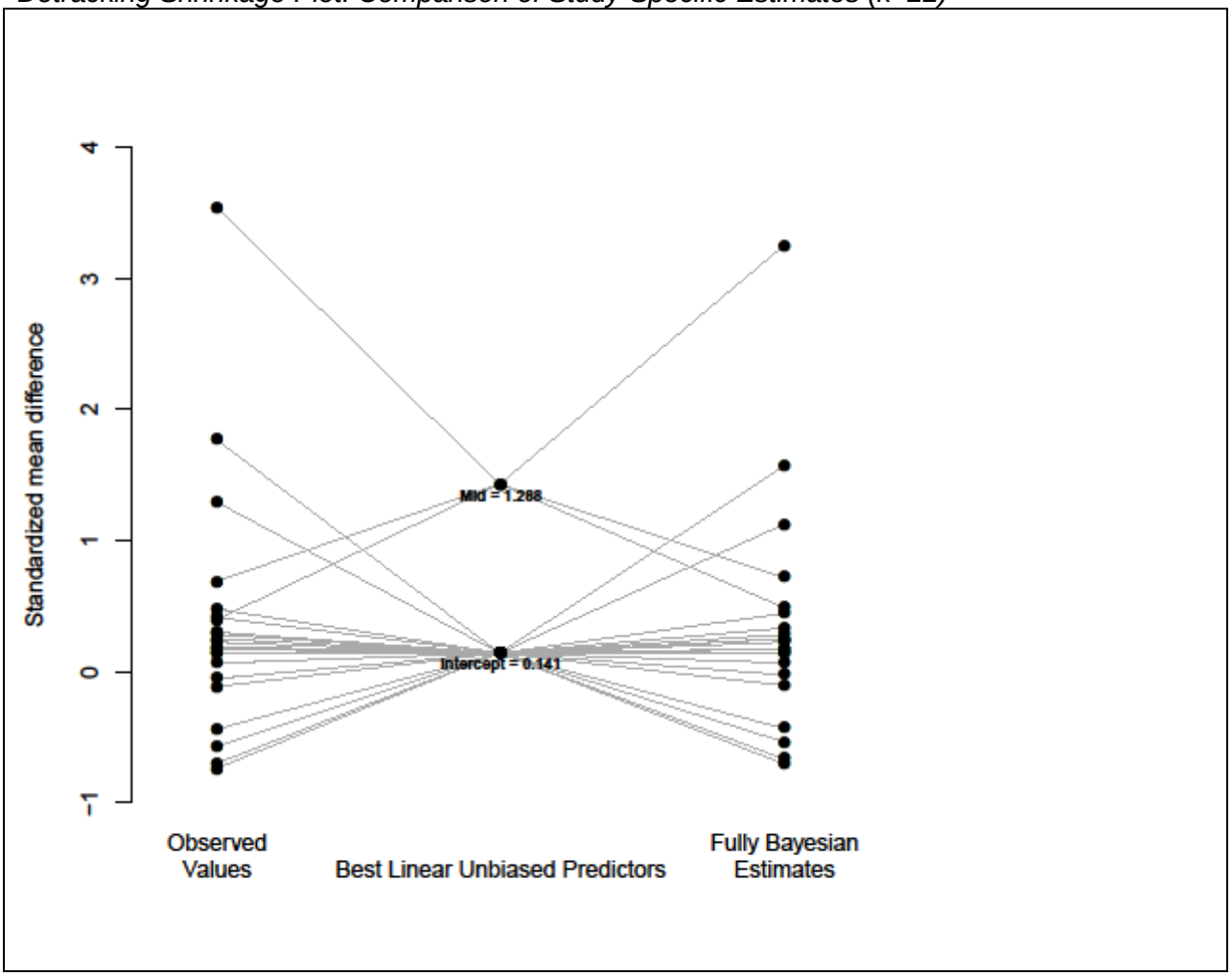
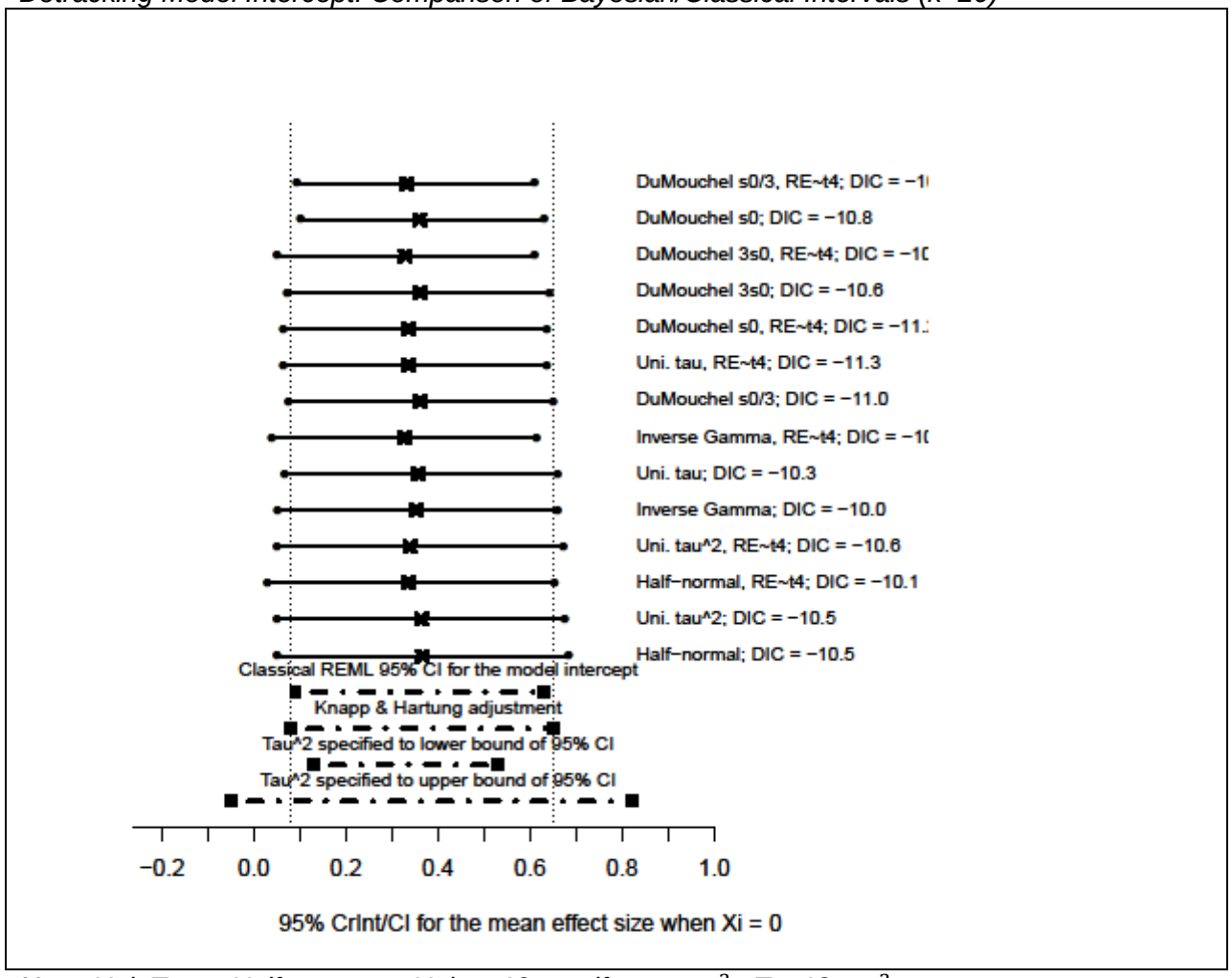


Figure 25

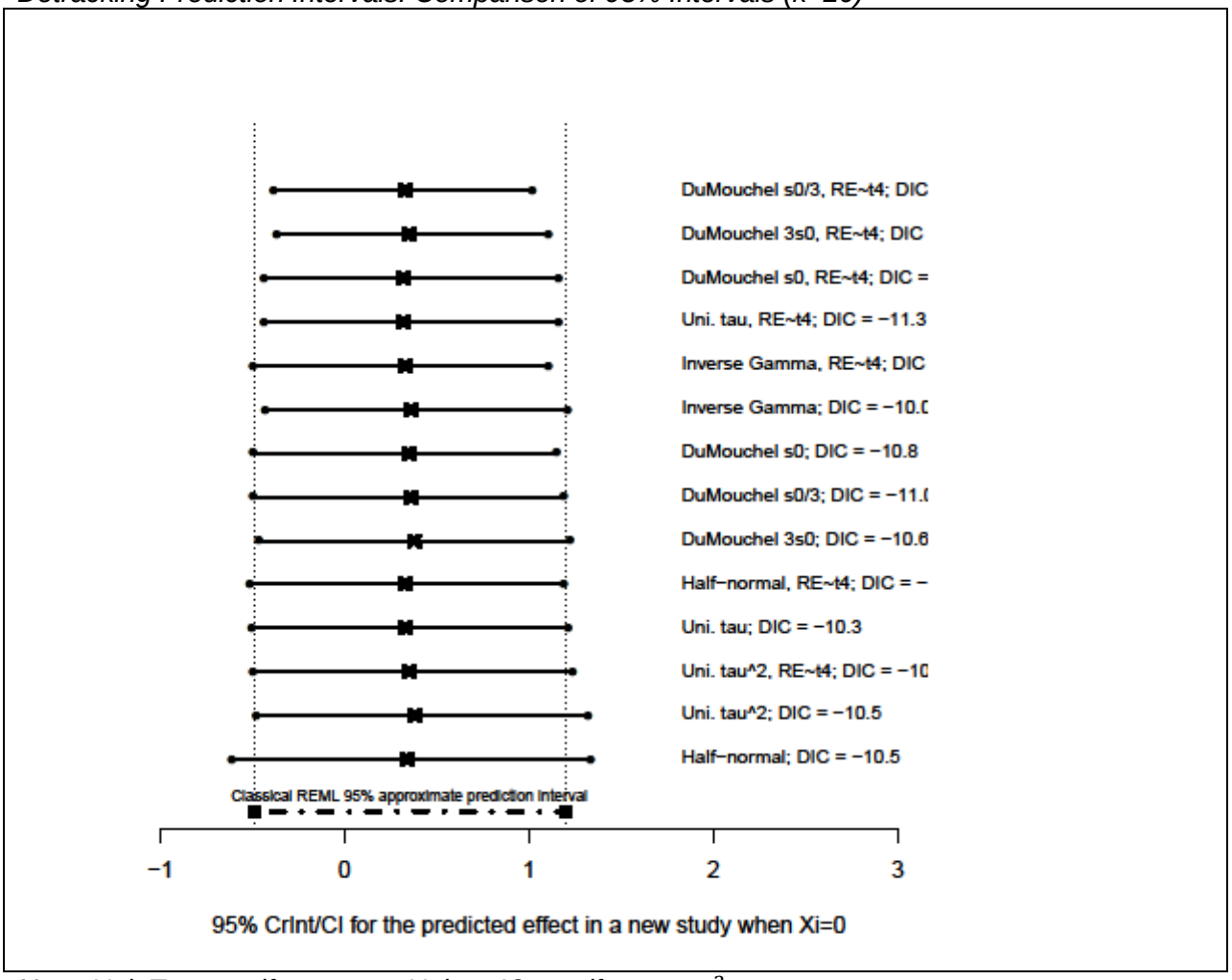
Detracking Model Intercept: Comparison of Bayesian/Classical Intervals (k=20)



Note. Uni. Tau = Uniform on τ. Uni tau^2 = uniform on τ^2. Tau^2 = τ^2.

Figure 26

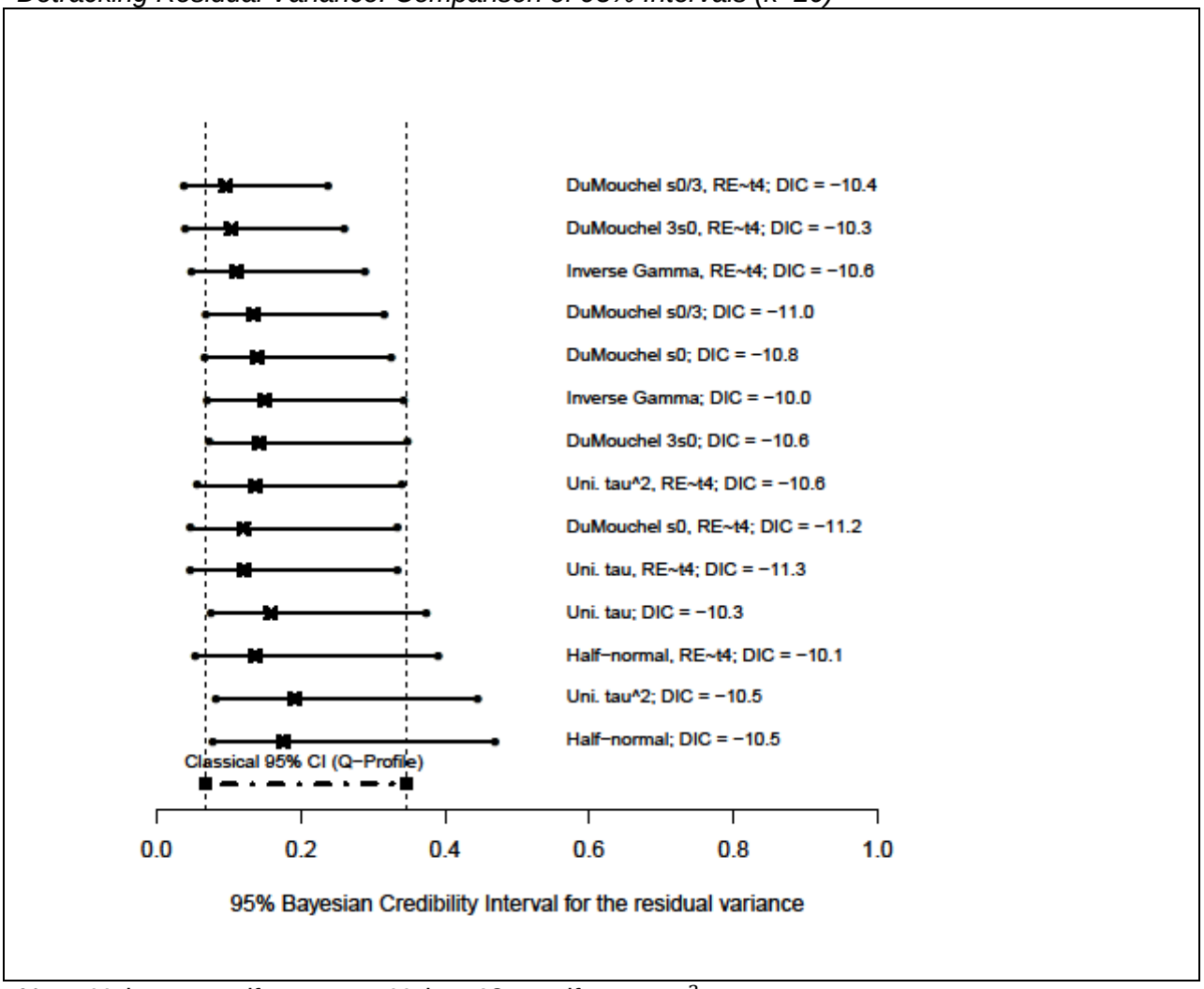
Detracking Prediction Intervals: Comparison of 95% Intervals (k=20)



Note. Uni. Tau = uniform on τ. Uni tau^2 = uniform on τ^2.

Figure 27

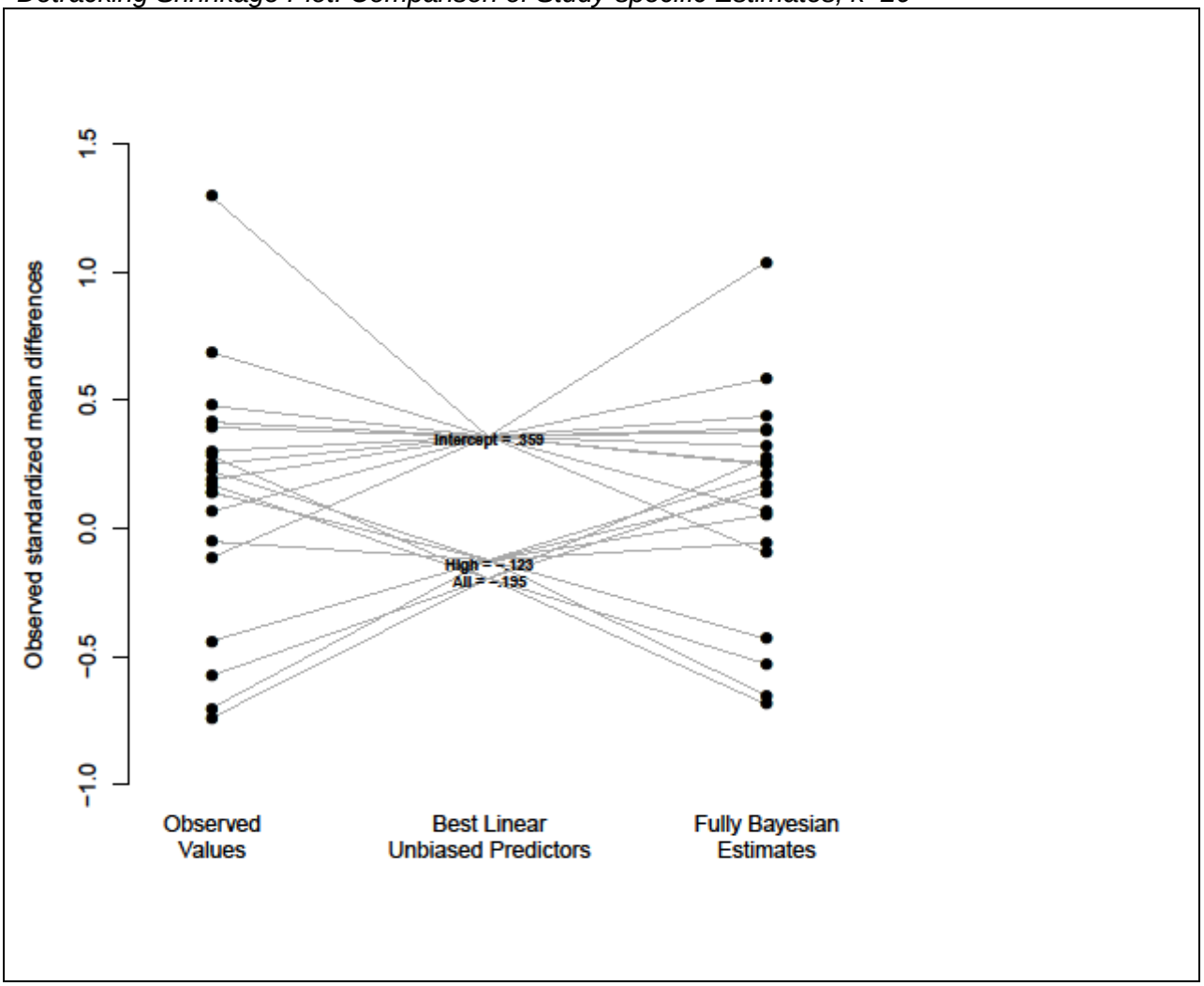
Detracking Residual Variance: Comparison of 95% Intervals (k=20)



Note. Uni. tau = uniform on τ. Uni tau^2 = uniform on τ^2.

Figure 28

Detracking Shrinkage Plot: Comparison of Study-specific Estimates, k=20



Data Set 3: Zinc for the Common Cold

Classical Meta-analysis Results for Zinc for the Common Cold

As shown in Table 21, the original simple RE meta-analysis results were successfully replicated with the *metafor* package in R 2.11.1 with D-L estimation, which was the method of estimation used for τ^2 in the original analysis. The *Q*-test for heterogeneity for the simple RE models was significant at $p < .001$, which is indicative of the presence of significant heterogeneity and implies that the estimates from a random/mixed-effects model would be more appropriate for this data set. In order to investigate potential sources of heterogeneity, both of the potential moderator variables were entered into a hierarchical linear model, and the moderators were removed sequentially using backward-stepwise model selection procedure with REML estimation. However, as shown in Table 21, neither moderator variable was significant, so the final model reduced to a simple RE model with the significance of the intercept (weighted mean effect size) dependent upon whether the Knapp and Hartung (2003) adjustment was applied to the standard error and test statistic for the intercept. The Knapp and Hartung adjustment produced a larger standard error for the intercept and more conservative hypothesis testing resulting in a 'modest to severe' impact among the models as the significance of the overall effect changed although the fitted meta-analytic model did not change. A forest plot for the fitted simple RE (REML) model is displayed in Figure 29. It is important to note that intercept and study-level covariates had very low statistical power and therefore are unlikely to be powerful enough to be able to detect a true underlying significant effect.

For the FE classical model the estimate for $\hat{\beta}_0$ is likely to be inaccurate and its standard error too small, given the large amount of heterogeneity that exists among the effect sizes. Furthermore, the D-L (method of moments) simple RE model produced estimates for $se_\mu$and τ that were slightly smaller than the estimates from the simple REML RE model, resulting in

more liberal hypothesis testing for the weighted mean effect size. For all of the classical (REML) RE models, the approximate prediction interval for an effect in a new study is very wide due to the large estimates for τ^2.

When considering the sensitivity of the mixed-effects REML classical meta-analytic models to different values of τ^2, it is important to note that the classical inferences change depending upon the value of τ^2. At a small value of τ^2(i.e., when τ^2 is set to the lower limit of its 95% CI), the study-level covariate, *adult*, is significant, (*p*=.027), but, when τ^2 is set to the upper limits of its 95% CI, it is no longer significant (*p*=.575). Comparisons of the model fit statistics indicated that the mixed-effects REML model had that best fit (i.e., the lowest AIC and BIC values), although traditional hypothesis testing resulted in non-significant *p* values for the mixed-effects study-level covariates.

In terms of detecting the presence of publication bias, the funnel plot for the simple RE model shows asymmetry, and the regression test for funnel plot asymmetry using se_i as a predictor was significant at $p < .001$, suggesting the presence of some asymmetry. The Orwin Fail-safe *N* value for a simple RE model was 6, indicating that 6 studies with an average of a null result would have to be added to this data set in order to reduce the simple RE unweighted average effect size of -1.060 in half, to a target effect size value of -0.530. However, in contrast to the above indicators of publication bias, the trim and fill method estimated that 0 studies were missing on either side of the funnel plot.

The leave-one-out case diagnostics and *Q*-*Q* normal plot were examined in order to investigate whether outlying studies that may have influenced the meta-analytic results. As displayed in Table 23, the case deletion diagnostics indicate that Study 6 appears to be a potential outlier. It has an *r*-student value of -2.27, a covariance value of 0.615, and, when it is excluded from the data set, the estimate for β_0 changes by -1.054 standard deviation units. However, examination of the *Q*-*Q* normal plot as a diagnostic indicator for outliers does not indicate the presence of an outlier, as all of the data points fall within their confidence envelopes

of the theoretical quantiles of the normal distribution plotted against the externally standardized residuals.

Table 21

Zinc: Classical Meta-analysis Results

Estimation Method	FE	D-L	REML	$REML_{KH}$	$REML_L$ τ^2 = .338	$REML_H$ τ^2= 6.70
Model	Simple	Simple	Simple	Simple	Simple	Simple
$\hat{B}_0$	-0.525	-0.972	-1.012	-1.012	-0.943	-1.052
s.e.	0.076	0.302	0.419	0.426	0.258	1.060
p value	<.001	.001	.016	.064	<.001	0.321
95% CI low	-0.674	-1.564	-1.833	-2.108	-1.448	-3.131
95%CI High	-0.377	-0.380	-0.192	0.083	-0.438	1.026
τ		0.696	0.994	0.994	0.581	0.581
95% CI low		0.582	0.582	0.582	0.582	0.582
95% CI high		2.588	2.588	2.588	2.588	2.588
τ^2		0.485	0.988	0.988	0.338	0.338
s.e.			0.665	0.665	0.251	0.251
95% CI low		0.338	0.338	0.338	0.338	0.338
95%CI high		6.697	6.697	6.697	6.697	6.697
Q	68.812,	68.812,	68.812,	68.812,	68.812,	68.812,
p value	< .001	< .001	< .001	< .001	< .001	< .001
PI $\hat{B}_0$ Lower			-4.007,	-4.105,	-2.708	-8.809,
PI $\hat{B}_0$ Upper			1.982	1.990	0.823	6.704
I^2		92.73%	96.30%	96.30%	89.90%	99.43%
Log-like	-30.736	-8.527	-8.212	-8.212	-9.634	-10.680
Deviance	61.473	17.054	16.424	16.424	19.268	21.360
AIC	63.473	21.054	20.424	20.424	23.268	25.360
BIC	63.265	20.638	19.642	19.642	22.487	24.579
Min, max $\theta_i^*(\tau)$			-2.437, -.017	-2.437, -.017	-2.126 -0.044	-2.627, -0.003

Note. FE= Fixed-effect model. D-L = DerSimonian & Laird estimation. REML = Restricted Maximum Likelihood Estimation. $REML_L$ = τ^2 was set equal to the lower bound of its 95% CI from the simple REML model. $REML_H$ = τ^2 was set equal to the upper bound of its 95% CI from the simple REML model. $REML_{KH}$ = Knapp and Hartung adjustment made to the standard errors of the regression coefficients. The permutation test *p* value was .062 for the REML simple model.

Table 21 (continued)

Zinc for the Common Cold: Classical Meta-analysis Results(continued)

Estimation Method	REML	$REML_L$ τ^2 = .338	$REML_H$ τ^2= 6.70
Model	Mixed	Mixed	Mixed
$\hat{B}_0$	0.000	-0.001	-0.000
s.e.	0.971	0.595	2.591
p value	.999	.999	.999
95% CI	(-1.903 ,1.903)	(-1.167 ,1.167)	(-5.078, 5.078)
$\hat{B}_1$ Syrup	-0.651	-0.653	-0.650
s.e.	1.191	0.733	3.174
p value	.585	0.373	0.838
95% CI	(-2.986, 1.684)	(-2.090, 0.784)	(-6.872, 5.571)
$\hat{B}_2$ Adult	-1.637	-1.567	-1.679
s.e.	1.134	0.708	2.997
p value	.149	.027	.575
95% CI	(-3.860, 0.586)	(-2.955 , -0.179)	(-7.553, 4.195)
τ	0.962	0.581	2.588
95% CI	(.493, 3.727)	(.493 3.727)	(.493 3.727)
τ^2	0.926	0.338	6.697
s.e.	0.821	0.337	5.535
95% CI	(0.243, 13.891)	(0.243, 13.891)	(0.243, 13.891)
Q	39.680	39.680	39.680
p value	< .001	< .001	< .001
PI $\hat{B}_0$	-3.795, 3.795	-2.310, 2.310	-10.16, 10.16
PI $\hat{B}_1$Lower	-4.903, 3.601	-3.251, 1.945	-12.021, 10.720
PI $\hat{B}_2$	-5.767, 2.493	-4.111, 0.977	-12.673, 9.314
Q_{Mods}	2.495	5.617	0.386
p value	.287	.060	.824
Log-like Deviance	-5.162 10.325	-5.956 11.912	-6.746 13.492
AIC	18.325 14.719	19.912 16.306	21.492 17.886
BIC			
Min, max $\theta_i^*(\tau)$	-2.514, .000	-2.317 .000	-2.637, .000

Note. FE= Fixed-effect model. D-L = DerSimonian & Laird estimation. REML = Restricted Maximum Likelihood Estimation. $REML_L$ = τ^2 was set equal to the lower bound of its 95% CI from the simple REML model. $REML_H$ = τ^2 was set equal to the upper bound of its 95% CI from the simple REML model. $REML_{KH}$ = Knapp and Hartung adjustment made to the standard errors of the regression coefficients.

Table 22

Zinc: Power for Two-sided Test of the Regression Coefficients

	Power for Individual Regression Coefficients Backward Selection Regression Models		
Coefficient	Full Model	Model 1	Fitted Model
Intercept	.07	.11	.13
Syrup	.06		
Adult	.06	.07	

Note. α = .05 with a clinically meaningful effect set equal to 0.33 standard deviations.

Figure 29

Zinc for the Common Cold Meta-analysis: Forest Plot (REML)

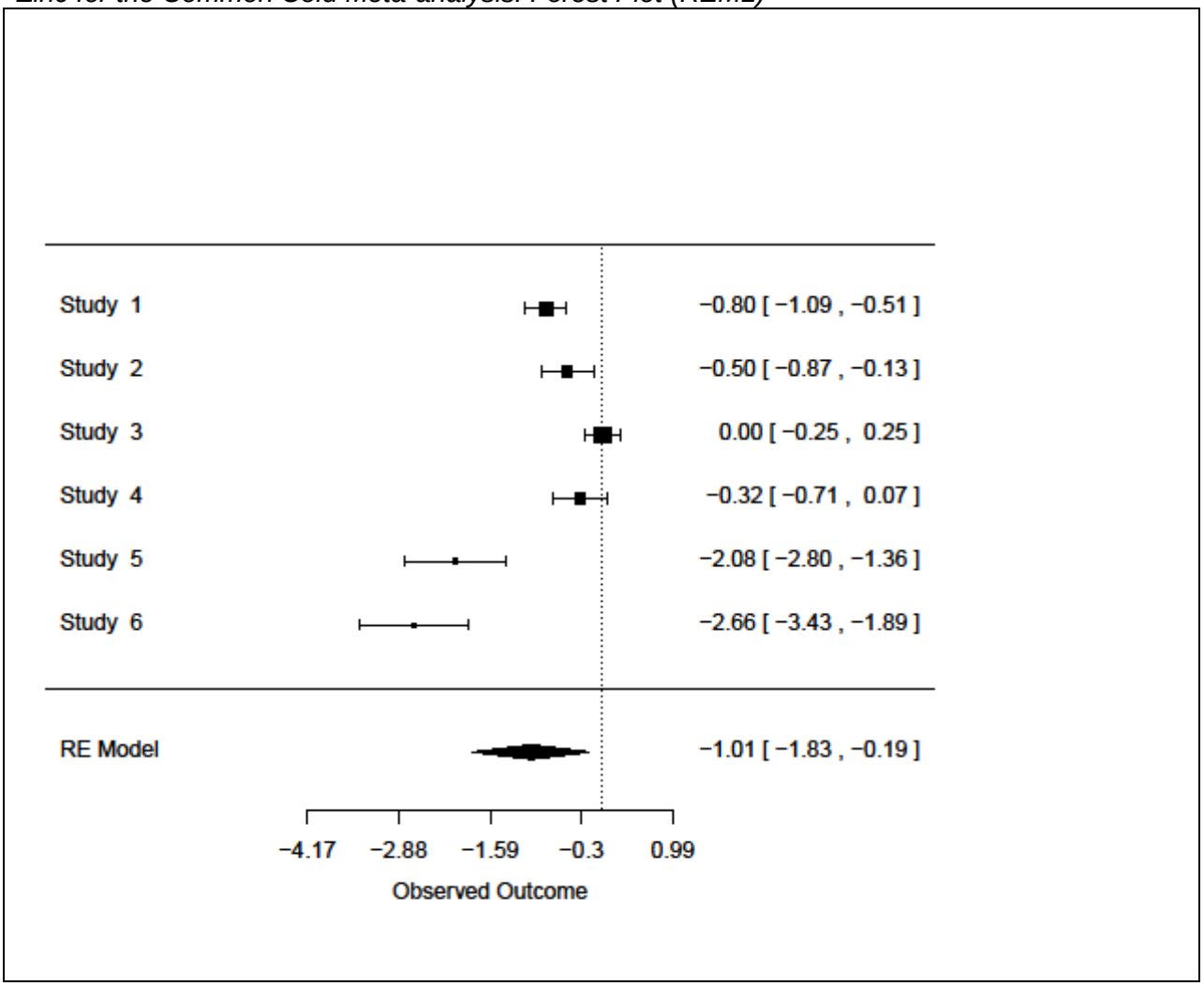


Figure 30

Zinc for the Common Cold Meta-analysis: Funnel Plot

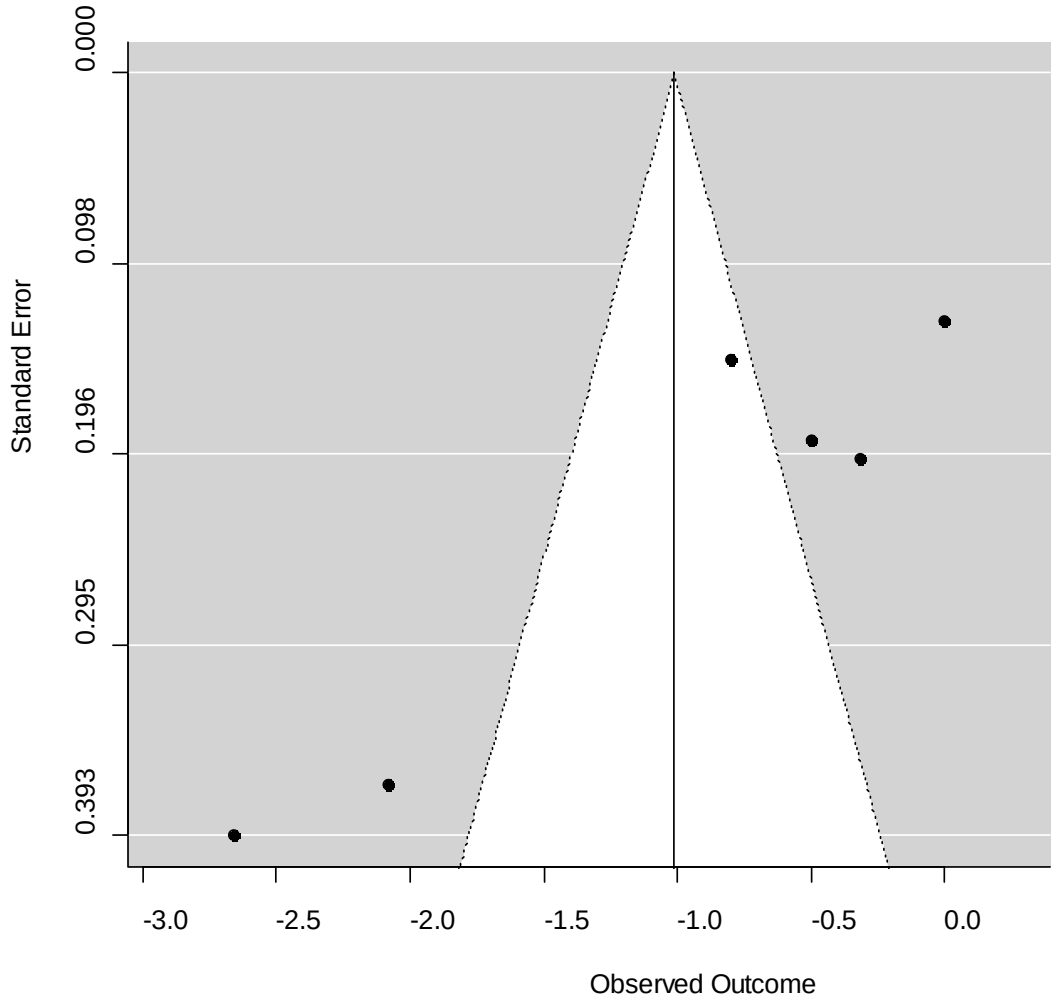


Table 23

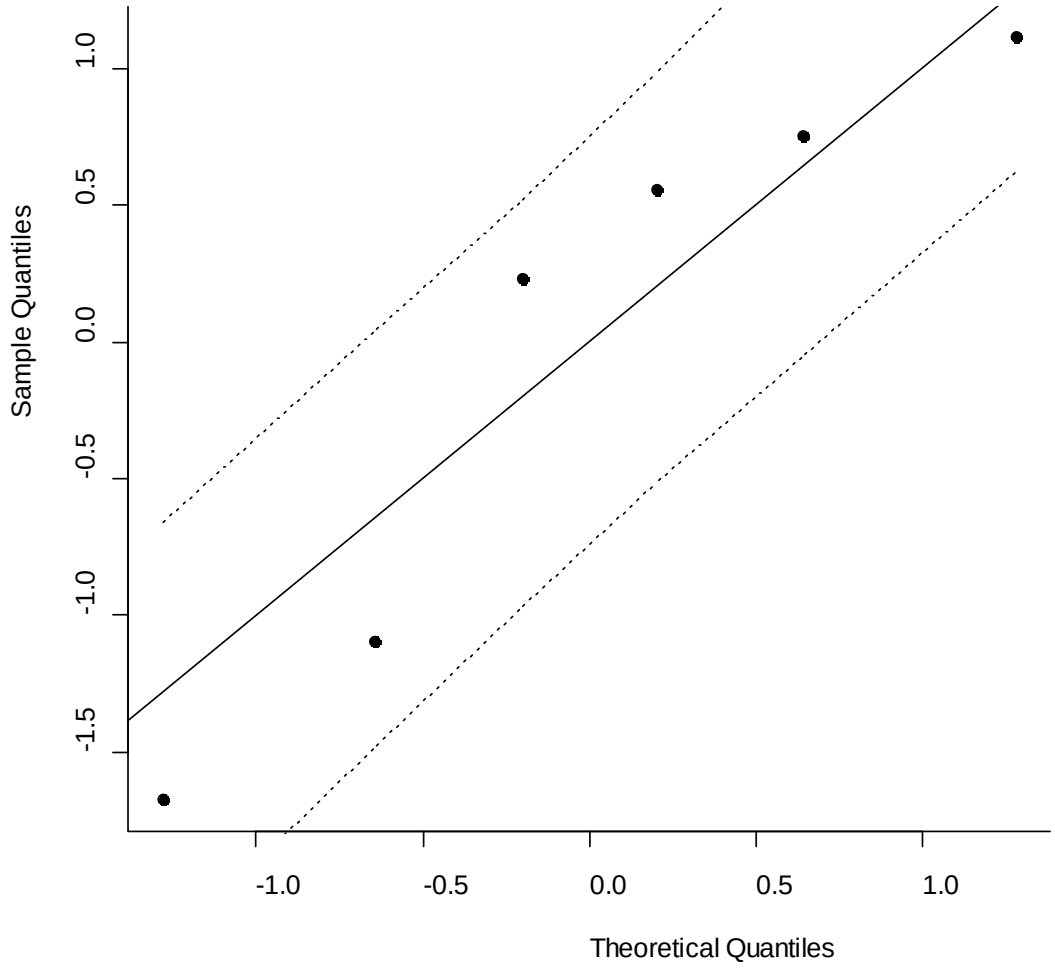
Zinc: Leave-one-out Case Diagnostics

Study	rstudent[a]	cov.r[b]	τ^2.del[c]	QE.del[d]	B_0[e]
1	0.215	1.531	1.269	64.136	0.119
2	0.514	1.448	1.199	68.790	0.25
3	1.136	1.160	0.943	42.900	0.518
4	0.708	1.370	1.131	67.567	0.333
5	-1.140	1.096	0.910	50.065	-0.498
6	-2.265	0.615	0.493	38.153	-1.054

Note. [a]denotes the r student (externally standardized residual) value with the i^{th} study excluded. [b]denotes the covariance ratio. [c]denotes the estimate value of tau-squared based on the data set with the i^{th} study excluded. [d]denotes the *Q* statistic for the test of heterogeneity based on the data set with the i^{th} study removed. [e]denotes by how many standard deviation units by which the intercept changes when the i^{th} study is excluded from the data set (Viechtbauer, 2009).

Figure 31

Zinc Meta-analysis: Normal Q-Q Plot



Bayesian Meta-analysis Results for Zinc for the Cold

A Bayesian hierarchical linear model was fitted in WinBUGS with a backward stepwise model selection procedure and a uniform prior distribution on τ (0,5) with $\delta_i \sim Normal$ and $\delta_i \sim t_4$. A sample of the Bayesian quality assurance checklist that was used to fit the model, monitor the parameters, and check the items related to model convergence is displayed in Figure 32 (Bayesian Quality Assurance Checklist for the Zinc for the Common Cold Meta-analysis). As shown in Table 24, none of the study-level covariates were significant, and the model reduced to a simple (no covariates) model. For the Zinc for the Common Cold data set, there is a 96.1% probability that there is an effect that is less than zero (i.e., a beneficial effect in favor of reducing the duration of cold symptoms) for those participants in RCT who took zinc. In a new study there is an 80% chance that there will be a beneficial predicted effect of zinc in reducing cold duration. The estimate and 95% Cr Int for τ is large (τ = 1.372, 95% Cr Int .621, 3.006), reflecting the large amount of uncertainty in the heterogeneity. The trace plot (Figure 34) displays how the meta-analytic inferences are dependent upon the posterior predictive values of τ and its uncertainty. The Bayesian results for the 14 different meta-analytic models (e.g., the seven different prior distributions on the heterogeneity variance for which δ_i ~ Normal or $\delta_i \sim t_4$ are displayed in the Appendix).

Bayesian cross-validation was conducted in S-Plus 2000 with the use of the *hblm* function on the fitted model using a DuMouchel prior. As depicted in Table 25 and Figure 33, Bayesian cross-validation revealed that Study 6 was a potential outlying data point, as its predictive probability was .026; however, as depicted in the Q-Q normal plot, the residual is not considered to be too extreme, and the Bonferroni significance for the residual of Study 6 is not significant (*p* = .308).

Table 24

Zinc for the Common Cold Meta-analysis: Bayesian Results

Model 1. Full Model, DIC = 4.627				
Parameter	Mean	s.d.	Prob < 0	95% Cr Int
$\hat{\beta}_0$	-0.133	1.856	.520	-3.755, 3.683
$\hat{\beta}_1$	-0.422	2.339	.632	-5.248, 4.364
$\hat{\beta}_2$	-1.577	2.179	.819	-6.155, 3.351
$\hat{\beta}_0 + \hat{\beta}_1$	-0.555	1.340	.724	-3.358, 2.548
$\hat{\beta}_0 + \hat{\beta}_2$	-1.711	1.142	.946	-1.656, 0.854
$\theta_{new.\beta 0}$	-0.166	2.565	.518	-5.658, 4.909
$\theta_{new.\ \beta 0+\beta 1}$	-0.580	2.279	.643	-5.386, 4.377
$\theta_{new.\ \beta 0+\beta 2}$	-1.757	2.203	.847	-6.278, 2.711
τ	1.671	0.955		0.591, 4.272
Model 2. Fitted Model, DIC = 4.640				
Parameter	Mean	s.d.	Prob < 0	95% Cr Int
$\hat{\beta}_0$	-1.084	0.652	.961	-2.448, .251
$\theta_{new.\beta 0}$	-1.121	1.589	.800	-4.549, 2.158
τ	1.372	.63		.621, 3.006

Note. Prior for τ ~ uniform (0,5). Prior for $\boldsymbol{\beta}$ ~ (0,1000). δ_i ~ Normal.

Figure 32

Zinc: Bayesian Quality Assurance Meta-analysis Checklist

☒ Cross-validation Outlier identified? ☒ No ☐ Yes

Fitted model: $X_i\beta = \hat{\beta}_0$

Prior Distribution for τ:

☒ Uni. on τ (0,5) ☐ Uni. on τ^2 ☐ Half-norm ☐ Inverse Gamma ☐ DuM. s0 ☐ DuM. s0/3 ☐ DuM. 3s0

Form for the random-effects distribution: ☒ Normal distribution ☐ t_4-distribution

Prior Distribution for **β**: ☒ Normal (0,1000)

Model specification:

☒ Run 3 chains ☒ 50,000 iterations ☒ Discard half ☒ DIC = 4.6

Do meta-analytic inferences change from the fitted hblm: ☐No ☒Yes. If yes, impact: Minimal

Node	Mean	95% Cr Int	Set parameter	MC error < 5% of s.d.	Rhat proximity to 1
β_0	-1.084	-2.448, .251	☒	☒	☒
τ	1.372	.621, 3.006	☒	☒	☒
$\theta_{new\beta0}$	-1.121	-4.549, 2.158	☒	☒	☒
$p\ \beta_0 < 0$	.961		☒	☒	☒
$p\ \theta_{new\beta0} < 0$	.800		☒	☒	☒
θ_i^*	θ_3^* min = -.016,	θ_6^* max = -2.49	☒	☒	☒

Items pertaining to relevant parameter diagnostic checks:

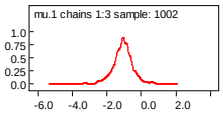
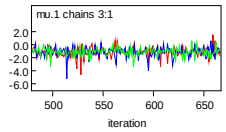
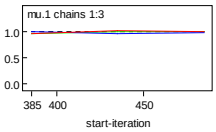
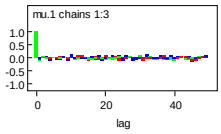
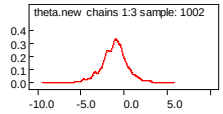
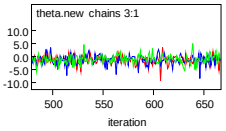
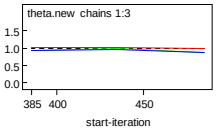
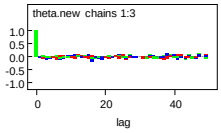
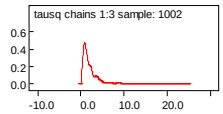
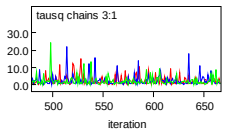
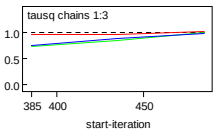
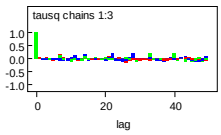
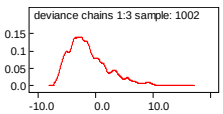
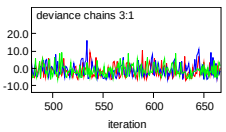
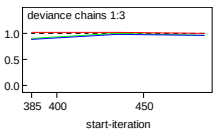
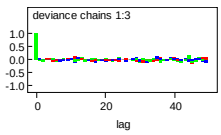
	Kernel Density	History	BGR	Autocorrelation
☒ β_0				
☒ θ_{new}				
☒ τ^2				
☒ Dev.				

Table 25

Zinc: Bayesian Cross-validation

Study	Y	SE	Prior Mean	Prior SD	Pred. Prob	τ^2	Post. Mean	Post.SD	Prob > 0
Kurugol 06	-0.80	0.148	-1.058	1.462	.583	1.769	-0.805	0.146	.000
Kurugol 07	-0.50	0.189	-1.117	1.419	.700	1.666	-0.521	0.186	.003
Macknin98	0.00	0.128	-1.213	1.266	.868	1.322	-0.020	0.128	.436
Petrus 98	-0.32	0.199	-1.150	1.379	.764	1.572	-0.352	0.196	.036
Prasad 00	-2.08	0.367	-0.800	1.226	.127	1.243	-1.925	0.357	.000
Prasad 08	-2.66	0.393	-0.679	0.909	.026	0.679	-2.395	0.396	.000

Figure 33

Zinc for the Common Cold: Bayesian Q-Q Plot

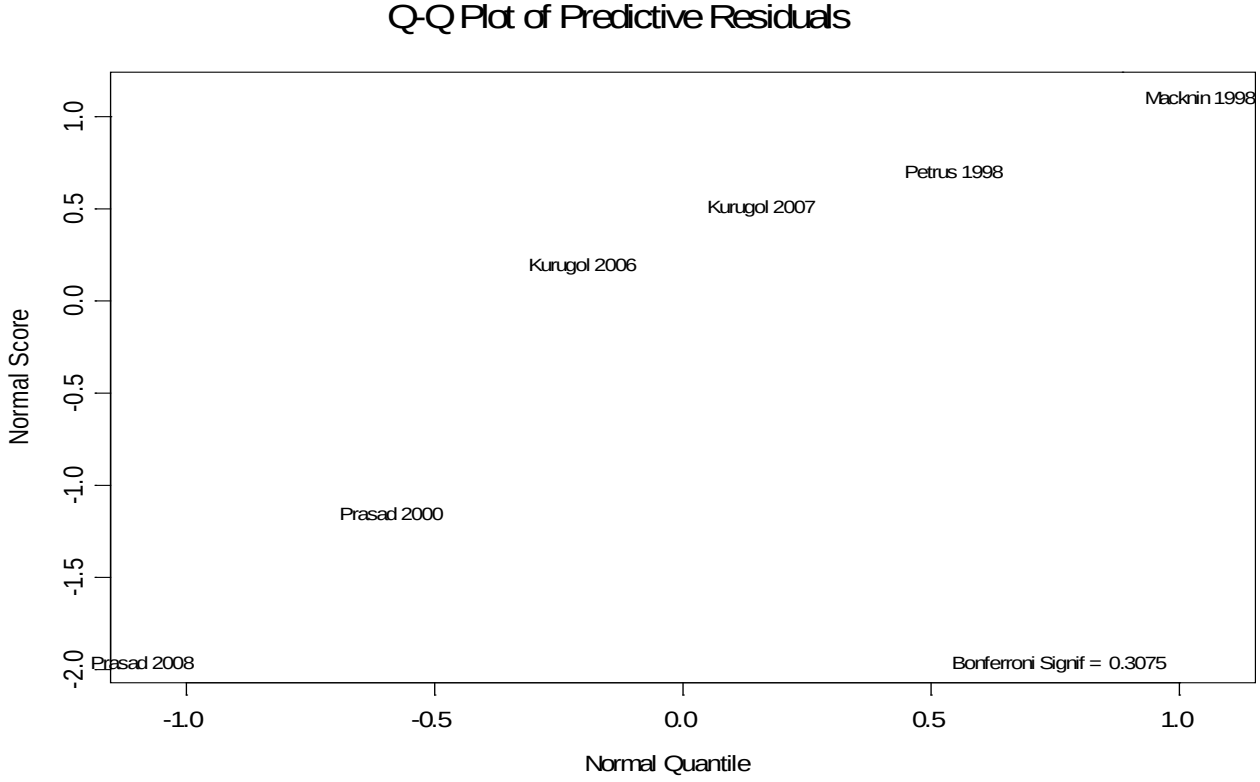
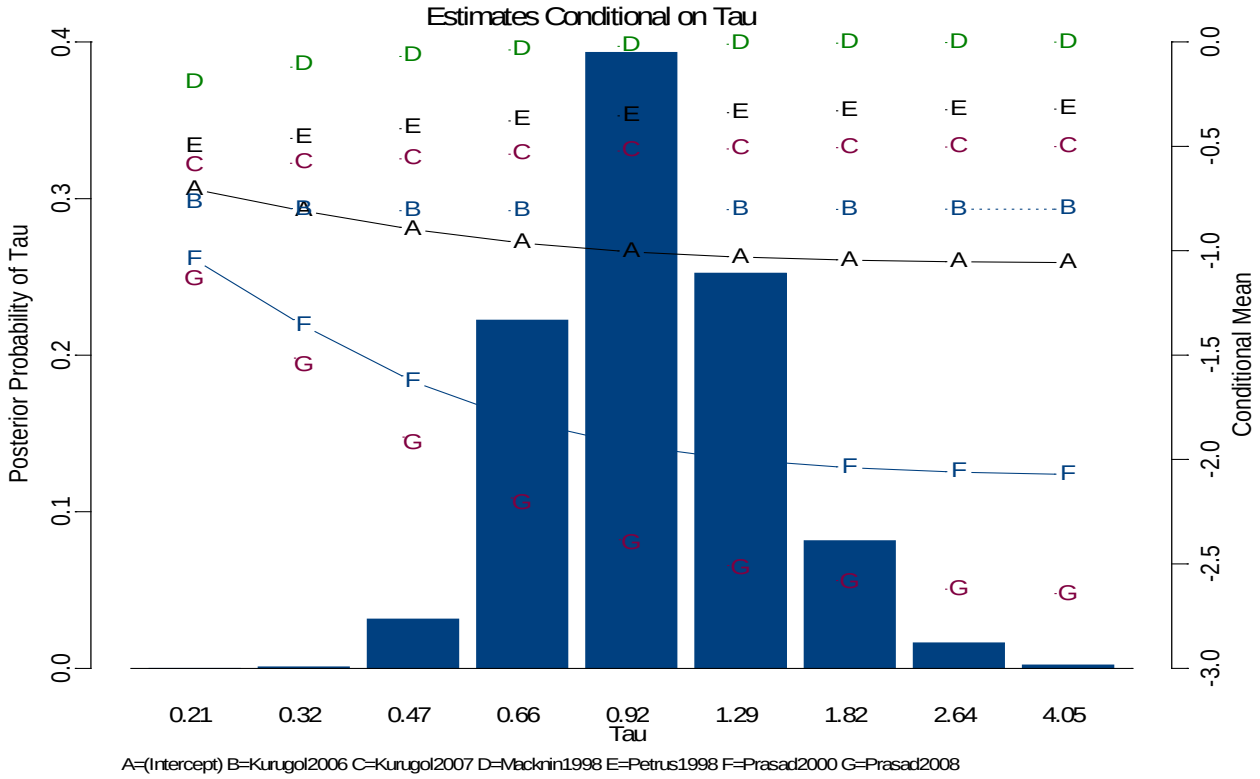


Figure 34

Zinc for the Common Cold: Bayesian Trace Plot of Tau



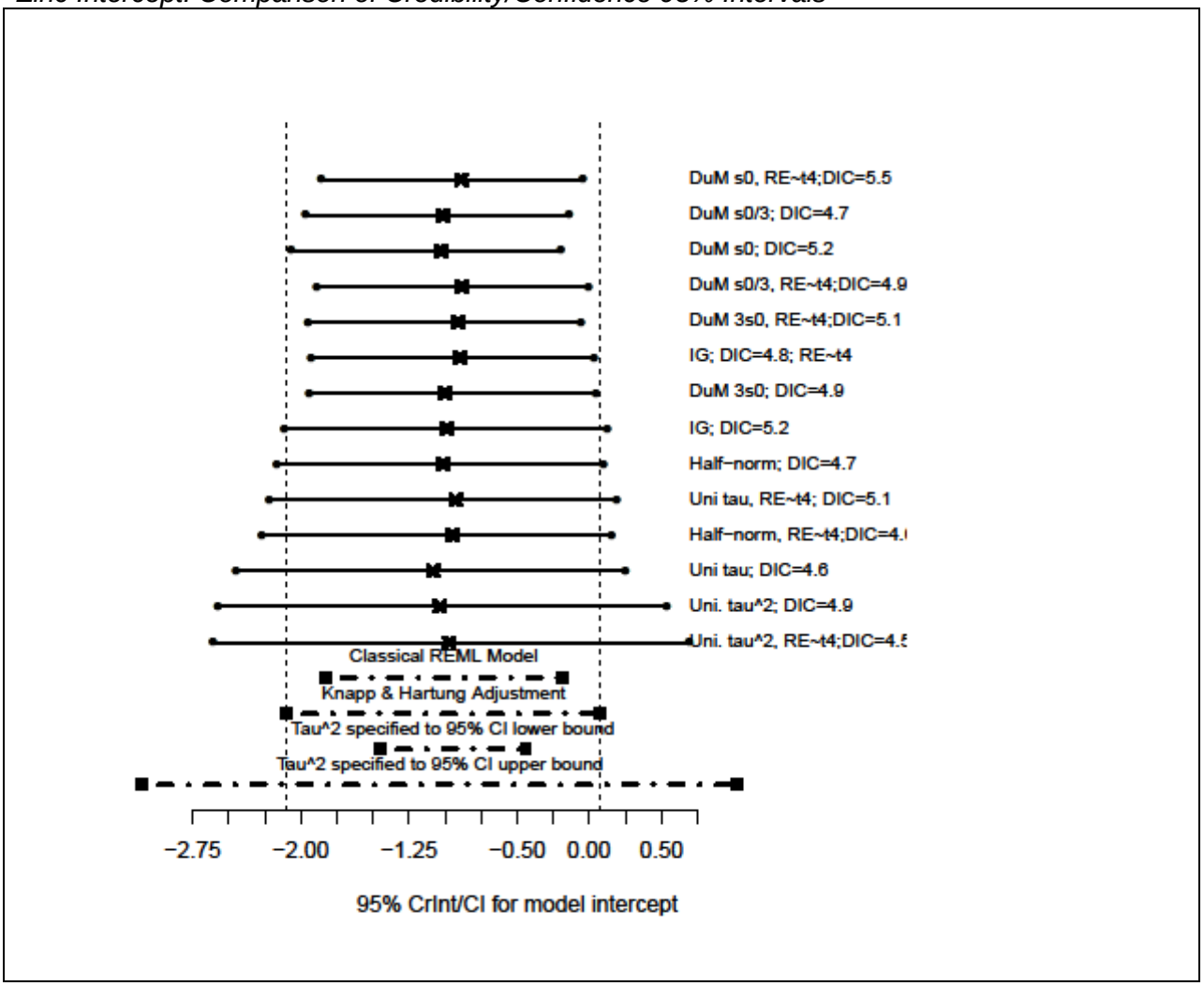
Comparison of Classical and Bayesian Results

Figures 35-37 display the widths of the Bayesian 95% credibility intervals compared to the classical 95% CI widths with REML estimation for the fitted models for the: (a) mean effect size, (b) the predicted effect in a new study, and (c) τ^2, the residual variance. With regard to the estimation of the predicted effect in a new study and residual variance, all of the Bayesian models produce estimates that are very similar to those inferences made from the classical REML method with use of the *Q*-profile method for quantifying the uncertainty in the heterogeneity. The confidence intervals are all extremely wide for θ_{new} and τ^2. The only Bayesian model that does not produce estimates similar to those from the classical REML method is the Bayesian model with a prior distribution that favors extremely large values of τ^2 (i.e., the uniform on τ^2 prior distribution). Somewhat surprisingly, the Bayesian DIC measures of model fit are the lowest for the uniform on τ^2prior distribution. With regard to the estimation of the weighted mean effect size, the classical REML model produces a 95% CI that is too short. The confidence interval for the weighted mean effect size with the classical REML model with the Knapp and Hartung (2003) adjustment has greater similarity to the Bayesian 95% Cr Int for the model intercept.

With regard to the ability of the models to estimate study-specific effects, the values of the Bayesian θ_i^* (as depicted in Figure 34 and 37) provide better estimates because the Bayesian study-specific estimates consider the full uncertainty in all the auxiliary parameters as compared to the classical empirical Bayes study-specific estimates. However, as displayed in Figure 37, the fully Bayesian estimates experience minimal shrinkage due to the considerable heterogeneity.

Figure 35

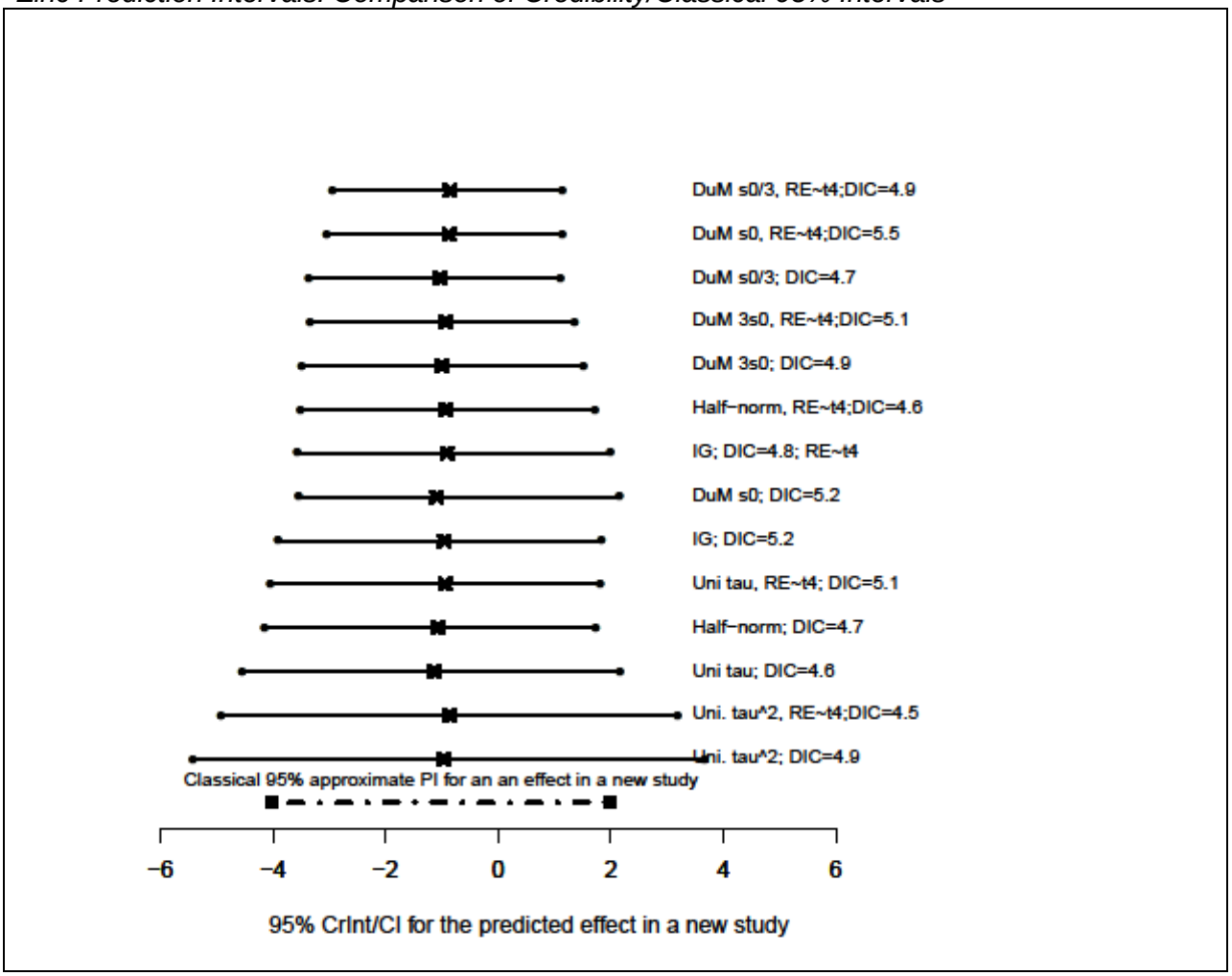
Zinc Intercept: Comparison of Credibility/Confidence 95% Intervals



Note. Du. = DuMouchel prior distribution. Uni. Tau = Uniform on tau. Half-norm = half-normal. Uni tau^2 = uniform on τ^2. IG = inverse gamma. Tau^2 = τ^2.

Figure 36

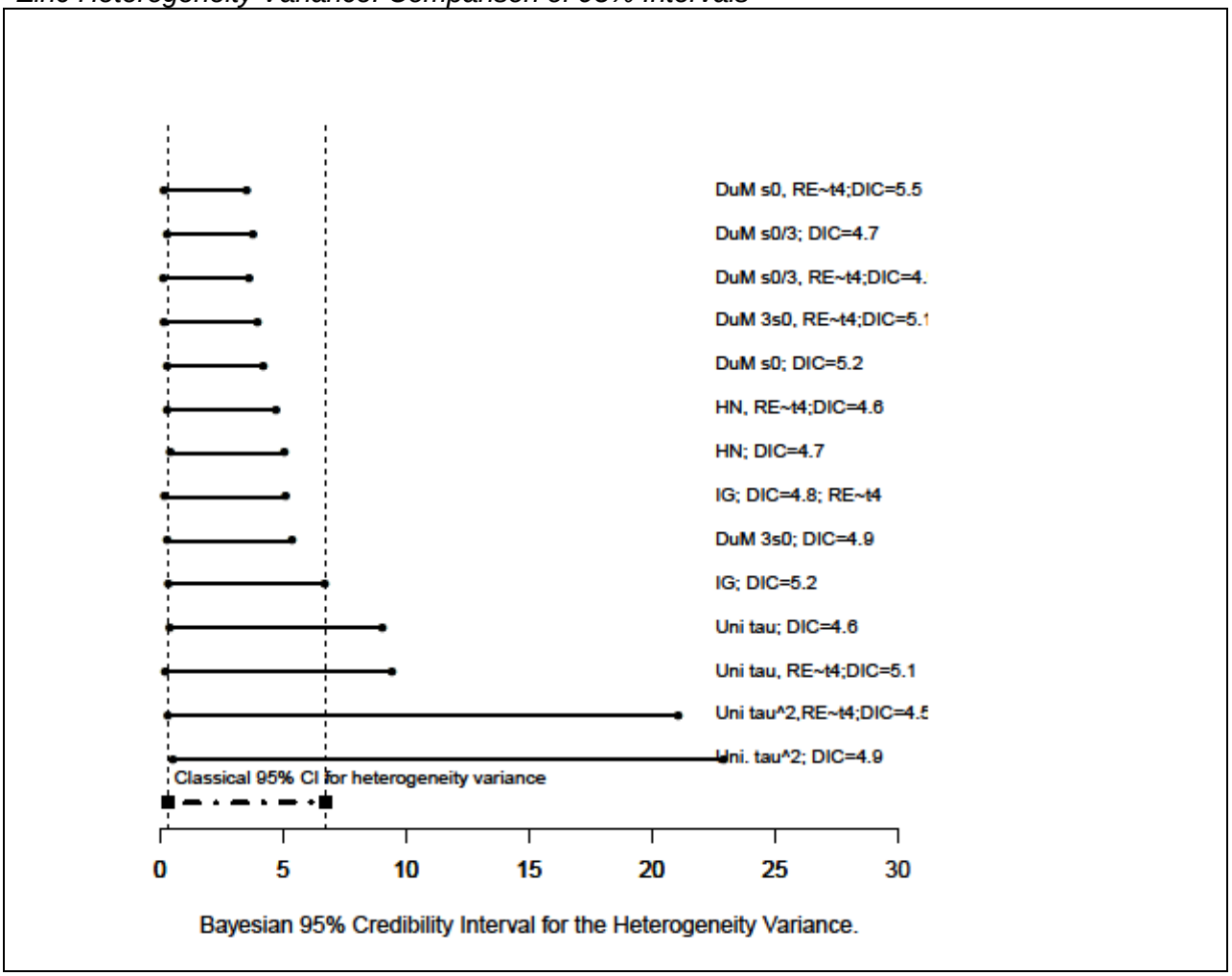
Zinc Prediction Intervals: Comparison of Credibility/Classical 95% Intervals



Note. Du. = DuMouchel. Uni. Tau = Uniform on τ. HN = half-normal. Uni tau^2 = uniform on τ^2. IG = inverse gamma.

Figure 37

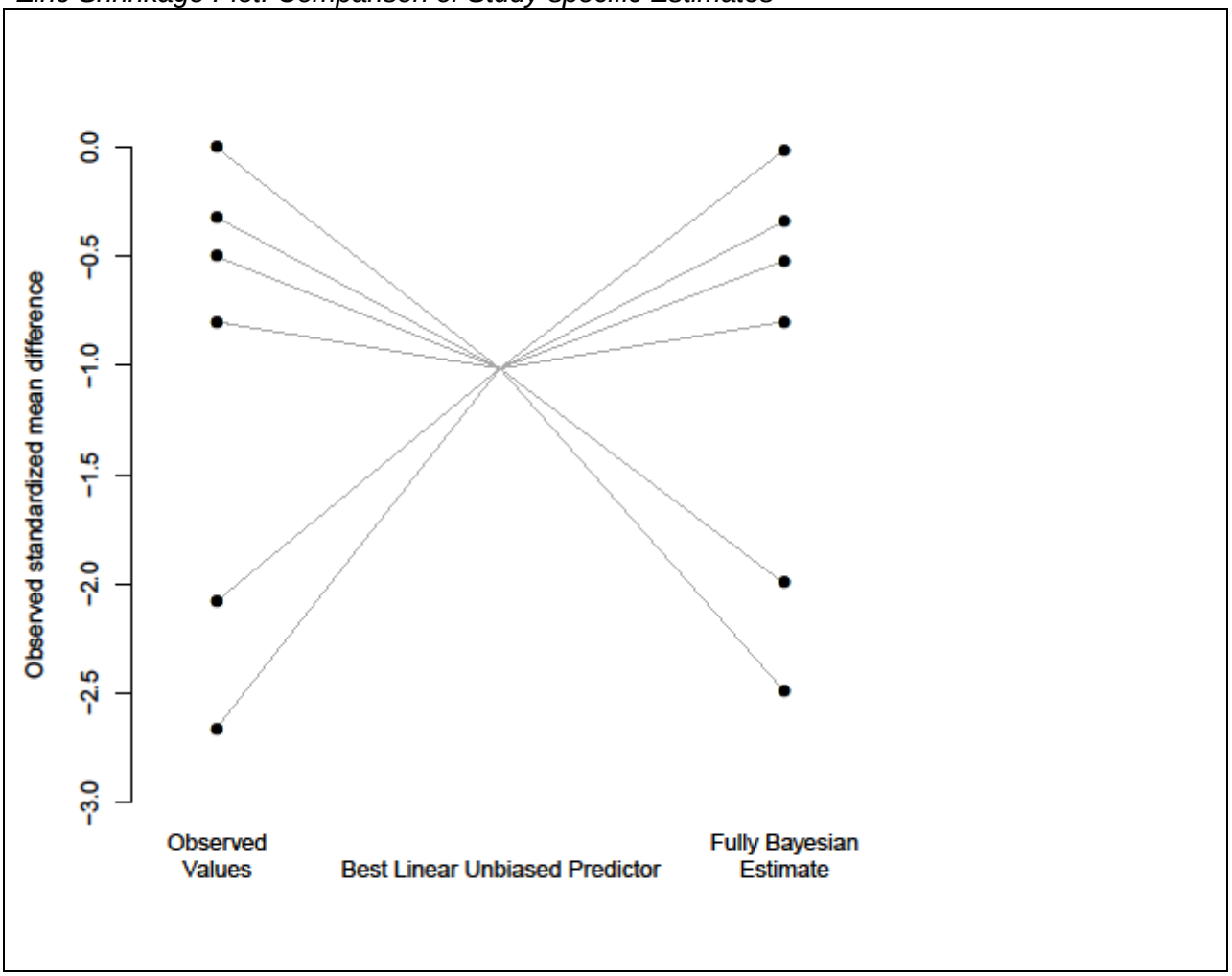
Zinc Heterogeneity Variance: Comparison of 95% Intervals



Note. Du. = DuMouchel. Uni. Tau = Uniform on τ. HN = half-normal. Uni tau^2 = uniform on τ^2. IG = inverse gamma.

Figure 38

Zinc Shrinkage Plot: Comparison of Study-specific Estimates



Data Set 4: Effects of Chondroitin in reducing pain in osteoarthritis

Classical Meta-Analysis Results for Chondroitin

Reichenbach et al.'s (2007) meta-analytic results were successfully replicated with the *metafor* package using a simple RE model with D-L estimation for: (a) all of the *k*=20 studies ($\hat{d}$ = -0.745, $p < 0.000$, 95% CI (-0.989. -0.501), I^2=91.8%) and (b) the *k* = 3 ($\hat{d}$ = -0.041, *p* = .812, 95% CI (-0.381, 0.298, I^2=0%), studies that were summarized by Reichenbach et al. to be the methodologically sound studies. The results from the different classical meta-analytic models for the *k* = 20 studies are displayed in Table 27. I^2, the proportion of variability that is due to heterogeneity, was estimated from the simple RE model as 93.1% with REML estimation and 91.8% using the standard D-L method of estimation. The *Q*-test for heterogeneity was significant at $p<.001$, indicating the presence of significant variation between effect sizes.

New coding of the chondroitin moderator variables for the mixed-effects model

For the reanalysis of the Chondroitin meta-analytic data with a hierarchical linear model, the coding of the study-level variable, *duration of follow-up*, was changed from a categorically coded variable of greater or less than 6 months duration to a continuous variable using the original value (i.e., *weeks*) for the time of the outcome assessment. Here, in the recoded data set, *weeks*, represents the deviation in weeks of each study from the overall mean study duration value of 41.11 weeks. Furthermore, the coding of the study-level variable, trial size, was changed from a categorically coded variable of greater than or equal to 200 patients or less than 200 patients to a continuous variable using the value of *N* for each study that was provided in the original systematic review. R 2.11.1 was used to find the appropriate transformations for the linear variables to improve their linearity and interpretability. The study-level variable *N* (trial size) was transformed to $100/\sqrt{N}$ in order to make the relationship between Y_i and *N* more linear and achieve greater symmetry about the regression of sample size on Y_i. In addition, a dummy variable was created for the two experimental studies (Kerzberg, 1987; Rovetta, 1991)

that used an intramuscular route of injection to administer the chondroitin dosage. Table 26 displays the recoded moderator variables.

Mixed-Effects Model with REML, k = 20

In order to investigate the effects of the study-level covariates, instead of conducting separate univariate meta-analyses of the moderator variables as Reichenbach et al. (2007) did in their original meta-analysis, a backward-stepwise model selection procedure with REML estimation was used with all of the study-level covariates. As shown in Table 27, when all of the k = 20 studies were included, the final fitted model resulted in a non-significant intercept $\hat{\beta}_0$= -0.049, and two significant study-level covariates, *intention to treat* (ITT) and *route of drug administration*. Here, for those studies that did not include an ITT analysis, the linear combination estimate for this group was -0.809, *se* = .123, $p < .001$, indicating that this group experienced a significant reduction in pain with the use of chondroitin. For the group of intramuscular injection studies (k=2), the linear combination estimate was -0.902, *se* = .506, p = .075.

Regarding the sensitivity of the different mixed-effects REML classical meta-analytic models, the coefficients for the study-level characteristics (*ITT* and *Route*) are no longer significant when $\tau^2$was set to the upper bound of its 95% CI. As shown in Table 27, when the Knapp and Hartung (2003) adjustment is applied to the standard errors and test statistics for the regression coefficients, the standard errors are larger, and the testing of significance of the regression coefficients is more conservative so that *Route* is no longer a significant study-level covariate.

Although the original meta-analysis included both the intramuscular and oral dosage experiments in a single meta-analysis, for the purposes of this research project, a decision was made to also analyze this meta-analytic data set by excluding the two intramuscular injection studies from the classical and Bayesian reanalysis for the following reasons: (a) the studies that

used an intramuscular injection of chondroitin were considered to be *substantively different* from the other (over-the-counter oral supplement) studies to be considered in one meta-analysis , (b) the *Q*-*Q* plot shows that for the classical REML fitted model the residual for the Rovetta study is outside its expected range, (c) the leave-one-out case diagnostics (see Table 28) from the simple RE model indicated that the Rovetta study may be an outlier, and (d) as will be discussed in the Bayesian results section, the cross-validated predictive probability for the residual from the Rovetta (intramuscular) study was in the .007 tail of its Bayesian predictive distribution.

Table 26

Chondroitin: Recoded Moderator Variable Data

Study	Y_i	se_i	Weeks[a]	Weeks.Dev[b]	$100/\sqrt{N}$[c]
Kerzberg 1987	-1.01	0.474	6	-35.1	24.253
Rovetta 1991	-2.14	0.337	51	9.9	15.811
Conrozier 1992	-1.93	0.27	24	-17.1	13.363
Lhirondel 1992	-0.53	0.179	25	-16.1	8.804
Mazieres 1992	-0.64	0.194	21	-20.1	9.128
Morreale 1996	-1.81	0.179	25	-28.1	8.276
Bourgeois 1998	-0.87	0.184	13	-14.1	8.873
Busci 1998	-0.94	0.219	26	-15.1	10.846
Conrozier 1998	-0.57	0.199	52	10.9	9.805
Uebelhart 1998	-1.17	0.296	52	10.9	14.744
Alekseeva 1999	-0.57	0.204	39	-2.1	10.000
Malaise 1999	-0.42	0.189	52	10.9	9.128
Pavelka 1999	-1.23	0.204	13	-28.1	9.759
Mazieres 2001	-0.23	0.179	27	-14.1	8.703
Nasonova 2001	-0.86	0.107	26	-15.1	4.244
Soroka 2002	-0.34	0.199	52	10.9	10.000
Michel 2005	-0.14	0.112	103	61.9	5.773
Clegg 2006	0.01	0.082	24	-17.1	3.981
Kahan 2006	-0.02	0.082	132	90.9	4.010
Mazieres 2006	-0.30	0.112	34	-7.1	5.670

Note. [a]Weeks at pain assessment was used as a continuous variable in this re-analysis rather than greater than or less than 6 months until pain assessment.
[b]Weeks.Dev represents each study's deviation from the overall mean value of 41.1 weeks.
[c]The reciprocal of the square root of *N* multiplied by 100 was used as the transformed variable in this re-analysis rather than using trial size of greater than or less than 200.

Figure 39

Chondroitin: Forest Plot for Mixed-effects Model (all studies, k=20) REML

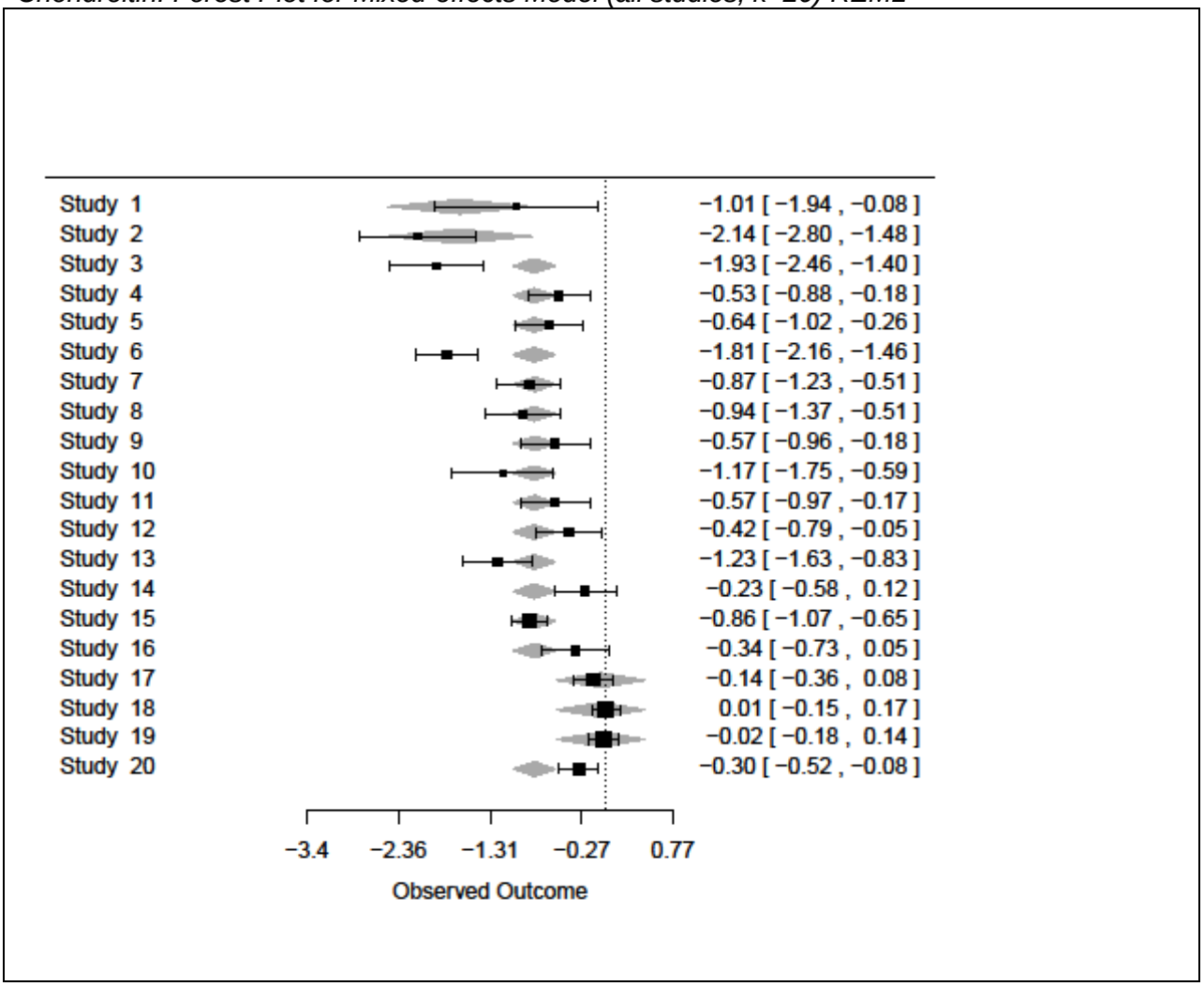


Table 27

Chondroitin: Classical Meta-analytic Results (k=20)

Estimation Method	FE	D-L	REML	REML	$REML_{KH}$	$REML_L$ τ^2 =.171	$REML_H$ τ^2=.766
Model	Simple	Simple	Simple	Mixed	Mixed	Mixed	Mixed
$\hat{B}_0$	-0.458	-0.745	-0.751	-0.049	-0.049	-0.049	-0.050
s.e.	0.034	0.125	0.136	0.257	0.264	0.245	0.508
p value	<.001	<.001	<.001	.848	.855	0.841	0.922
95% CI low	0.525	-0.990	-1.016	-0.552	-0.605	-0.529	-1.046
95%CI High	-0.391	-0.501	-0.485	0.454	0.507	0.431	0.946
$\hat{B}_1$ ITT				-0.760	-0.760	-0.759	-0.772
s.e.				0.285	0.292	0.272	0.559
p value				.008	.019	.005	.167
95% CI low				-1.318	-1.377	-1.292	-1.867
95% CI high				-0.202	-0.143	-0.226	0.323
$\hat{B}_2$ Route				-0.853	-0.853	-0.859	-0.787
s.e.				0.436	0.448	0.424	0.721
p value				.050	.074	.043	0.275
95% CI low				-1.708	-1.798	-1.690	-2.200
95% CI high				0.001	0.091	-0.028	0.626
τ		0.518	0.569	0.434	0.434	0.414	0.875
95% CI low		0.413	0.413	0.303	0.303	0.303	0.303
95% CI high		0.875	0.875	0.723	0.723	0.723	0.723
τ^2		0.268	0.324	0.189	0.189	0.171	0.766
s.e.			0.119	0.078	0.078	0.072	0.092
95% CI low		0.171	0.171	0.092	0.092	0.092	0.523
95%CI high		0.766	0.766	0.523	0.523	0.523	
Q	229.503	229.503	229.503	103.484	103.484	103.484	103.484
p value	< .001	< .001	< .001	< .001	< .001	<.001	<.001
PI $\hat{B}_0$ Lower			-1.980	-1.109	-1.117	-1.059	-2.176
PI $\hat{B}_0$ Upper			0.478	1.011	1.019	0.961	2.077
PI $\hat{B}_1$ Lower				-1.851	-1.860	-1.798	-2.953
PI $\hat{B}_1$ Upper				0.331	0.340	0.281	1.409
PI $\hat{B}_2$ Lower				-2.146	-2.164	-2.103	-3.169
PI $\hat{B}_2$ Upper				0.439	0.457	0.384	1.600
I^2		91.72%	93.04%				
Q_{Mods}				12.425	5.886(F)	13.463,	3.567
p value				.002	.011	.001	.168
Log-like	-99.157	-18.155	-19.188	-14.363	-14.363	-14.389	-18.742
Deviance	198.314	36.310	38.376	28.726	28.726	28.778	37.485
AIC	200.314	40.310	42.376	36.726	36.726	36.778	45.485
BIC	201.310	42.301	44.265	40.059	40.059	40.111	48.817
Min, max			-1.778	-1.960	-1.960	-1.951	-2.071
$\theta_i^*(\tau)$			-0.006	0.008	0.008	0.008	0.009

Note. FE= Fixed-effect model. D-L = DerSimonian & Laird estimation. REML = Restricted Maximum Likelihood Estimation. $REML_L$ = τ^2 was set equal to the lower bound of its 95% CI from the simple REML model. $REML_H$ = $\tau^2$was set equal to the upper bound of its 95% CI from the simple REML model. $REML_{KH}$ = Knapp and Hartung adjustment made to the standard errors of the regression coefficients. Permutation test *p* values for $\hat{B}_1$ and $\hat{B}_2$: <.001 and .096.

Figure 40

Chondroitin Q-Q Normal Plot, k=20

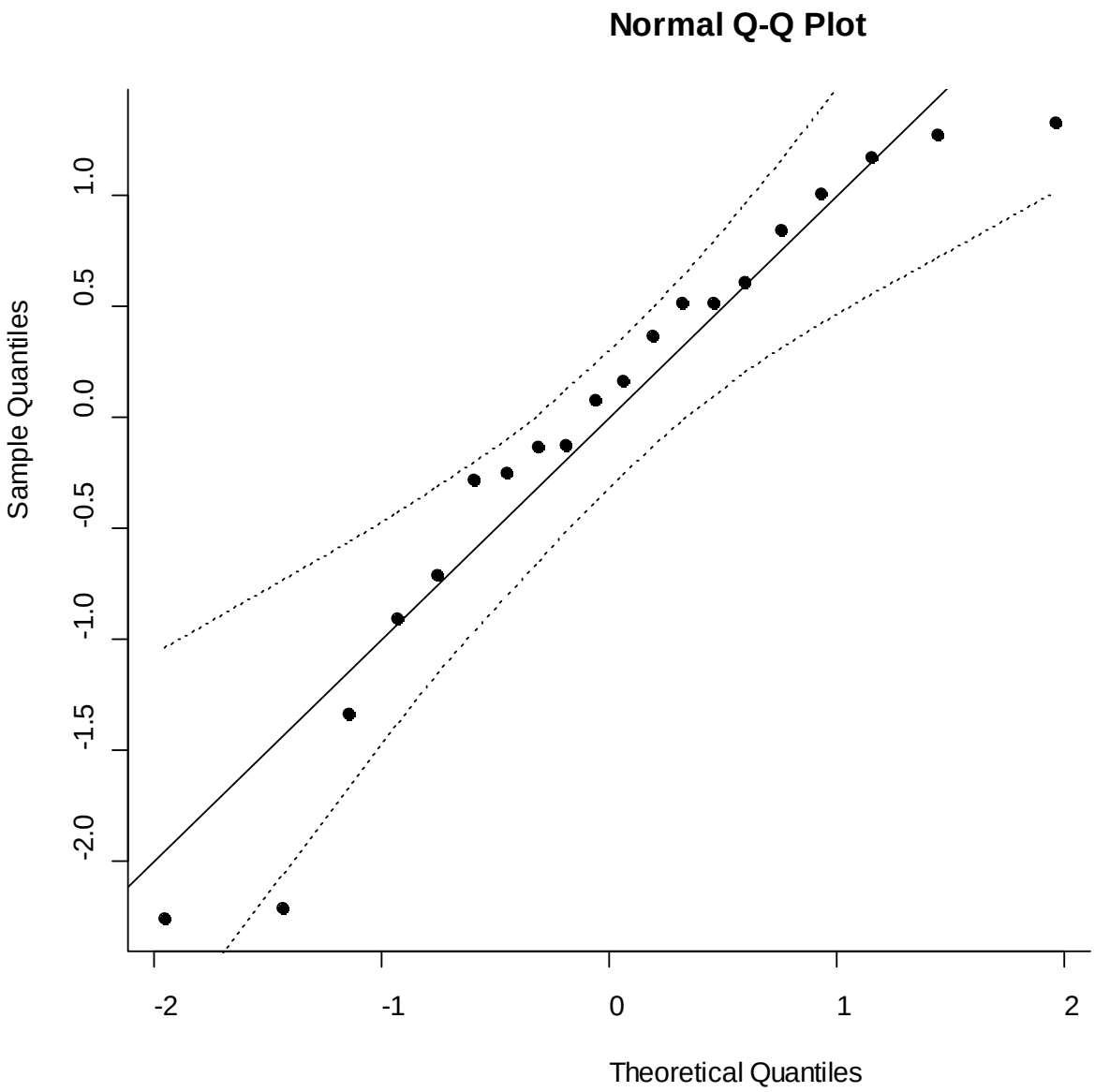


Table 28

Chondroitin Leave-one-out Case Diagnostics for Simple RE Model (k=20)

Study	rstudent[a]	cov.r[b]	τ^2.del[c]	QE.del[d]	B_0[e]
Kerzberg*	-0.351	1.069	0.336	228.142	-0.059
Rovetta*	-2.329	0.860	0.260	204.436	-0.523
Conrozier	-2.076	0.887	0.268	199.349	-0.496
Lhirondel	0.373	1.111	0.343	229.336	0.099
Mazieres	0.187	1.115	0.345	228.605	0.056
Morreale	-2.004	0.887	0.266	170.252	-0.517
Bourgeois	-0.196	1.113	0.344	224.338	-0.033
Busci	-0.308	1.104	0.342	224.546	-0.059
Conrozier	0.302	1.111	0.344	229.181	0.082
Uebelhart	-0.660	1.073	0.333	223.668	-0.137
Alekseeva	0.301	1.110	0.344	229.197	0.081
Malaise	0.556	1.100	0.340	229.461	0.139
Pavelka	-0.806	1.068	0.329	214.917	-0.182
Mazieres	0.890	1.073	0.330	227.815	0.211
Nasonova	-0.185	1.121	0.346	213.097	-0.031
Soroka	0.689	1.089	0.336	229.144	0.167
Michel	1.085	1.053	0.322	220.956	0.260
Clegg	1.395	1.009	0.307	191.995	0.328
Kahan	1.334	1.018	0.310	196.648	0.315
Mazieres	0.789	1.087	0.334	227.391	0.195

[a]denotes the r student (externally standardized residual) value with the i^{th} study excluded. [b]denotes the covariance ratio. [c]denotes the estimate value of τ^2based on the data set with the i^{th} study excluded. [d]denotes the *Q* statistic for the test of heterogeneity based on the data set with the i^{th} study removed. [e]denotes by how many standard deviation units the regression coefficients change when the i^{th} study is excluded (Viechtbauer, 2009). *= trial that used intramuscular injection.

Mixed-Effects Model with REML, k = 18

After exclusion of the two intramuscular injection studies, as shown in Table 29, the fitted model resulted in a non-significant intercept and two significant study-level covariates, $\hat{\beta}_1$(*Analgesic Use*), and $\hat{\beta}_2$(*Weeks*). Here, by including these two study-level covariates the residual heterogeneity is equal to τ^2 = 0.162, resulting in a 40% reduction in the size of the variance that can be attributed to heterogeneity.

Let $X_i\beta = \hat{\beta}_0 + \hat{\beta}_{i1}(Analgesic\ Use)_i + \hat{\beta}_{i2}(Weeks)_i$, where the estimated intercept, $\hat{\beta}_0 = -0.319$, represents the average reduction in pain in terms of standardized mean differences for those study participants who took oral chondroitin supplement and were average on the mean value of weeks (41.1) and used analgesic medicine similar to the control group. In this hierarchical linear model the intercept is non-significant at *p* = .090. However, there was a large benefit of chondroitin for the studies that were categorized as either reporting a greater or "unclear" amount of analgesic drug use among the control group. The linear combination estimate for those studies which had control groups that used a greater or "unclear" amount of analgesic medicine was -0.826, *se* = 0.102, $p < .001$. This indicates that there was an appreciable effect in pain reduction for chondroitin, especially for those studies of shorter duration (less than 41 weeks). In this model the *Analgesic Use* study-level covariate is likely to be a confounding variable because it is correlated with both the independent and dependent variable.

Comparison of the model fits statistics for the different models revealed that for the simple RE (D-L) model the AIC and BIC values are at their lowest, which is indicative of the best fitting model. The mixed-effects REML model had the next lowest values. An examination of the sensitivity of the meta-analytic model to different values of τ^2, as shown in Table 29, reveals that, when τ^2 is set to its upper 95% CI boundary, none of the study-level covariates are significant predictors of effect size.

For this data set the Orwin Fail-safe *N* value was 18, which indicates that 18 studies with an average of a null result would be required in order to reduce the average unweighted effect size (from a simple RE model) by half, from -0.698 to a target effect size value of -0.349. The funnel plot here is nearly symmetric and Egger's test for the mixed-effects regression model was not significant (*p*=.075). The trim and fill method indicated that there is only one study missing on the left side of the funnel (e.g., a study that was in favor of pain reduction). If such a study were to be imputed, then the simple RE model would still yield similar results with the following estimates: $\hat{\beta}_0 = -0.720$, *p* = <.000, 95% CI(-0.975, -0.466), τ= 0.538 in favor of the use of chondroitin for pain reduction in people with osteoarthritis.

Table 30 displays the statistical power of the covariates (moderators) to be able to detect differences that may exist among the effect sizes due to the study characteristics. The statistical power is low for all of the potential moderator variables with the exception of *Study Size* and *Weeks*. It is important to note that the low power of covariates may lead to invalid moderator tests because the tests may not be sufficiently powerful enough to detect real effects, even if such true effects were present (Hedges and Pigott, 2004).

Table 29

Chondroitin: Classical Meta-analysis Results (k = 18)

Estimation	FE	D-L	REML	REML	REML KH	$REML_L$ τ^2= .137	$REML_H$ τ^2= .660
Model	Simple	Simple	Simple	Mixed	Mixed	Mixed	Mixed
$\hat{B}_0$	-0.4379	-0.673	-0.675	-0.319	-0.319	-0.319	-0.318
s.e.	0.0343	0.123	0.130	0.188	0.190	0.174	0.369
p value	<.000	<.000	<.000	.090	.114	.067	.389
95% CI low	-0.505	-0.915	-0.929	-0.688	-0.724	-0.661	-1.042
95%CI High	-0.371	-0.431	-0.421	0.049	0.086	0.022	0.405
$\hat{B}_1$ Analgesic				-0.507	-0.507	-0.505	-0.520
s.e.				0.226	0.228	0.210	0.437
p value				.025	.042	0.016	.234
95% CI low				-0.949,	-0.993	-0.916,	-1.376
95% CI high				-0.065	-0.021	-0.094	0.336
$\hat{B}_2$ Weeks.Dev				0.008	.008	0.008	0.008
s.e.				0.003	0.003	0.003	0.006
p value				.015	.030	.009	.208
95% CI low				0.002,	0.001	0.002,	-0.005
95% CI high				0.014	0.015	0.014	0.021
τ		0.492	0.518	0.402	0.402	0.370	0.816
95% CI low		0.371	0.371	0.271	0.271	0.271	0.271
95% CI high		0.812	0.812	0.675	0.675	0.675	0.675
τ^2		0.242	0.268	0.162	0.162	0.137	0.666
s.e.			0.103	0.072	0.072	0.062	0.256
95% CI low		0.137	0.137	0.074	0.074	0.074	0.074
95%CI high		0.660	0.660	0.455	0.455	0.455	0.455
Q		202.9	202.9	112.351	112.351	112.351	112.351
p value		<.001	<.001	<.001	<.001	<.001	<.001
PI $\hat{B}_0$ Lower			-1.810,	-1.263,	-1.262	-1.187,	-2.217
PI $\hat{B}_0$ Upper			0.457	0.622	0.624	0.548	1.580
PI $\hat{B}_1$Lower				-1.485	-1.487	-1.407	-2.482
PI $\hat{B}_1$Upper				0.4705	0.473	0.397	1.442
PI $\hat{B}_2$Lower				-0.845	-0.845	-0.777	-1.722
PI $\hat{B}_2$Upper				0.861	0.861	0.793	1.738
I^2		91.63%	92.39%				
Q_{Mods}				11.12	5.445(*F*)	12.7	3.097,
p value				.004	.017	.002	.213
Log-likelihood	-85.894	-14.379	-15.514	-16.234	-16.234	-16.303	-20.276
Deviance	171.788	28.758	31.028	32.467	32.467	32.605	40.552
AIC	173.788	32.758	35.028	40.467	40.467	40.605	48.552
BIC	174.678	34.539	36.695	43.299	43.299	43.438	51.384
Min, max $\theta_i^*(\tau)$			-1.689, -0.007	-1.669, -0.009	-1.669 - 0.009	-1.648, 0.012	-1.951 0.008

Note. $REML_L$ = τ^2 was set equal to the lower bound of its 95% CI from the simple REML model. $REML_H$ = $\tau^2$was set equal to the upper bound of its 95% CI from the simple REML model. $REML_{KH}$ = Knapp and Hartung adjustment made to the standard errors of the regression coefficients. Permutation test *p* values for REML mixed-effects: $\hat{B}_1$ and $\hat{B}_2$: .018 and <.001.

Table 30

Chondroitin: Power for Two-sided Test of the Regression Coefficients (k = 18)

	Power for Individual Regression Coefficients Backward Stepwise Regression Models					
	Full Model	Model 1	Model 2	Model 3	Model 4	Final Model
Patient Blind	.123					
ITT	.064	.073				
Placebo controlled	.124	.175	.200			
N	1.000	1.000	1.000	1.000		
Conceal	.072	.084	.143	.155	.560	
Analgesic	.152	.064	.235	.252	.311	.310
Weeks	1.000	1.000	1.000	1.000	1.000	1.000
Intercept	.094	.111	.119	.125	.169	.419

Note. α = .05 with a clinically meaningful effect set equal to 0.33 standard deviations.

Table 31.

Chondroitin: Correlation among Variables (k=18)

	Conceal	Placebo	Blind	ITT	Analgesic	Weeks	100/sqrtN	Y_i
Conceal	1.000							
Placebo	.158	1.000						
Patient.Blind	-.081	.561	1.000					
ITT	.791	.200	.051	1.000				
Analgesic	.175	-.055	-.014	.388	1.000			
Weeks	-.433	-.031	.277	-.671	-0.035	1.000		
100/sqrt *N*	.190	-.088	-.429	.360	.396	-.401	1.000	
Y_i	-.451	.089	.270	-.534	-.450	.470	-.524	1.000

Bayesian Meta-analysis Results

Hierarchical Bayes Linear Model, k=20 (all studies included)

In WinBUGS with all of the studies included, a backward stepwise model selection procedure with a uniform prior distribution for τ (0,5) was used to fit the data. As detailed in Table 32 and the Appendix, when the $\delta_i \sim t_4$ the Bayesian estimate that the probability that the effect is less than zero and in favor of pain reduction in a new study is 52.5% for a study that uses oral supplements and an ITT analysis. For those studies that do not conduct ITT analyses, there is a 96.7% chance that the effect in the new study will be less than zero, and, for those studies that use intramuscular injections, there is a 92.3% probability that there will be a favorable effect of chondroitin in terms of pain reduction.

As depicted in Table 33 and Figure 41, Bayesian cross-validation of the fitted model for all of the studies (*k*=20) revealed that the Rovetta study, which used intramuscular injections, resulted in a predictive probability value of .007, indicating that this study result is in .7% of the tail of its predictive distribution, given all of the other studies. The Bonferroni bound for the probability of such an event was .263. There was some evidence suggesting that the two intramuscular injection studies may be too different from the oral supplement studies to be considered in one meta-analysis due to the: a) substantive differences in the route of chondroitin administration and b) predictive probability cross-validation evidence showing that the Rovetta study may indeed be an outlying data point. For further Bayesian hierarchical linear modeling a decision was made to synthesize the intramuscular studies separately and to narrow the conceptualization of the reanalysis of the chondroitin efficacy meta-analysis so that the reanalysis only included the studies that used oral supplements.

Table 32

Chondroitin: Bayesian results (k=20, all studies)

Parameter	Mean	s.d.	Prob < 0	95% Cr Int
$\hat{\beta}_0$	-0.046	0.248	.579	-0.575, 0.454
$\hat{\beta}_1$ (No ITT)	-0.695	0.281	.989	-1.267, -0.134
$\hat{\beta}_2$ (Intramuscular)	-0.941	0.489	.963	-1.847, 0.129
$\hat{\beta}_0 + \hat{\beta}_1$	-0.741	0.123	1.000	-0.982, -0.509
$\hat{\beta}_0 + \hat{\beta}_2$	-0.987	0.561	.963	-2.072, 0.176
$\theta_{new.\beta 0}$	-0.041	0.458	.525	-0.977, 0.859
$\theta_{new.\ \beta 0+\beta 1}$	-0.720	0.405	.967	-1.499, 0.044
$\theta_{new.\ \beta 0+\beta 2}$	-0.977	0.669	.929	-2.189, 0.359
τ	0.371	0.109		0.201, 0.614

Note. Prior for τ ~ uniform (0,5). Prior for **β** ~ (0,1000). $\delta_i \sim t_4(0, \tau^2)$. DIC = 7.5.

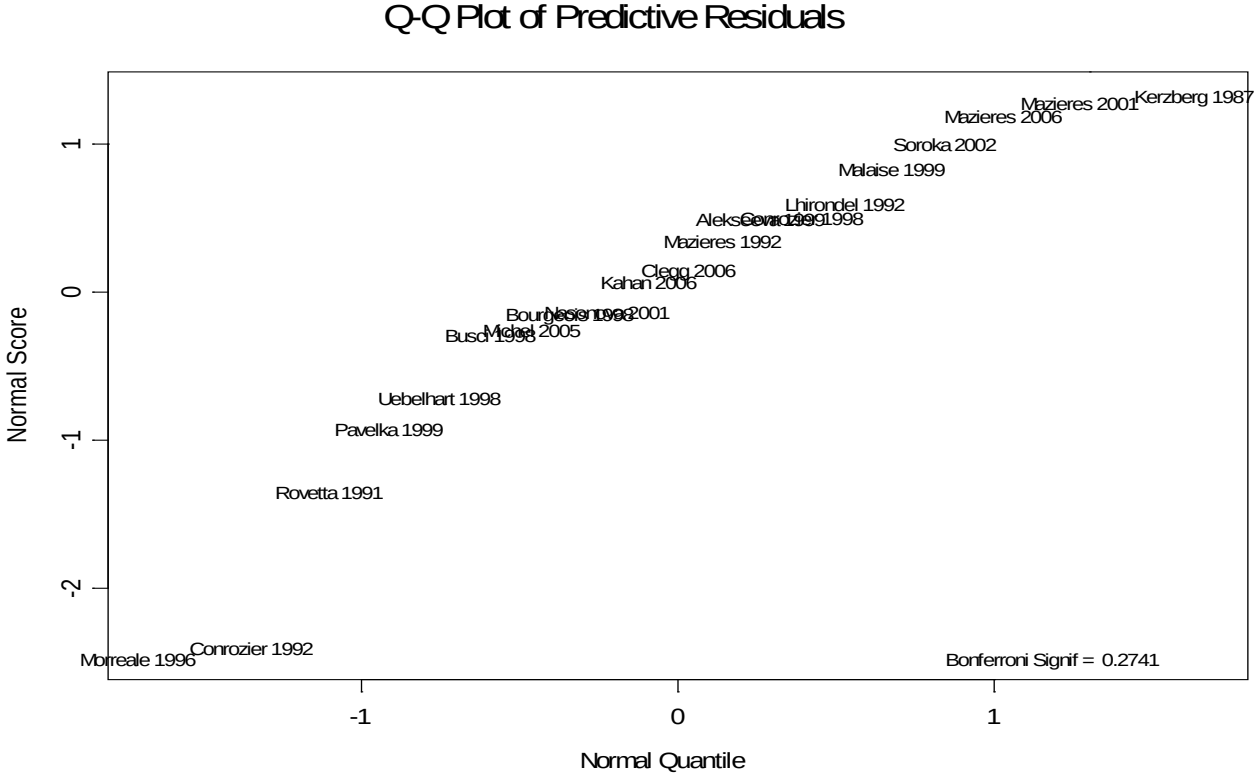
Table 33

Chondroitin: Bayesian cross-validation results, all studies (k=20)

Study	Y_i	se_i	Prior Mean	Prior s.d.	Pred. Prob.	τ^2	Post. Mean	Post. sd	Prob > 0
Kerzberg	-1.01	0.474	-1.170	0.532	0.590	0.242	-1.085	0.342	0.001
Rovetta	-2.14	0.337	-0.743	0.430	0.007	0.168	-1.668	0.303	0.000
Conrozier	-1.93	0.27	-0.949	0.464	0.034	0.194	-1.687	0.247	0.000
Lhirondel	-0.53	0.179	-1.063	0.506	0.849	0.231	-0.598	0.169	0.000
Mazieres 92	-0.64	0.194	-1.088	0.515	0.802	0.237	-0.705	0.181	0.000
Morreale	-1.81	0.179	-0.937	0.457	0.038	0.188	-1.698	0.172	0.000
Bourgeois	-0.87	0.184	-1.136	0.531	0.691	0.246	-0.904	0.172	0.000
Busci	-0.94	0.219	-1.018	0.524	0.556	0.249	-0.953	0.198	0.000
Conrozier	-0.57	0.199	-0.842	0.519	0.695	0.246	-0.611	0.183	0.000
Uebelhart	-1.17	0.296	-0.797	0.512	0.256	0.239	-1.063	0.253	0.000
Alekseeva	-0.57	0.204	-0.942	0.513	0.759	0.242	-0.630	0.188	0.000
Malaise	-0.42	0.189	-0.854	0.512	0.797	0.238	-0.481	0.176	0.003
Pavelka	-1.23	0.204	-1.091	0.532	0.398	0.248	-1.208	0.188	0.000
Mazieres 01	-0.23	0.179	-0.468	0.564	0.665	0.247	-0.257	0.169	0.064
Nasonova	-0.86	0.107	-0.306	0.545	0.148	0.229	-0.836	0.105	0.000
Soroka	-0.34	0.199	-0.860	0.506	0.840	0.233	-0.420	0.185	0.012
Michel	-0.14	0.112	0.332	0.646	0.223	0.239	-0.123	0.110	0.132
Clegg	0.01	0.082	-0.578	0.541	0.869	0.224	-0.005	0.081	0.473
Kahan	-0.02	0.082	-0.547	0.774	0.761	0.244	-0.027	0.081	0.370
Mazieres 06	-0.3	0.112	-0.387	0.566	0.564	0.252	-0.304	0.109	0.003

Figure 41

Chondroitin: Bayesian Q-Q Plot of Predictive Residuals for fitted model, (k=20)



Hierarchical Bayes Linear Model, k=18 (only the oral supplement studies included)

In WinBUGS with only the oral administration studies included (*k* = 18), a backward stepwise model selection procedure with a uniform prior distribution for τ was used to fit the data. As detailed in the Bayesian Quality Assurance checklist and in the Appendix, this Bayesian model resulted in a mixed-effects model with two different covariates that were significant (*Analgesic Use* and *Weeks*). Here, there is a 79.2% chance that in a new study the use of oral chondroitin supplements will result in a treatment effect that is less than zero (i.e., result in a pain reduction) for people with osteoarthritis who are at the average value on weeks and use an amount of pain medicine that is equivalent to that used by the control group. For those studies characterized by control groups that used a greater or "unclear" amount of analgesic medicine, there is a 99.99% chance that in a new study the chondroitin group will experience a treatment effect in favor of pain reduction. However, it is important to note that in this meta-analytic model, the *Analgesic Use* study-level covariate may be a confounding variable, because it is correlated with both the independent and dependent variable.

Cross-validation of the Bayesian fitted model for the *k* =18 data set revealed that the two studies, (Conrozier 1992; Morreale, 1996) had predictive probability values that were within 1.3 to 1.5% of the tail of their predictive distribution given all of the other studies, however, the Bonferroni bound for the probability of these events was not significant.

In a sensitivity analysis for the seven different prior distributions for τ, whenever $\delta_i \sim$ normal, the regression coefficients and estimates for τ were larger for all of the prior distributions on τ, as shown in the tables in the Appendix. All of the Bayesian posterior 95% credibility intervals for the model estimates are similar, with the exception of the 95% credibility intervals for the residual variance produced by the two prior distributions, which allow for a stronger preference for larger values of τ (e.g., the half-normal on τ and the uniform on τ^2).

A trace plot of τ that depicts the dependency of the intercept (curve labeled A) and the study-specific estimates (curves labeled B-S) on the different values of τ is displayed in Figure

44. Overall, as depicted in the trace plot, the majority of the studies experience minimal shrinkage and borrowing of strength from the entire ensemble of oral supplement chondroitin studies. However, the two studies that had extreme leave-one-out predictive probabilities and lower reliabilities (Conrozier, 1992; Morreale, 1996) experienced a large amount of shrinkage.

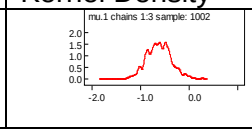
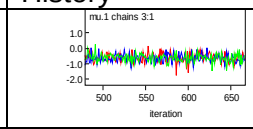
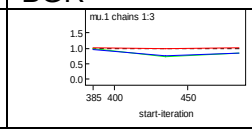
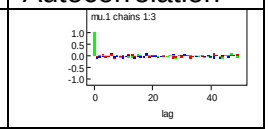
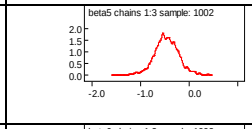
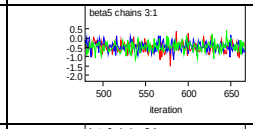
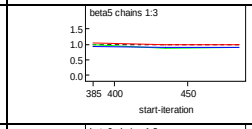
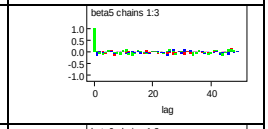
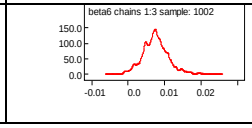
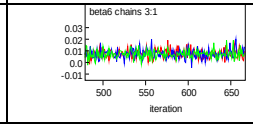
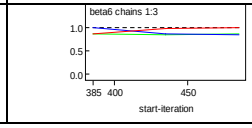
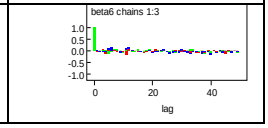
Figure 42

Chondroitin: Bayesian Quality Assurance Meta-analysis Checklist, k=18

Cross-validation ☒ Outlier Detected?: ☒No ☐Yes
If yes, study = Predictive Prob. = Bonferroni significance =
Fitted model: $X_i\beta = \hat{\beta}_0 + \hat{\beta}_{i1}$(Analgesic Use) + $\hat{\beta}_{i2}$ (Deviation from mean week)
Prior Distribution for τ:
☒ Uniform on τ ☐ Uniform on τ^2 ☐ Half-norm ☐ Inverse Gamma ☐ DuM. s0 ☐ DuM. s0/3 ☐ DuM. 3s0
Random effect modeled from: Prior Distribution for **β**:
☐ Normal distribution ☒ t_4-dist. ☒ Normal (0,1000)
Model specification:
☒ Run 3 chains ☒ 50,000 iterations ☒ Discard first half ☒ DIC = 2.79
Do meta-analytic inferences change from fitted hblm?: ☒No ☐ Yes. If yes, impact: _____

Node	Mean	95% Credibility Interval	Set parameter	MC error <5% of s.d.	Rhat proximity to 1
β_0	-0.311	-0.719, 0.086	☒	☒	☒
β_1	-0.452	-0.961, 0.033	☒	☒	☒
β_2	0.007	.000, 0.015	☒	☒	☒
τ	0.358	0.204, 0.619	☒	☒	☒
$\beta_0 + \beta_1$	-0.763	-1.034, -0.490	☒	☒	☒
$\theta_{new\cdot\beta0}$	-0.311	-1.161, 0.602	☒	☒	☒
$\theta_{new\ .\beta0 + \beta1}$	-0.769	-1.536,-0.018	☒	☒	☒
$\theta_{new\ .\beta1}$	-0.444	-1.401, 0.498	☒	☒	☒
$\theta_{new.\ \beta2}$	0.002	-0.813, 0.701	☒	☒	☒
p $\beta_0 < 0$	.941	0, 1	☒	☒	☒
p $\beta_1 < 0$	.969	0,1	☒	☒	☒
p $\beta_2 < 0$	.027	0,1	☒	☒	☒
p $\beta_0+\beta_1 < 0$	1.000	1,1	☒	☒	☒
p $\theta_{new.\beta0} < 0$	.792	0,1	☒	☒	☒
p $\theta_{new.\beta0 + \beta1} < 0$	.978	1,1	☒	☒	☒
p $\theta_{new.\beta1} < 0$	.838	0,1	☒	☒	☒
p $\theta_{new.B2} < 0$	.498	0,1	☒	☒	☒
Deviance	-13.039	-22.977, 1.178	☒	☒	☒
θ_i^*	Min θ_4^* = -1.7, Max θ_{16}^* = -.007		☒	☒	☒

Items pertaining to parameter diagnostic checks:

	Kernel Density	History	BGR	Autocorrelation
☒ β_0				
☒ β_1				
☒ β_2				

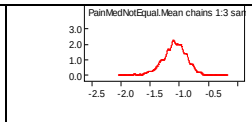
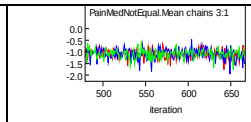
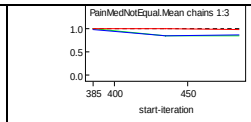
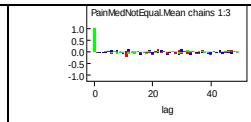
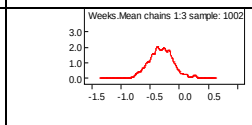
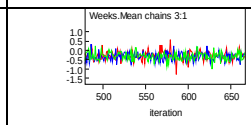
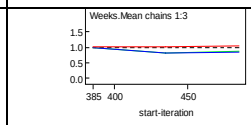
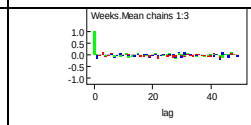
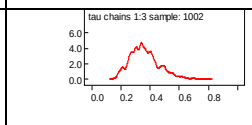
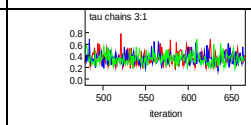
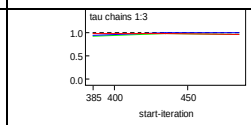
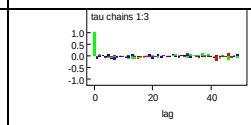
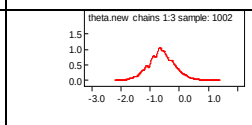
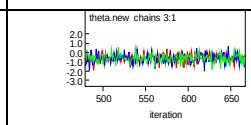
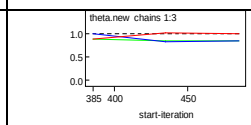
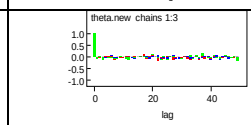
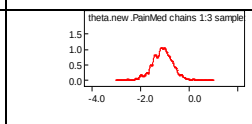
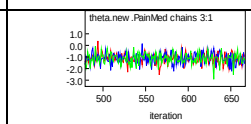
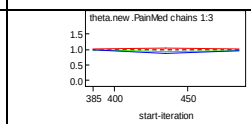
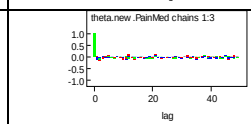
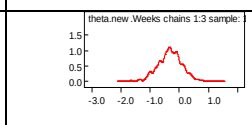
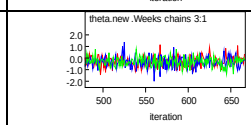
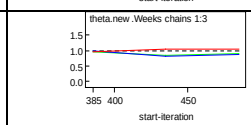
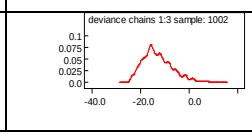
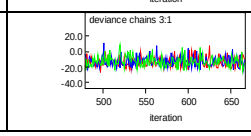
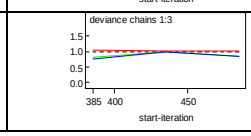
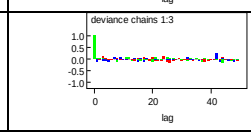
☒ $\beta_0+\beta_1$				
☒ $\beta_0+\beta_2*41$				
☒ τ				
☒θ_{new} for β_0				
☒θ_{new} for $\beta_0 + \beta_1$				
☒θ_{new} for $\beta_0 + \beta_2$				
☒ Deviance				

Figure 43

Chondroitin: Bayesian Q-Q Plot of Predictive Residuals for fitted model (k=18)

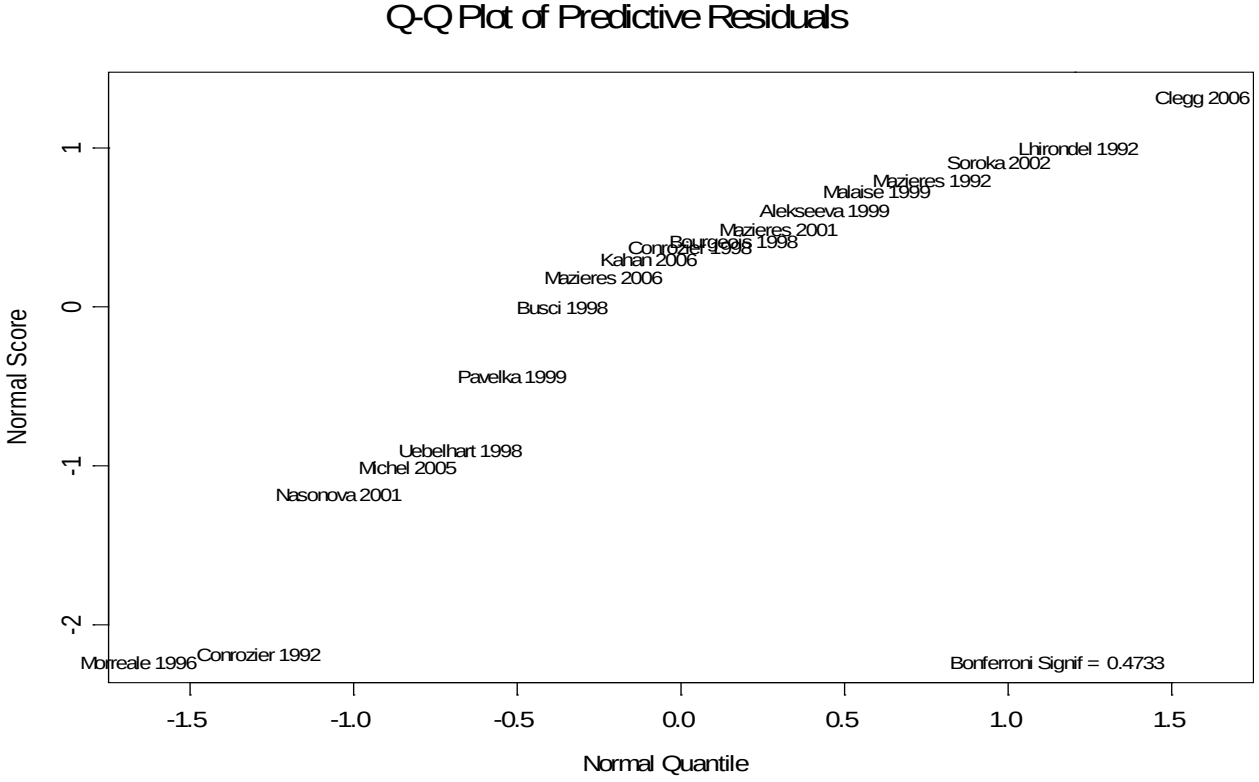
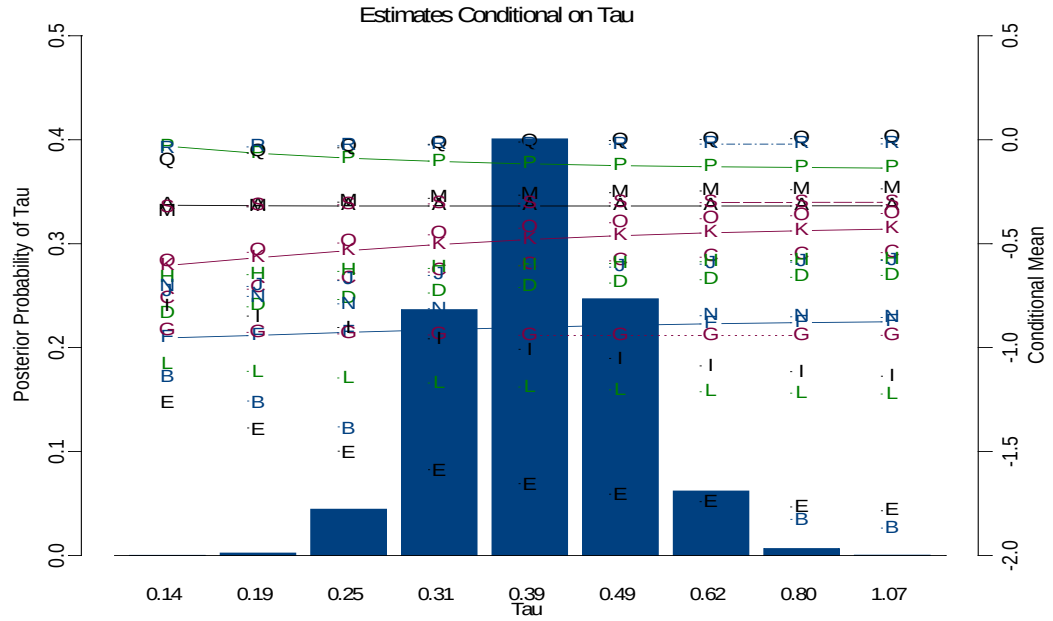


Figure 44

Chondroitin: Trace Plot of Tau for Fitted Mixed-effects Model, k = 18



A=(I B=C1992 C=L1992 D=M1992 E=M1996 F=Br1998 G=Bs1998 H=C1998 I=U1998 J=A1999 K=M1999 L=P1999 M=M2001 N=N2001 O=S2002 P=M2005 Q=C2006 R=K2006 S=M2006

Comparison of Results for Classical and Bayesian Meta-analytic Models

Figures 44 - 46 display the widths of the Bayesian 95% CrInts compared to the classical 95% CI widths with REML estimation for the fitted models for the: (a) mean effect size, (b) the predicted effect in a new study, and (c) τ^2, the residual variance. All of the Bayesian parameter estimates and meta-analytic inferences were similar to those inferences made from the classical REML analyses, with the exception of the classical mixed-effects $REML_{High}$ model.

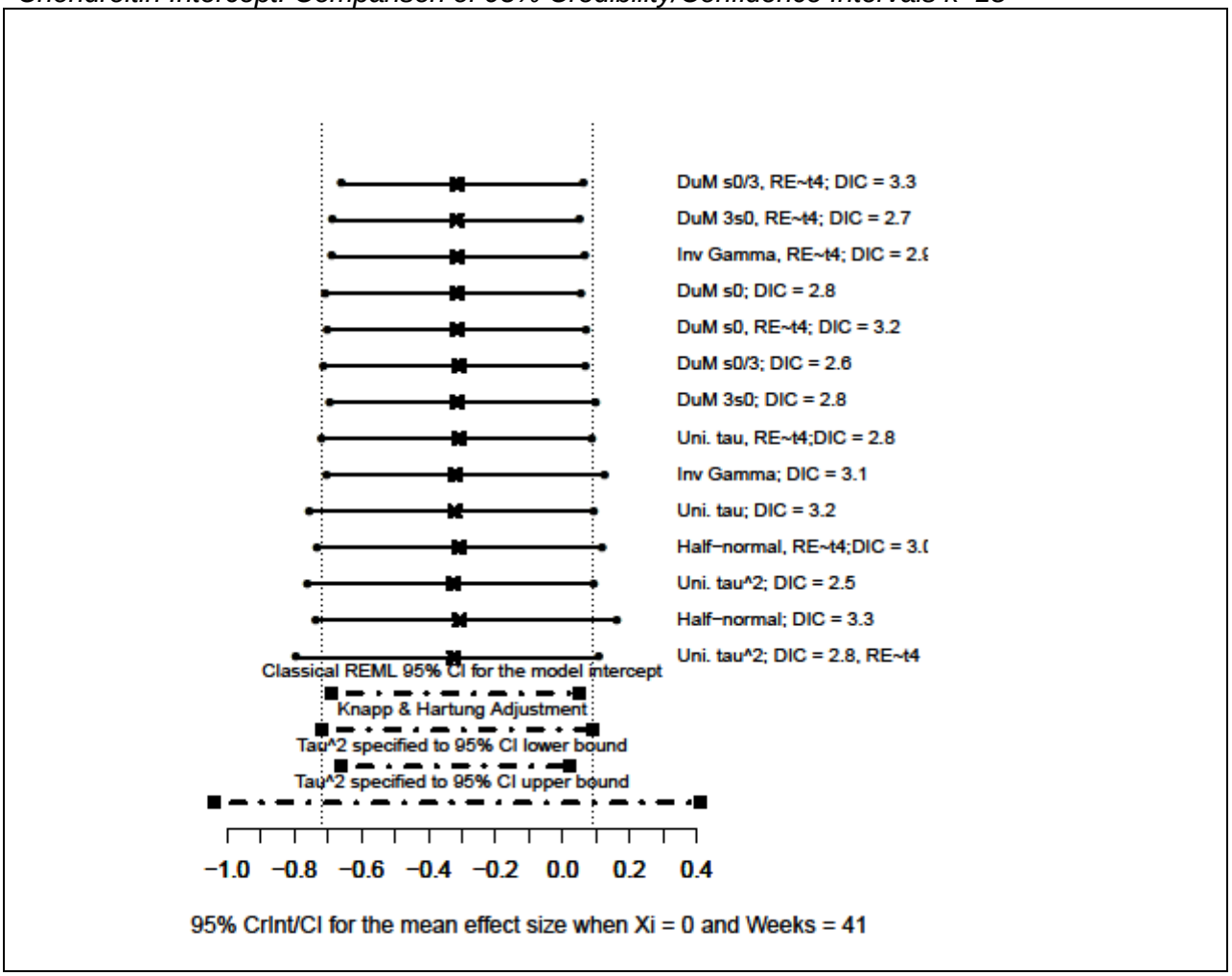
A comparison of the classical and Bayesian study-specific estimates is depicted in Figure 48. The Bayesian θ_i^* estimates provide unique estimates for each observed effect size because the θ_i^* estimates consider the full uncertainty in all of the auxiliary parameters as compared to the classical study-specific estimates.

The meta-analytic inferences made from both the classical mixed-effects (REML) and hierarchical Bayesian linear model differ from those made by the original systematic review. In the original review, Reichenbach et al. (2007) used simple RE (D-L) meta-analytic models to synthesize the results for: a) all the *k*=20 studies and b) only the *k*=3 studies that were considered to be "methodologically sound." With both hierarchical linear modeling and Bayesian modeling, when all of the studies were included in the model (*k* = 20), although there was a non-significant intercept, both intramuscular injection and lack of ITT analyses were found to reduce the between-study variance component and significantly predict effect size. From a Bayesian interpretation, there is a 96% probability that intramuscular injections of chondroitin and a 99% probability that studies that did not use an ITT analysis will result in an effect of pain reduction that is less than zero for patients in the chondroitin treatment group. The studies that included ITT analyses as part of their study design may have introduced less bias in their statistical analyses and effect size computations, because studies that use ITT analyses include all participants in the analysis regardless of treatment compliance. Here, the role of ITT as a moderating variable may be a de facto indicator of study quality for a well-designed (and less biased) experiment.

A sensitivity analysis of the fitted meta-analytic model of the *k* = 20 studies revealed that one of the intramuscular injection studies may be an outlier. When the intramuscular injection studies were excluded from the data set, the meta-analytic conclusions again differ from the original simple RE (D-L) model. This fitted mixed-effects model from the reanalysis of the *k* = 18 studies indicates that for the model intercept, there is a 94% probability that chondroitin will result in a reduction of pain that is less than zero in patients with osteoarthritis with two study-level characteristics moderating the effect: (a) duration of chondroitin treatment (e.g., the shorter duration treatments had larger effects) and (b) analgesic pain medicine use (e.g., chondroitin participants who used less pain medicine or an "unclear" amount of pain medicine as compared to their controls experienced greater amounts of pain reduction). However, it is important to note that the results of this meta-analytic model are difficult to interpret and may be invalid because the covariate, *Analgesic Use*, is confounded with both the treatment and outcome variables.

Figure 45

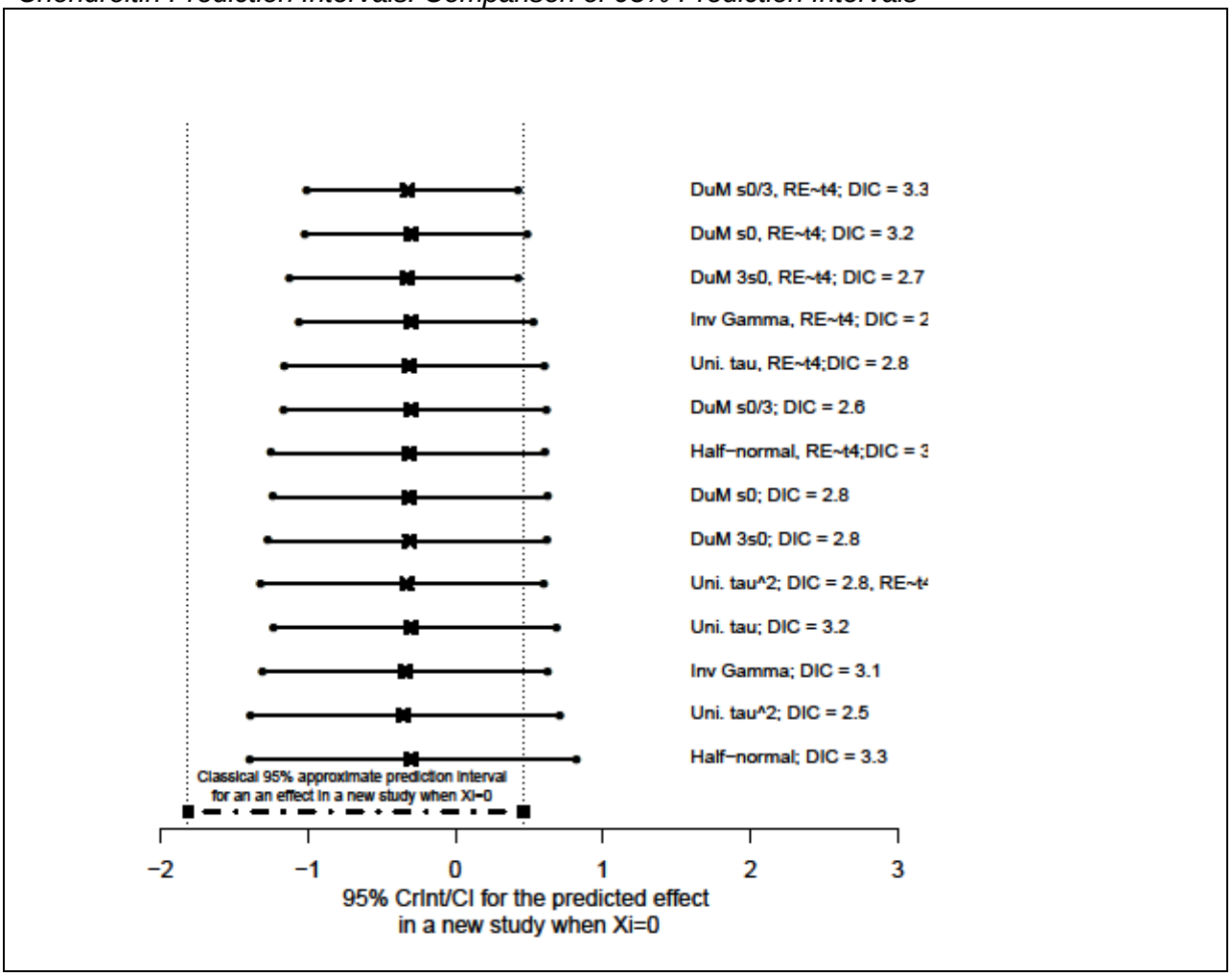
Chondroitin Intercept: Comparison of 95% Credibility/Confidence Intervals k=18



Note. DuM. = DuMouchel. Uni. Tau = Uniform on τ. HN = half-normal. Uni tau^2 = uniform on τ^2. IG = inverse gamma.

Figure 46

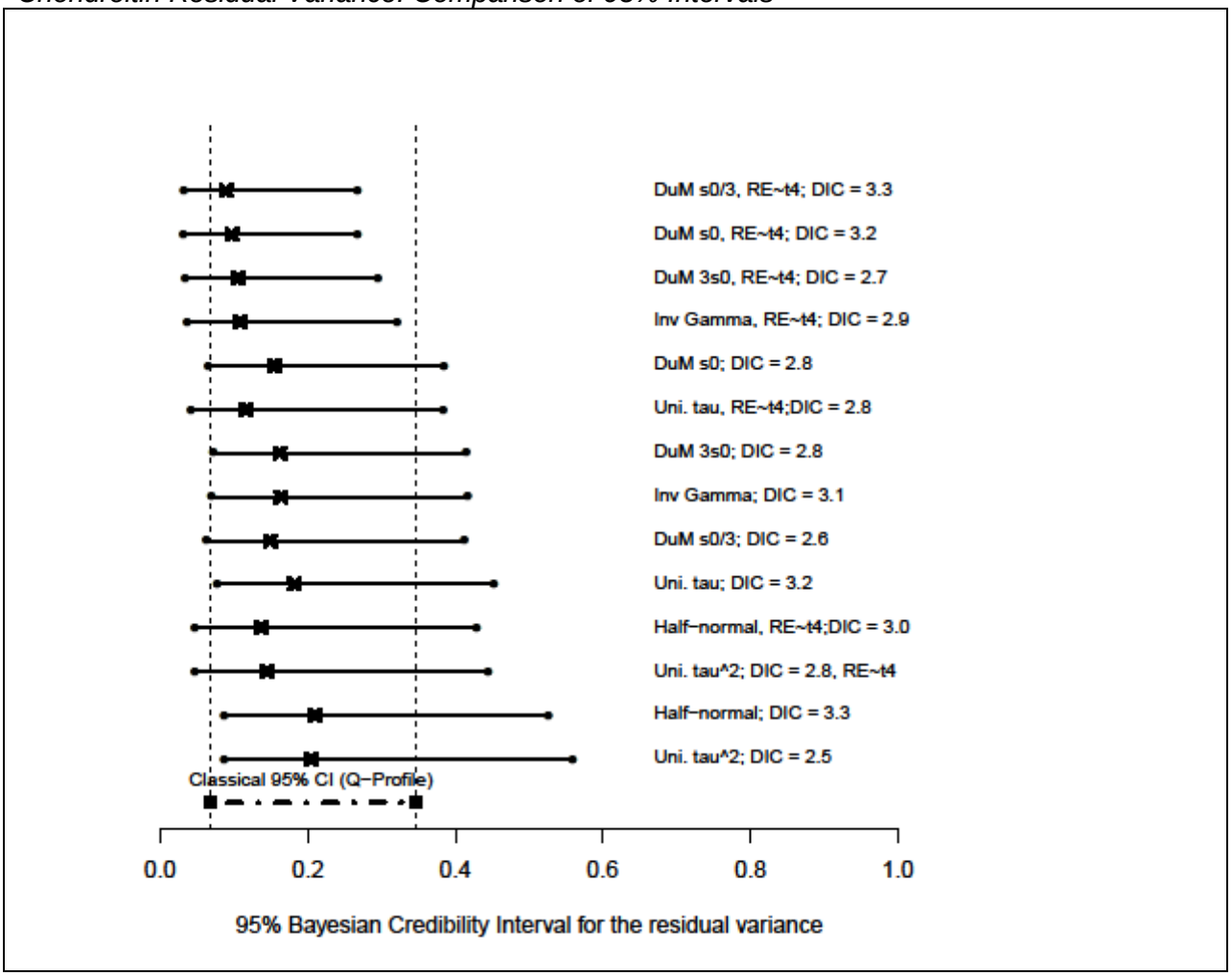
Chondroitin Prediction Intervals: Comparison of 95% Prediction Intervals



Note. DuM. = DuMouchel. Uni. Tau = Uniform on τ. HN = half-normal. Uni tau^2 = uniform on τ^2. IG = inverse gamma.

Figure 47

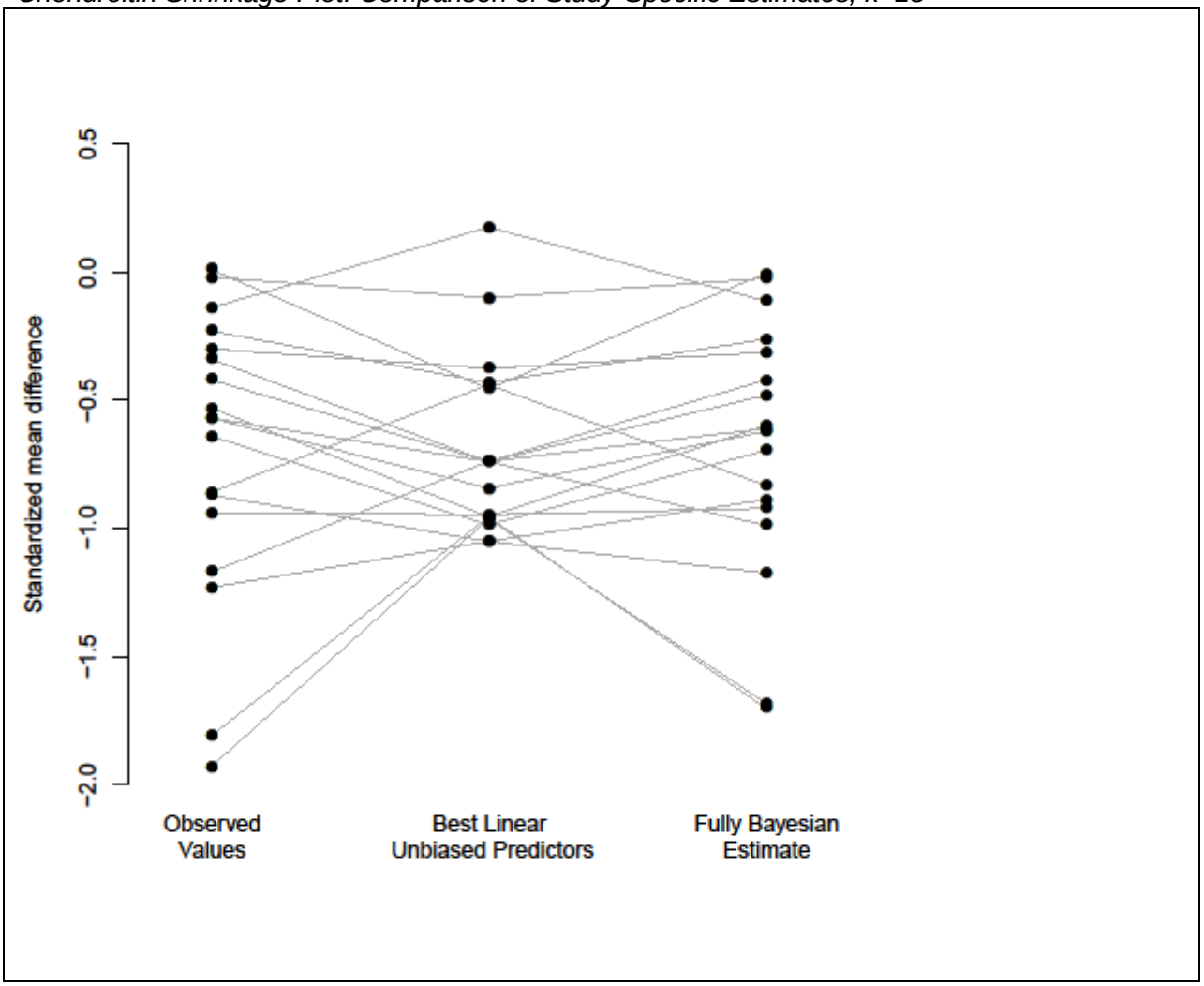
Chondroitin Residual Variance: Comparison of 95% Intervals



Note. DuM. = DuMouchel. Uni. Tau = Uniform on τ. HN = half-normal. Uni tau^2 = uniform on τ^2. IG = inverse gamma.

Figure 48

Chondroitin Shrinkage Plot: Comparison of Study-Specific Estimates, k=18



Data Set 5: Effects of Stress Management Interventions

Classical Meta-analysis Results for SMI

The original SMI meta-analytic results were successfully replicated with the *metafor* package in R 2.11 with the use of a simple RE model ($\hat{d}$ = 0.526, 95% CI 0.364, 0.687, *k* = 55, *p* < .001, I^2=73.4%, τ = 0.501) using D-L estimation, which was the method of estimation used in the original meta-analysis. The test for heterogeneity in the simple RE model was significant at *p* <.001, indicating the presence of significant variation between effect sizes. Similar to Richardson and Rothstein's findings, the sensitivity analysis conducted in *metafor* revealed that the study by Taylor was an outlier as determined by inspection of the forest plot, Q-Q Plot, and leave-one-out case diagnostics. The leave-one-out case diagnostics for the two treatment-control contrasts from the Taylor study revealed r-student values of 4.25 and 0.731, DFFITS values of 1.70 and 0.314, and covariance values of 0.246 and 1.218, respectively, indicating that, when this study is excluded from the model, the resulting model may have better model fit and greater precision. The Taylor study was also excluded from this reanalysis so that the reanalyzed data set remained identical to the data set used in the original meta-analysis.

Recoding of the SMI moderator variables for the mixed-effect model

The coding of the moderator variable *duration* was changed from a categorically coded (i.e., greater than or less than 6 months duration) variable to a continuous variable termed *weeks*, as the numerical value for the number of weeks of intervention duration was available within the SMI data set (the value for *weeks* was missing for four of the effect size data points; however, weeks was not found to be a significant study-level covariate with backward stepwise elimination). For the moderator variable *industry sector*, Richardson and Rothstein (2008) performed subgroup meta-analyses of this moderator for three of the industry categories: (a) Office (*k*=19); (b) Health Care (*k*=8); and (c) Education (*k*=7). For this reanalysis, industry sector was recoded so that information regarding the value of industry sector from the other 26

studies could be included in the model. The additional industry categories that were created were as follows: (a) the *education* category was expanded to included child service workers (*k* = 12); (b) a category termed *Manual and Technical* was created in order to capture the studies that included factory workers, maintenance workers, and army personnel (*k* = 9); and (c) a category termed *Mixed* was created to include the mixed worker studies (*k*=7).

Mixed-Effects Model with REML

In order to investigate the effects of the intervention-level and sample-level covariates, instead of conducting separate subset meta-analyses of the data as Richardson and Rothstein (2008) did, all of the potential moderator variables were entered into a hierarchical linear model, and then the moderators were removed sequentially using a backward-stepwise model selection procedure with REML estimation. The results for the mixed-effects model are displayed in Table 34 and are illustrated in the forest plot shown in Figure 49. The final mixed-effects model with REML revealed statistically significant coefficients for: $\beta_0 = 0.969$, $p < .001$, indicating that on average, there is a benefit of cognitive-behavioral and other types of SMI; for the relaxation interventions $\beta_1 = -0.458$, $p = .018$, resulting in a linear combination estimate for the relaxation interventions of 0.511, p <.001 ; for the multimodal interventions $\beta_2 = -0.674$, $p = .003$, which results in a linear combination estimate for the multimodal interventions of 0.295, $p = .012$; and for the organizational interventions, $\beta_3 = -0.852$, $p = .002$, which results in a linear combination estimate of 0.117, $p = .610$ for organizational interventions. Here, by including these study-level covariates in the model, τ^2, the estimated amount of residual variance equals 0.166, resulting in a 31% reduction in the size of the heterogeneity from the simple RE (REML) model.

In comparing among the models, the AIC and BIC values were at their lowest values for the mixed-effects REML and $REML_L$ models, indicating that these models are the best fitting models. Regarding the sensitivity of the different mixed-effects REML classical meta-analytic

models, the coefficients for relaxation is no longer significant (*p*=.094) when τ^2 was set to the upper bound of its 95% CI ($REML_H$), resulting in a 'modest to severe' impact on meta-analytic conclusions as the fitted model changed. When the Knapp and Hartung (2003) adjustment is applied to the standard errors and test statistics for the regression coefficients, the standard errors are larger and the testing of significance of the regression coefficients is slightly more conservative. For the FE model the estimate for $\hat{\beta}_0$ is likely to be inaccurate and its standard error too small, given the large amount of heterogeneity that exists among the effect sizes.

In terms of publication bias, the funnel plot for the mixed-effects model shows some asymmetry, and Egger's regression test for funnel plot asymmetry using se_i as a predictor was significant at *p* = .00. The Orwin Fail-safe *N* value for a simple RE model was 55, indicating that 55 studies with an average of a null result would have to be added to this data set in order to reduce the unweighted average effect size of 0.585 in half, to a target effect size value of 0.292. For a simple RE model, the trim and fill method estimated that 6 studies were missing on the right side (i.e., in favor of SMI). If these missing studies were imputed, the meta-analytic conclusions would not change, but the effect of SMI would be slightly larger with estimates for the simple RE model with REML of $\hat{\beta}_0 = 0.596$, *p* = <.000, 95% CI (0.441, 0.752), and τ = 0.509.

When interpreting the results of the final meta-analytic model, it is important to consider the recommendations of Hedges and Pigott (2004) regarding power and to note that that most of the statistical tests for the covariates had low power (see Table 36) for detecting a minimal substantive importance difference (for this reanalysis the difference was set equal to 0.33). The statistical tests of the covariates performed in *metafor* may not have been powerful enough to detect a significant moderator effect even if a true moderating difference were to exist among the population of studies. In fact, for the fitted REML mixed-effects model the statistical test for residual heterogeneity was significant (Q =155.4, *p* <.001, τ^2 = 0.166), indicating that there is still remaining unexplained heterogeneity in the residual variance.

Figure 49

SMI: Forest Plot Mixed-effects Model, k = 55

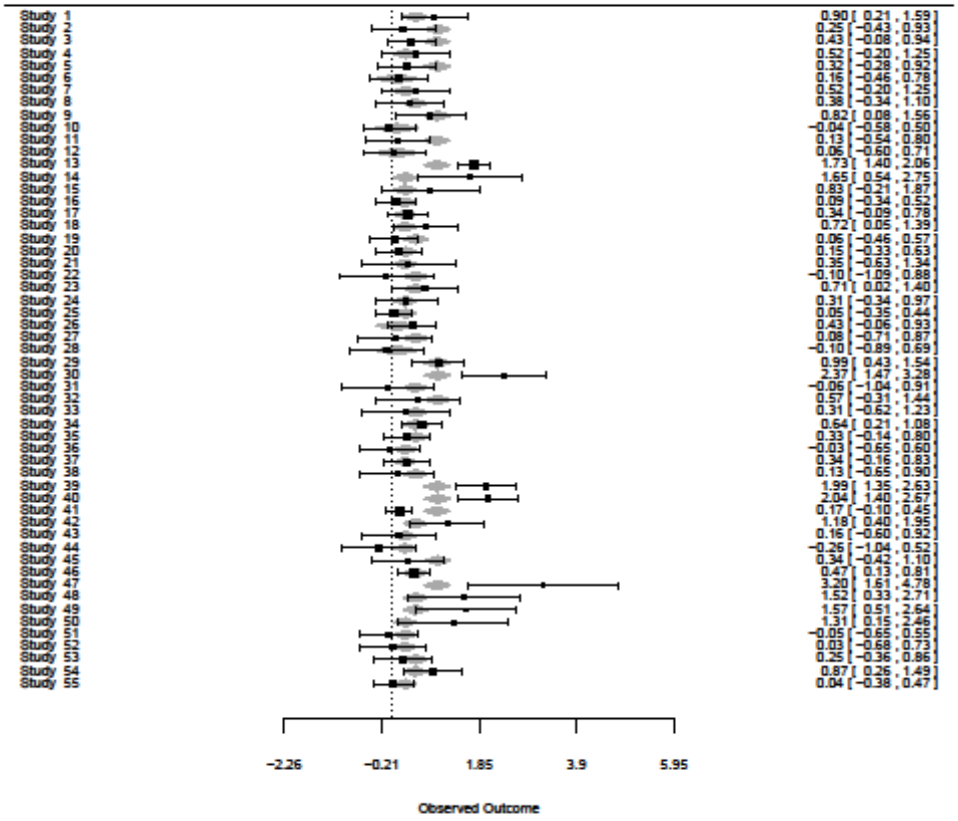


Figure 50

SMI: Normal QQ Plot, Mixed-effects Model, k = 55

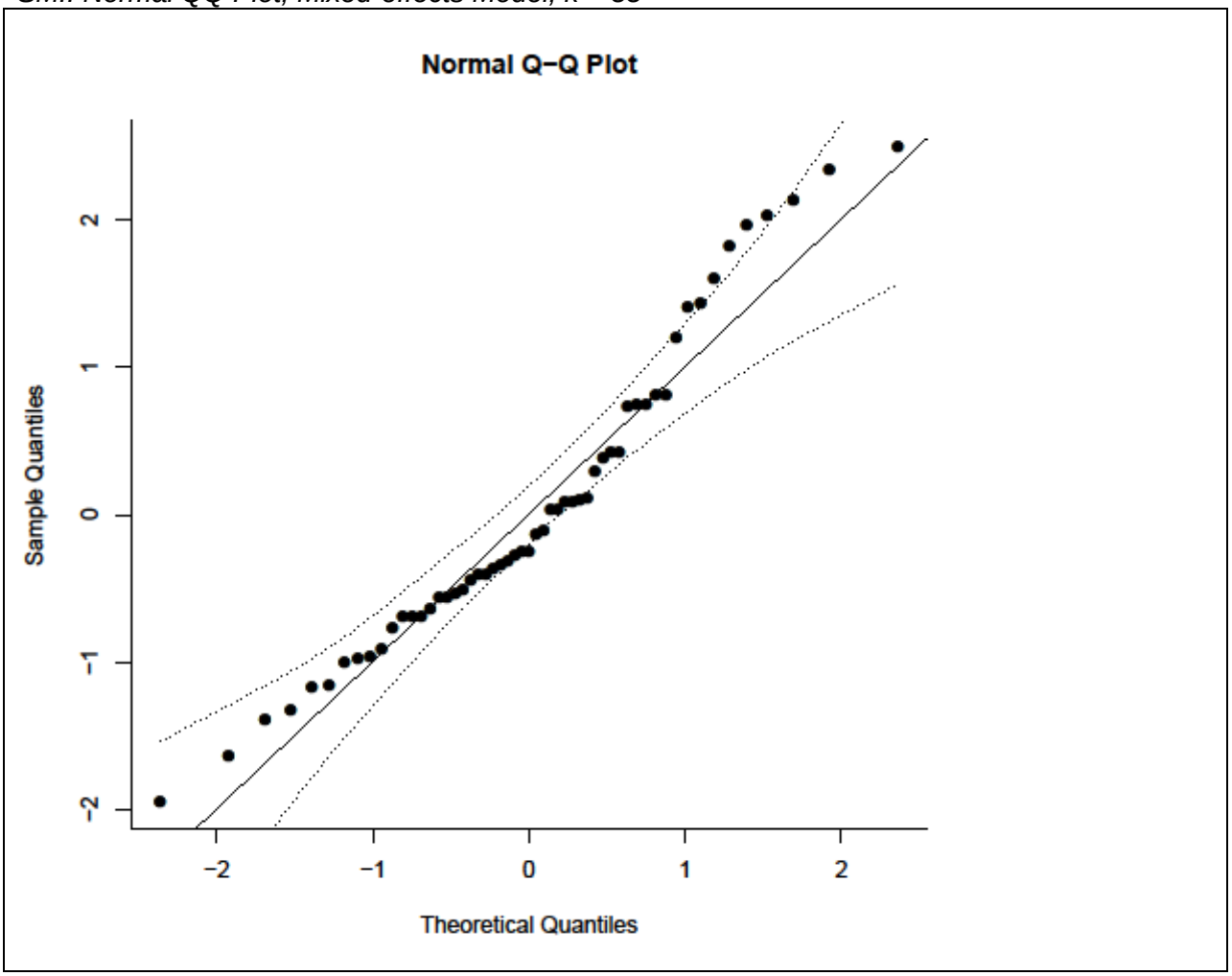


Table 34

SMI: Classical Meta-analysis Results, k = 55

Estimation	FE	D-L	REML	REML	REML (KNHA)	$REML_L$ τ^2= .155	$REML_H$ τ^2 =.521
Model	Simple	Simple	Simple	Mixed	Mixed	Mixed	Mixed
$\hat{B}_0$	0.487	0.526	0.525	0.969	0.969	0.967	1.014
s.e.	0.040	0.083	0.081	0.141	0.149	0.138	0.215
p value	<.001	.001	.001	.001	.001	.001	.001
95% CI low	0.407	0.36,	0.365	0.692	0.670	0.696	0.592
95%CI High	0.566	0.687	0.684	1.246	1.269	1.238	1.436
$\hat{B}_1$ Relaxation				-0.458	-0.458	-0.457	-0.489
s.e.				0.194	0.205	0.190	0.292
p value				.018	.029	.016	.094
95% CI low				-0.838	-0.869	-0.828	-1.062
95% CI high				-0.079	-0.048	-0.085	0.084
$\hat{B}_2$ Multimodal				-0.674	-0.674	-0.674	-0.683
s.e.				0.187	0.197	0.183	0.284
p value				.003	.001	.000	.016
95% CI low				-1.040	-1.070	-1.032	-1.239
95% CI high				-0.308	-0.278	-0.316	-0.127
$\hat{B}_3$ Organiz.				-0.852	-0.852	-0.849	-0.905
s.e.				0.270	0.285	0.264	0.413
p value				.002	.004	.001	.028
95% CI low				-1.381	-1.424	-1.367	-1.715
95% CI high				-0.323	-0.280	-0.331	-0.096
τ		0.501	0.491	0.407	0.407	0.394	0.722
95% CI low		0.394,	0.394	0.315	0.315	0.315	0.394
95% CI high		0.722	0.722	0.638	0.638	0.639	0.722
τ^2		0.2508	0.241	0.166	0.166	0.155	0.521
s.e.			0.0685	0.055	0.055	0.052	0.155
95% CI low		0.1549	0.1549	0.099	0.099	0.099	0.521
95%CI high		0.5208	0.5208	0.408	0.408	0.408	
Q	*Q*=202.8	*Q*=202.8	*Q*=202.8	155.4	155.4	155.4	202.84,
p value	*p*<.001	*p*<.001	*p*<.001	p< .001	p< .001	p< .001	p< .001
PI $\hat{B}_0$ Lower			-0.473,	0.105	0.099,	0.130	-0.921
PI $\hat{B}_0$ Upper			1.522	1.834	1.840	1.803	2.007
PI $\hat{B}_1$ Lower				-1.363	-1.372,	-1.333	
PI $\hat{B}_1$ Upper				0.446	0.456	0.419	
PI $\hat{B}_2$ Lower				-1.573	-1.582	-1.544,	
PI $\hat{B}_2$ Upper				0.224	0.233	0.196	
PI $\hat{B}_3$ Lower				-1.832	-1.850	-1.800	
PI $\hat{B}_3$ Upper				0.128	0.145	0.102	
I^2		73.38%	72.57%				
Q_{MODS}				16.5,	4.955(F)	17.26,	
p value				*p*= .001	*p*= .004	*p* = .001	
Log-likelihood	-92.327	-53.163	-54.729	-49.351	-49.351	-49.370	-58.095
Deviance	184.654	106.325	109.459	98.702	98.702	98.740	116.190
AIC	186.65	110.325	113.459	108.702	108.702	108.740	120.190
BIC	188.661	114.340	117.437	118.361	118.361	118.399	124.169

Min, max $\theta_i^*(\tau)$	.049	.007	.007	.009	-.075
	1.605	1.626	1.626	1.614	1.843

Note. FE= Fixed-effect model. D-L = DerSimonian & Laird estimation. REML = Restricted Maximum Likelihood Estimation. $REML_L = \tau^2$ was specified to the lower bound of its 95% CI from the simple REML model. $REML_H = \tau^2$ was specified to the upper bound of its 95% CI from the simple REML model. $REML_{KH}$ = Knapp and Hartung adjustment made to the standard errors of the regression coefficients. Permutation test *p* values for REML mixed-effects model: $\hat{B}_1, \hat{B}_2, \hat{B}_3$: .026, .001, .001.

Table 35

SMI: Correlation among Variables

	Rel.	Multi.	Other	Parts	Office	HC	Mix	Man.	Org.	y_i	
Rel.[a]	1										
Multi[b]	-0.486	1.000									
Other	-0.255	-0.277	1.000								
Parts	-0.333	0.879	-0.373	1.000							
Office	0.093	-0.045	-0.048	-0.087	1.000						
HC[c]	0.170	-0.191	-0.158	-0.109	-0.300	1.000					
Mix.[d]	-0.137	0.296	0.018	0.293	-0.277	-0.158	1.000				
Man.[e]	-0.083	0.195	0.126	0.168	-0.321	-0.182	-0.169	1.000			
Org.[f]	-0.212	-0.230	-0.121	-0.258	-0.097	0.228	-0.121	-0.140	1.000		
y_i	-0.040	-0.221	0.162	-0.191	0.149	-0.020	-0.193	-0.216	-0.218		1

Note. [a] Rel. = relaxation interventions. [b] Multi. = multimodal interventions. [c] HC = health care workdes. [d] Mix. = mixed workers. [e] Man. = manual/technical workers. [f] Org = organizational interventions.

Table 36

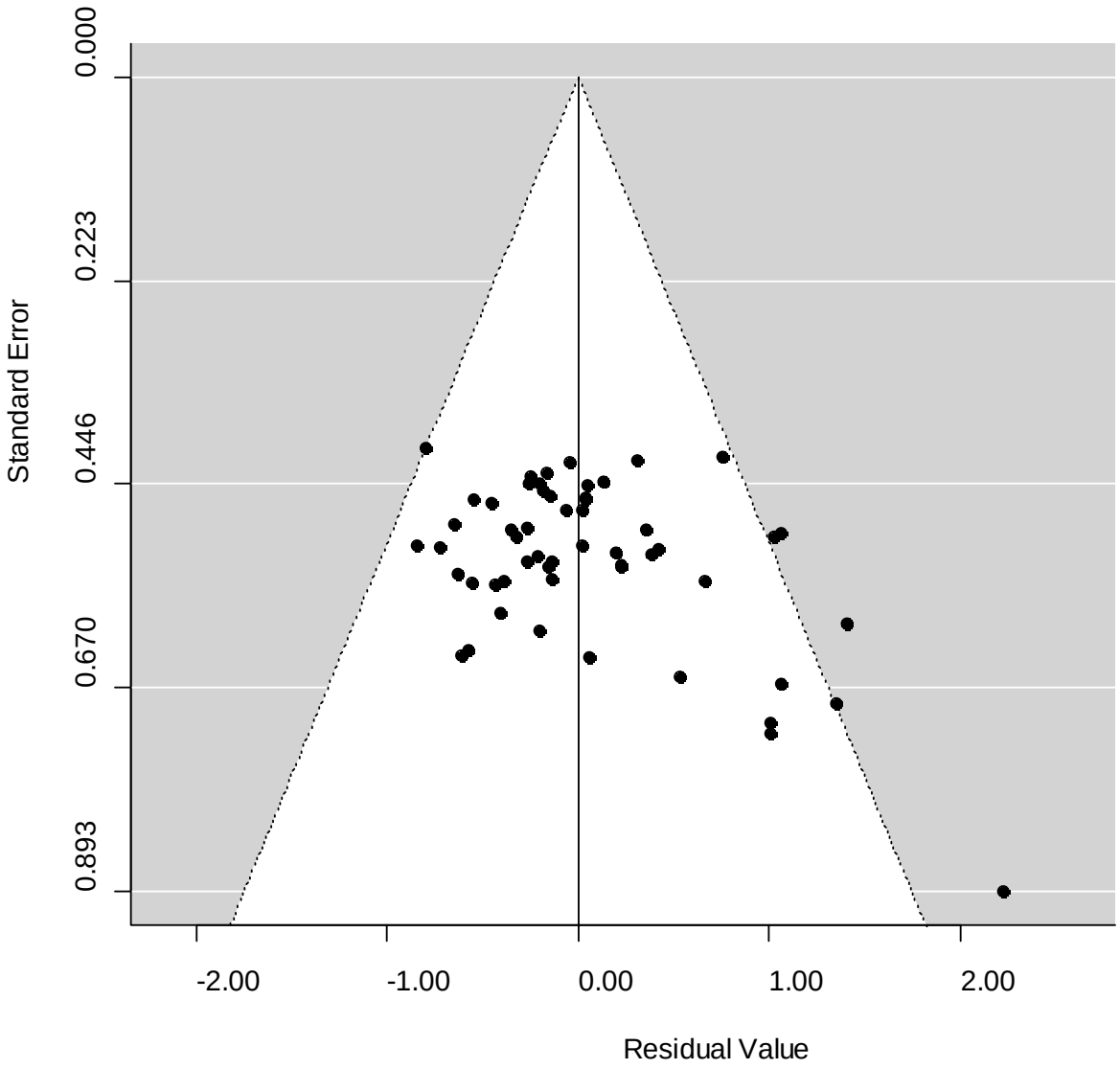
SMI: Power for Regression Coefficients

	Power for Individual Regression Coefficients Backward Stepwise Regression Models							
	Full	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Final
Other	.166							
Parts	.593	.619						
Weeks	.850	.877	.919					
Health	.209	.212	.220	.246				
Office	.256	.261	.268	.337	.448			
Mixed	.171	.180	.185	.220	.254	.310		
Manual	.180	.191	.195	.239	.283	.349	.370	
Org.[a]	.155	.175	.183	.210	.220	.231	.230	.231
Relax	.228	.312	.331	.370	.398	.404	.399	.399
Multi	.126	.133	.321	.363	.372	.383	.415	.424
Intercept	.135	.199	.227	.394	.479	.631	.635	.647

Note. [a] Organ. = organizational interventions. α = .05 with a clinically meaningful effect set equal to 0.33 standard deviations.

Figure 51

SMI: Funnel Plot Mixed-Effects Model



Bayesian Results for SMI

A backward stepwise model selection procedure with a uniform prior distribution for τ (0,5) and the random effect ~ t_4 distribution resulted in a fitted model with a significant intercept and three significant study-level covariates. Here, as shown in the Bayesian Quality Assurance Checklist and the Appendix, the estimates are as follows:

$\hat{\beta}_0 = 0.892$, 95% CI (0.533, 97.50), *prob* $\hat{\beta}_0 > 0 = .000$;

$\hat{\beta}_1 = -0.388$, 95% CI (-0.815, -0.377), *prob* $\hat{\beta}_1 < 0 = .959$;

$\hat{\beta}_2 = -0.619$, 95% CI (-1.055, -0.181), *prob* $\hat{\beta}_2 < 0 = .998$;

$\hat{\beta}_3 = -0.765$, 95% CI (-1.327, -0.245), *prob* $\hat{\beta}_3 < 0 = .994$;

$\tau = 0.323$, 95% CI (0.204, .461), DIC = 76.98.

For a new study using an SMI classified as cognitive-behavioral or other type, the predicted effect in a new study is .883 with a 98.6% probability that the effect is greater than zero and in favor of stress reduction. For relaxation interventions ($\hat{\beta}_0 + \hat{\beta}_1$), the predicted effect for this group in a new study is 0.504 with a 99.9% probability that the effect in a new relaxation study is greater than zero. For multimodal interventions ($\hat{\beta}_0 + \hat{\beta}_2$), the predicted effect for this group in a new study is 0.273 with a 99% probability that the effect in a new study is greater than zero. Finally, for organizational interventions, ($\hat{\beta}_0 + \hat{\beta}_3$), the predicted effect in a new study is 0.126 with an 83% probability that the effect is greater than zero.

The results for the 14 different Bayesian meta-analytic models (e.g., the seven different prior distributions on the heterogeneity variance for which the random effect is modeled from both a normal distribution and a t_4-distribution) are displayed in the Appendix (Bayesian Summary Results). The DIC measures for model fit are smaller (indicating better model fit) and the study-level coefficient, $\hat{\beta}_1$, is no longer significant whenever the δ_i ~ t_4 distribution, resulting in ‘modest to severe’ differences as the fitted model changed.

A trace plot of τ that depicts the dependency of the intercept (curve labeled A) and the study-specific estimates on the different values of τ is displayed in Figure 54 (Bayesian Trace Plot). As shown in the trace plot, the spread of the estimates shrink as τ approaches smaller values, and there is greater spread among the estimates when τ is at larger values. As depicted in Table 37 and Figure 53, Bayesian cross-validation of all of the effect sizes (*k*=57) revealed that the Taylor study, which is comprised of two effect size estimates, resulted in predictive probabilities of .999 and .901. The Bonferroni bound for the probability of the more extreme residual (Taylor a) was .001, indicating that this study differs significantly from the other studies. When this study was removed from the model, as depicted in Figure 53, all of the normalized residuals fall within an expected range.

Figure 52

SMI: Bayesian Quality Assurance Meta-analysis Checklist

Cross-validation ☒ Outlier Detected?: ☐No ☒ Yes

If yes, study name = Taylor Predictive Prob. = 1.000 Bonferroni significance = .000

Fitted model: $X_i\beta = \hat{\beta}_0 + \hat{\beta}_{i1} + \hat{\beta}_{i2} + \hat{\beta}_{i3}$ (with outlier excluded, k = 55)

Prior Distribution for τ:

☒ Uniform on τ ☐ Uniform on τ^2 ☐ Half-norm ☐ Inverse Gamma ☐ DuM. s0 ☐ DuM. s0/3 ☐ DuM. 3s0

Random effect modeled from: Prior Distribution for **β**:

☐ Normal distribution ☒ t_4-distribution ☒ Normal (0,1000)

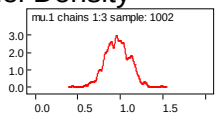
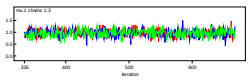
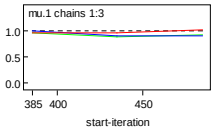
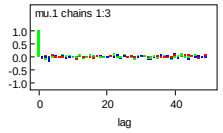
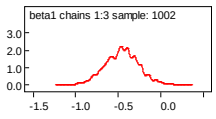
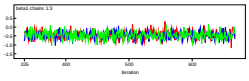
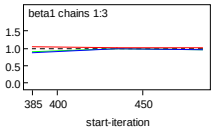
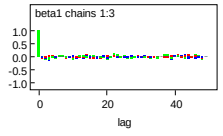
Model specification:

☒ Run 3 chains ☒ 50,000 iterations ☒ Discard first half ☒ DIC = 76.98

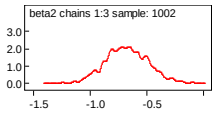
Do meta-analytic inferences change from the fitted hblm?: ☒No ☐ Yes. If yes, impact: _____

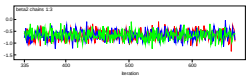
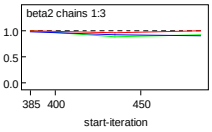
Node	Mean	95% Cred. Interval	Set parameter	MC error < 5% of s.d.	Rhat proximity to 1
β_0	0.892	0.533, 1.255	☒	☒	☒
β_1	-0.388	-0.815, 0.056	☒	☒	☒
β_2	-0.619	-1.055, 0.181	☒	☒	☒
β_3	-0.765	-1.327, -0.245	☒	☒	☒
τ	0.323	0.204, 0.461	☒	☒	☒
$\beta_0 + \beta_1$	0.504	0.254, 0.763	☒	☒	☒
$\beta_0 + \beta_2$	0.273	0.067, 0.499	☒	☒	☒
$\beta_0 + \beta_3$	0.126	-.0291, 0.543	☒	☒	☒
θ_{new}	0.883	0.153, 1.643	☒	☒	☒
$\theta_{new.\beta0 + \beta1}$	0.524	-.0190, 1.185	☒	☒	☒
$\theta_{new.\beta0 + \beta2}$	0.311	-0.454, 0.942	☒	☒	☒
$\theta_{new.\beta0 + \beta3}$	0.124	-0.640, 0.825	☒	☒	☒
$p\ \beta_0 < 0$	.000	0,0	☒	☒	☒
$p\ \beta_1 < 0$	.959	0,1	☒	☒	☒
$p\ \beta_2 < 0$	.998	0,1	☒	☒	☒
$p\ \beta_3 < 0$	.994	0,1	☒	☒	☒
$p\ \beta_0+\beta_1<0$	.000	0,1	☒	☒	☒
$p\ \beta_0+\beta_2<0$	.010	0,1	☒	☒	☒
$p\ \beta_0+\beta_3<0$	.267	0,1	☒	☒	☒
$p\ \theta_{new.\beta0}<0$	.015	0,1	☒	☒	☒
θ_i^*	Min θ_{44}^*= 0.010, Maxθ_{30}^*=1.687		☒	☒	☒

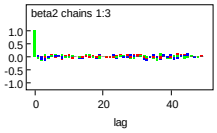
Items pertaining to parameter diagnostic checks:

	Kernel Density	History	BGR	Autocorrelation
☒ β_0				
☒ β_1				

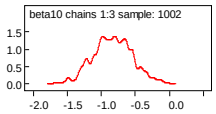
☒ β_2

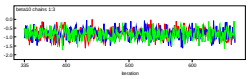
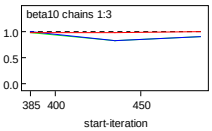


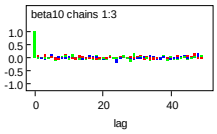





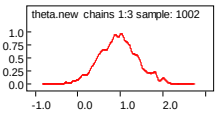
☒ β_3

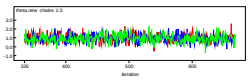
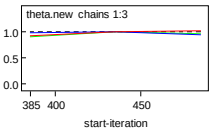


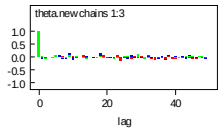





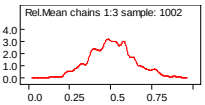
☒ θ_{new}

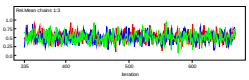
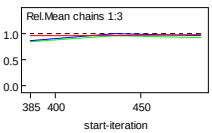


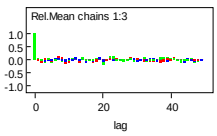





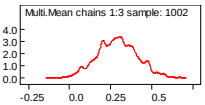
☒ $\beta_0+\beta_1$

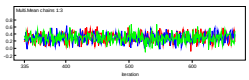
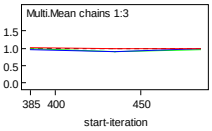


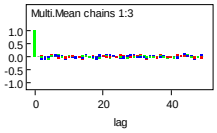





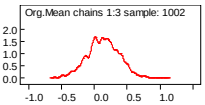
☒ $\beta_0+\beta_2$

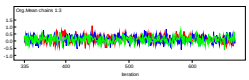
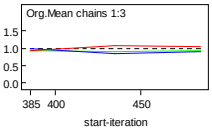


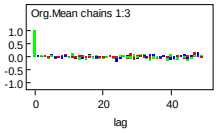





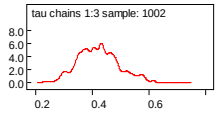
☒ $\beta_0+\beta_3$

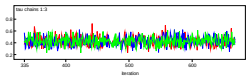
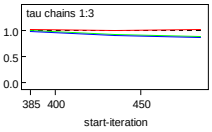


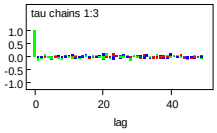





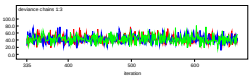
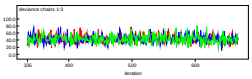
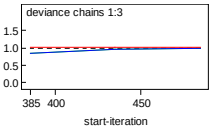
☒ τ





☒ Dev.

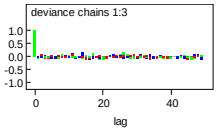


Table 37

SMI: Bayesian Cross-validation Results for the Outlying Study (k = 57)

Study	Y_i	se_i	Prior Mean	Prior s.d.	Pred. Prob	τ^2	Post. Mean	Post. s.d.	Prob > 0
Taylor a	3.988	0.43	1.045	0.459	0.999	0.191	2.939	0.396	1
Taylor b	2.015	0.318	1.165	0.582	0.901	0.310	1.814	0.282	1

Figure 53

SMI: Bayesian Q-Q Plot, k = 57

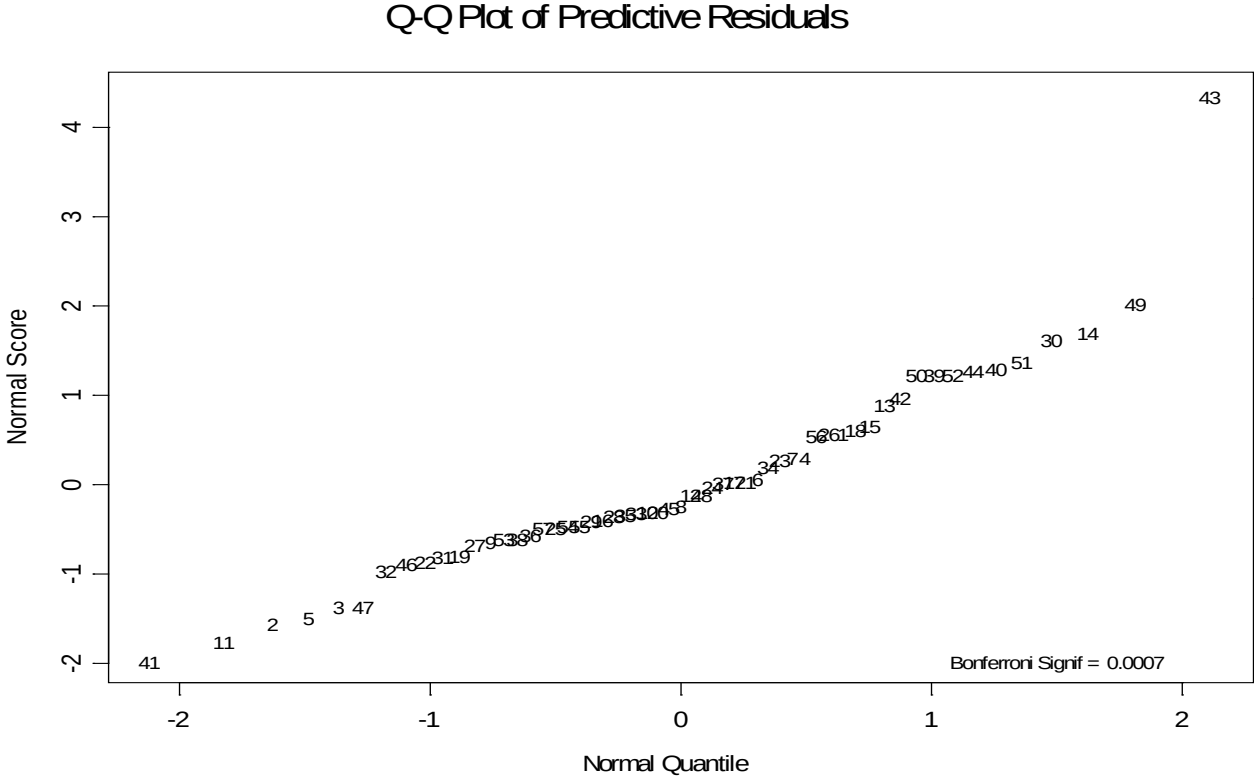


Figure 54

SMI: Bayesian Q-Q Plot, k = 55

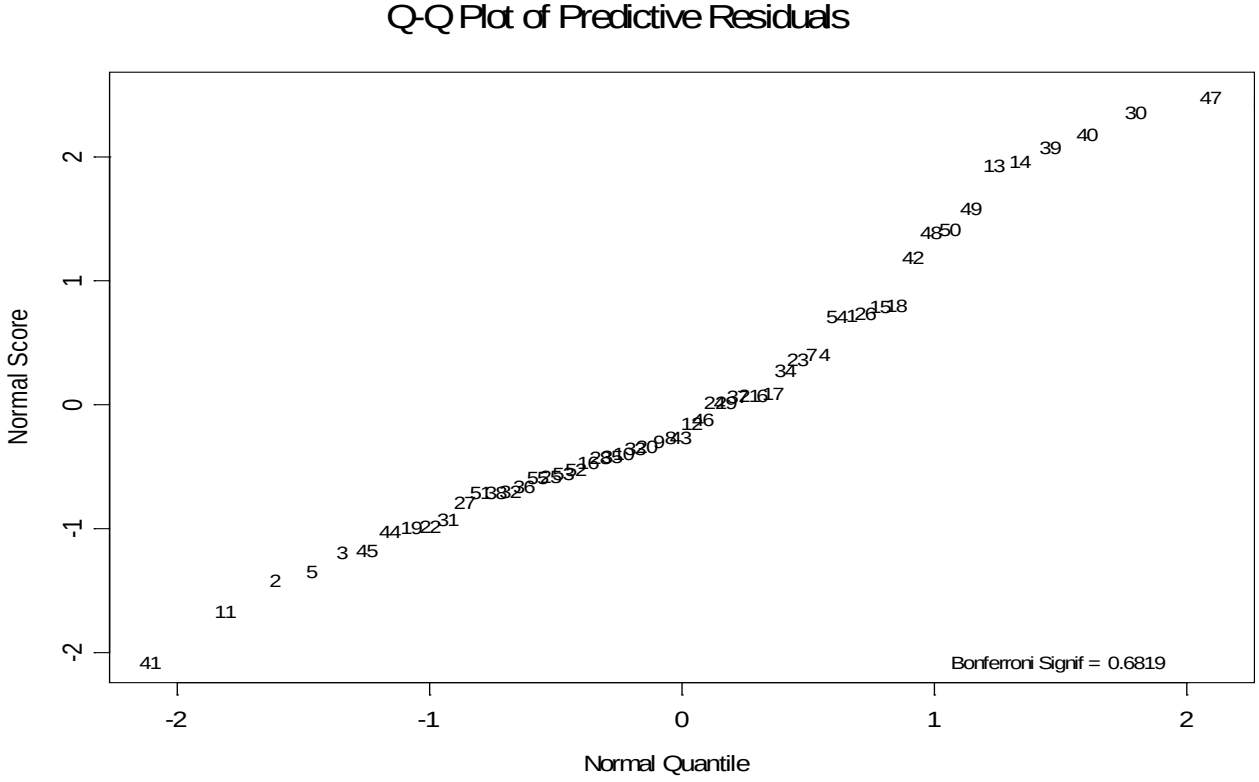
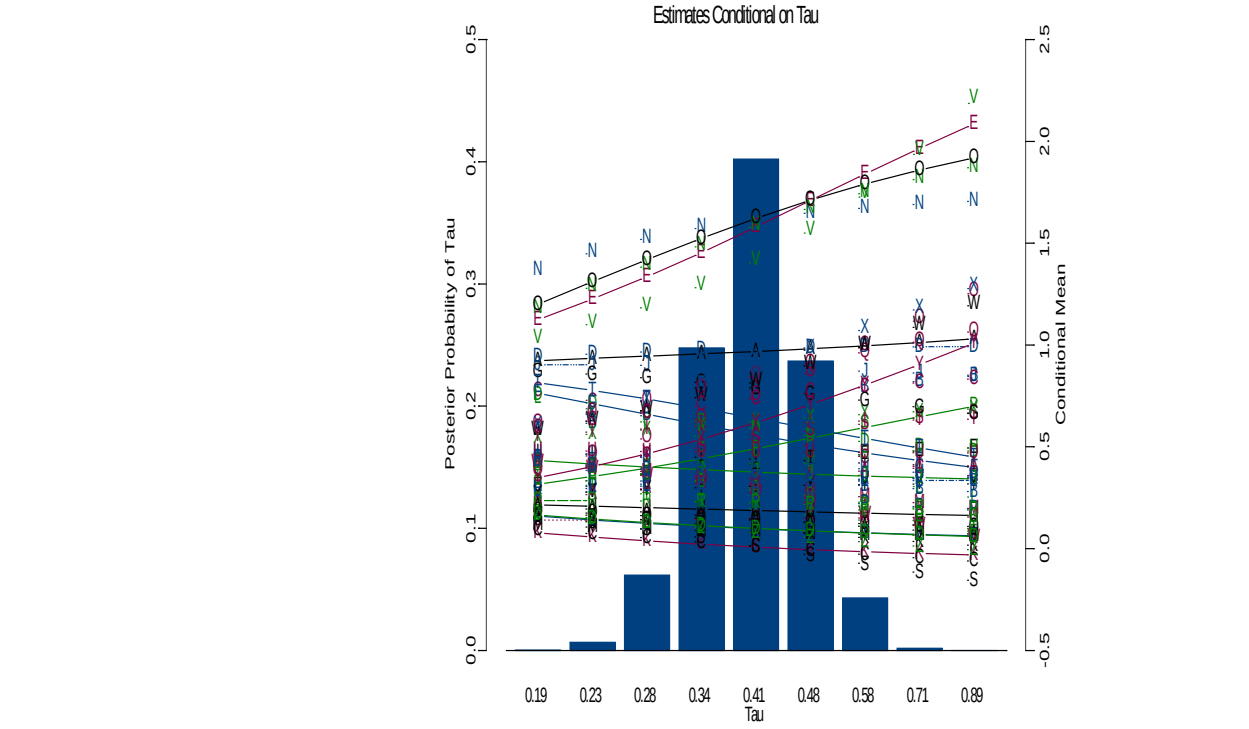


Figure 55

SMI: Trace Plot of Tau for Final Model (k=55)



Comparison of Classical and Bayesian Meta-analytic Results for SMI

As displayed in Figures 56-58, the widths of the Bayesian 95% credibility intervals for the regression coefficients and predicted effects in a new study are all similar to those from the classical REML fitted meta-analytic model. However, the classical REML estimate for the heterogeneity is larger, and the *Q*-profile 95% CI for the heterogeneity is wider.

The classical and Bayesian study-specific estimates in Figure 58 and the study-specific estimates displayed in the Bayesian trace plot (Figure 54) depict how the Bayesian values of θ_i^* estimates offer unique estimates for each observed effect size, because the θ_i^* estimates consider the full uncertainty in all of the auxiliary parameters as compared to the classical study-specific estimates.

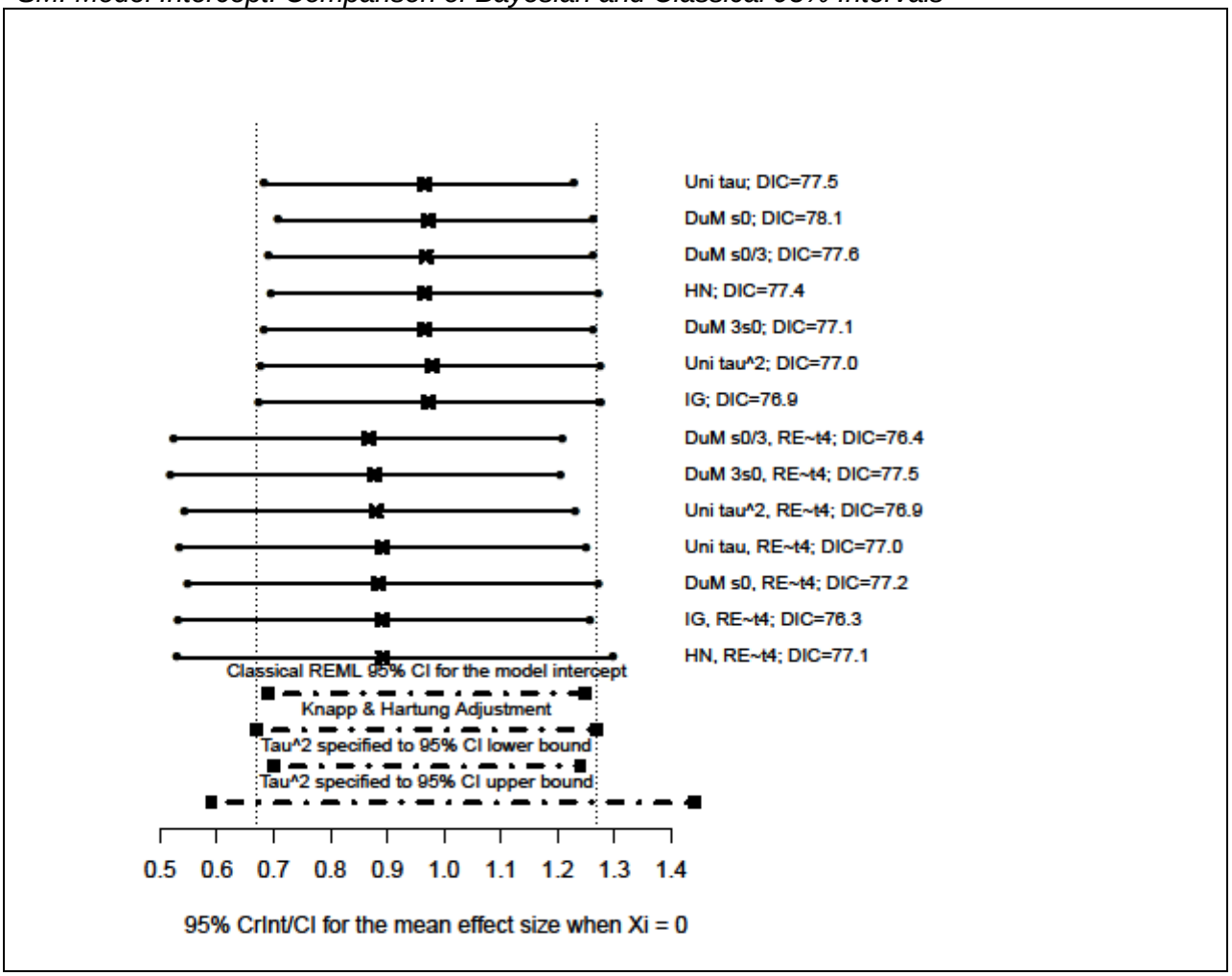
In the original SMI systematic review, a simple RE (D-L) meta-analysis was conducted which revealed a large and significant amount of heterogeneity between the studies. In order to explore the heterogeneity and identify moderator variables, Richardson and Rothstein (2008) performed multiple subgroup analyses of the following study-level characteristics: a) intervention type: cognitive-behavioral, relaxation, organizational, multimodal , and alternative interventions; b) number of intervention components: one, two, three, and four; c) length of intervention: 1-4, 5-8, 9-12, and greater than 12 weeks; and d) length of intervention x type of intervention; e) type of outcome variable: stress, anxiety, mental health, work-related, physiological, productivity, and absenteeism, f) type of outcome variable x type of intervention; g) industry sector: health care, office, and education; and h) industry sector x type of intervention.

The hierarchical linear modeling approach offers a concise, efficient, and unified approach for meta-analytic modeling of the SMI data set rather than conducting multiple univariate moderator tests. The meta-analytic inferences from the classical mixed-effects meta-analytic model with REML and the Bayesian hierarchical linear model differ slightly from those in the original SMI meta-analysis in terms of the estimated effect sizes for the intervention types and identification of the relevant moderating variables. Here, the results indicate in a concise

manner that there is a significant intercept and only three significant study-level coefficients: relaxation, multimodal, and organizational interventions. As shown in the Appendix, from a Bayesian perspective, the probabilities that the effect of an SMI will be beneficial and greater than zero are (a) 99.99% for cognitive-behavioral and other interventions, (b) 99.99% for relaxation interventions, (c) 99% for multimodal interventions, and (d) 73.3% for organizational interventions. Inclusion of the other study-level covariates (i.e., number of components, length of intervention, and type of industry sector) were not found to play a moderating role in the meta-analytic model.

Figure 56

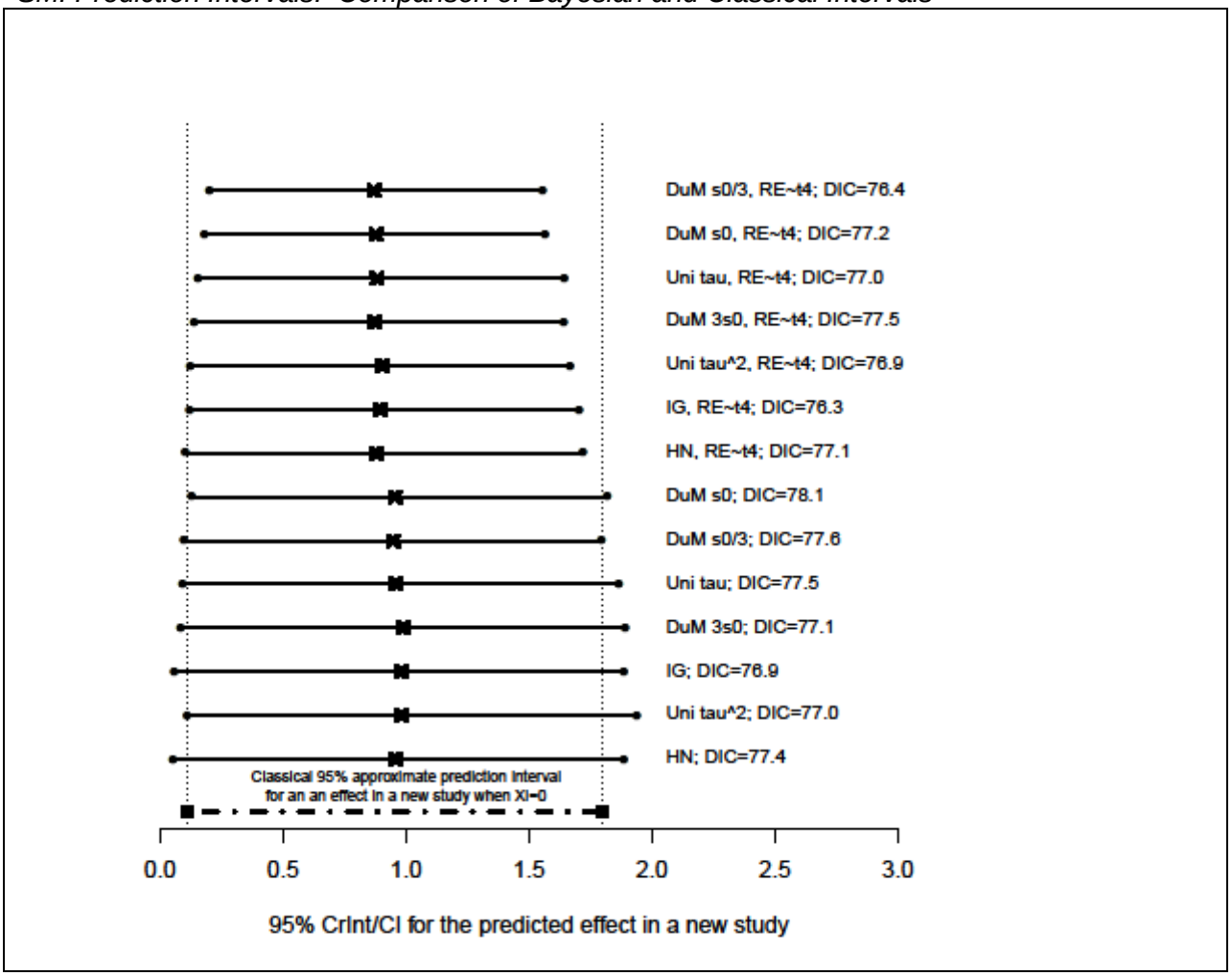
SMI Model Intercept: Comparison of Bayesian and Classical 95% Intervals



Note. DuM = DuMouchel. Uni. Tau = Uniform on τ. HN = half-normal. Uni tau^2 = uniform on τ^2. IG = inverse gamma.

Figure 57

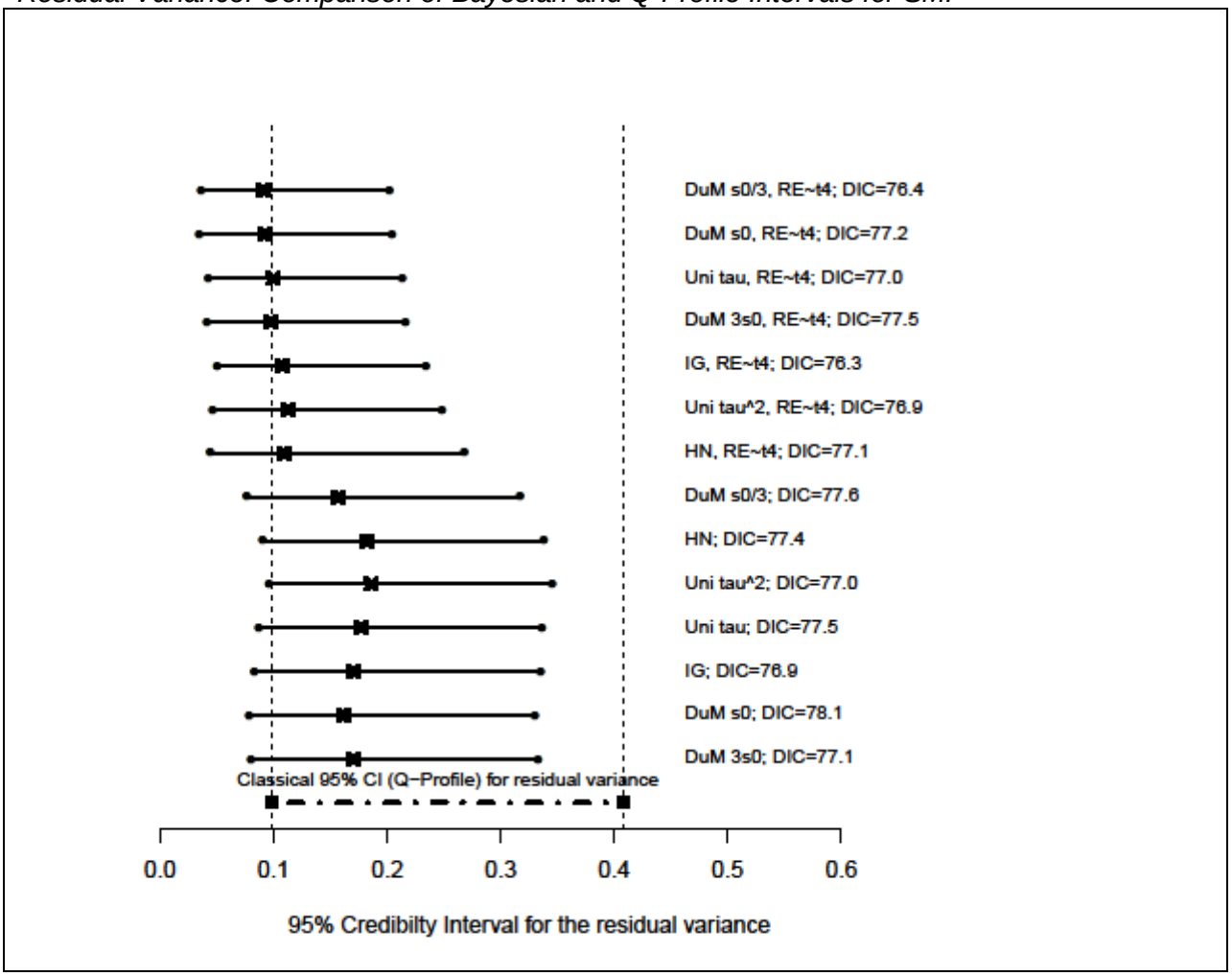
SMI Prediction Intervals: Comparison of Bayesian and Classical Intervals



Note. Du. = DuMouchel. Uni. Tau = Uniform on τ. HN = half-normal. Uni tau^2 = uniform on τ^2. IG = inverse gamma.

Figure 58

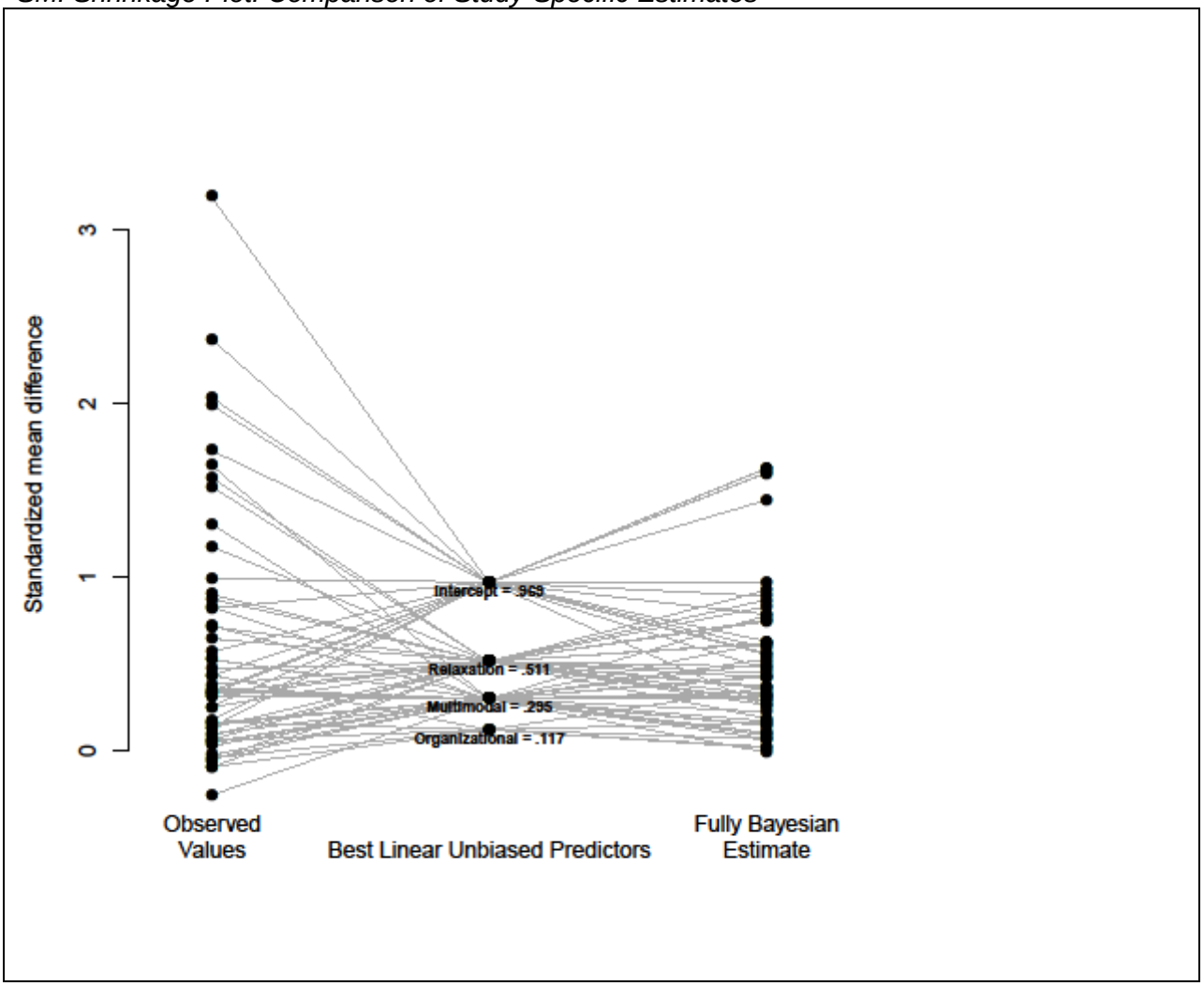
Residual Variance: Comparison of Bayesian and Q-Profile Intervals for SMI



Note. Du. = DuMouchel. Uni. Tau = Uniform on τ. HN = half-normal. Uni tau^2 = uniform on τ^2. IG = inverse gamma.

Figure 59

SMI Shrinkage Plot: Comparison of Study-Specific Estimates



Chapter V: Discussion

This project examined the efficacy of classical versus Bayesian meta-analytic models for addressing important meta-analytic objectives and demonstrated that meta-analytic inferences may change depending upon the type of meta-analytic model used. The Bayesian hierarchical linear models considered in this research project were the best meta-analytic models for addressing all relevant meta-analytic objectives, because Bayesian meta-analytic inferences are based on the joint posterior distribution of *all* of the unknown meta-analytic parameters. For the meta-analytic data sets reanalyzed in this research project, the Bayesian modeling approach with the commonly used, uniform prior distribution on τ; vague and noninformative prior distribution ($\boldsymbol{\beta}$ ~ 0, 1000) on the intercept and regression coefficients**;** and the random effects modeled from either a normal distribution (δ_i~ *Normal*) or from a *t*-distribution with 4 degrees of freedom ($\delta_i \sim t_4$) was the preferred model for addressing all of the relevant meta-analytic objectives, as this model typically yielded the lowest DIC values. Modeling of $\delta_i \sim t_4$ was particularly useful when there was an outlying data point in the data set, because for each of the seven prior distributions for the heterogeneity variance that were examined in the various sensitivity analyses, the estimates for the intercept, regression coefficients, and heterogeneity variance were subject to less influence from the outlier(s) with a Bayesian hierarchical meta-analytic model and $\delta_i \sim t_4$ compared to δ_i~ *Normal*. The *t*-distribution with 4 degrees of freedom proved to be more robust to outliers because the distribution allows for outliers to be more probable. Hence, the between-study variance component estimate was not as large as the estimate produced when the δ_i ~ *Normal*, because the normal distribution was more sensitive to the influence of the outliers.

This research suggests that Bayesian hierarchical linear modeling is the most complete and accurate approach for addressing all relevant meta-analytic objectives. However, it is important to recognize that the application of Bayesian hierarchical linear modeling to meta-analysis requires an adequate understanding of statistical model fitting and computer

programming, and the successful application of Bayesian methods may be too difficult for many non-statistician researchers who conduct meta-analyses. This research shows that the best classical meta-analytic alternative to Bayesian methods is the random/mixed-effects model with the use of: (a) the REML estimator for τ^2; (b) the Knapp and Hartung (2003) adjustment to the standard errors and test statistics, which is an adjustment made to help account for some of the uncertainty in estimating τ^2 (Viechtbauer, 2010); and (c) the Q-profile method (Viechtbauer, 2007) for constructing confidence intervals about τ^2. Of all of the classical meta-analytic models that were systematically examined in this research, the results of the above-mentioned REML random/mixed-effects models were most similar to and produced the most accurate confidence interval coverage probabilities when compared with those of the Bayesian methods. For the classical random/mixed-effects meta-analytic models, this research project conducted sensitivity analyses by specifying τ^2 to the upper and lower boundaries of its 95% CI (i.e., $REML_{Low}$ and $REML_{High}$) in order to determine if the fitted meta-analytic models and inferences changed. For the five meta-analytic examples reanalyzed in this project, whenever a $REML_{High}$ meta-analytic model was fit to the data, the meta-analytic fitted model did change resulting in a 'modest to severe' impact on meta-analytic inference, and the $REML_{High}$ 95% CIs for the model parameters were wider than all of the Bayesian 95% credibility intervals.

When conducting meta-regression with multiple study-level covariates, it is important to give consideration to the potential for an inflated type I error rate (i.e., false-positive results), especially in the presence of heterogeneity (Higgins & Thompson, 2004). The mixed-effects models that used the Knapp and Hartung adjustment (2003) allowed the tests for significance of the regression coefficients to be more conservative (by producing larger *p* values). Within the classical framework, the Knapp and Hartung adjustment may allow for the fitting of a more realistic and parsimonious meta-analytic model. The adjustment typically allows for a reduction in the statistical significance of the model coefficients and a reduction in the probability of type I error rates (Viechtbauer, 2010).

Results from the classical REML random/mixed-effects models that made use of the *Q*-profile method for constructing confidence intervals for τ^2 were similar to the Bayesian 95% credibility intervals for the heterogeneity (residual) variance in which the δ_i ~ *Normal*. However, an important limitation of the classical random/mixed-effects meta-analytic model with REML as compared to the Bayesian model is the inability of the classical model to consider fully the uncertainty in the between-study variance component. The Bayesian approach fully takes into account the uncertainty about all of the auxiliary model parameters. Furthermore, as this project illustrates and as has been shown by other research (e.g., Raudenbush, 2009), another limitation of the classical REML mixed-effects model is that, when the meta-analytic data set includes one or more outliers or when *k* is small, basing estimates on the classical framework assumption that $\tau^2$is equal to its estimated value and that $x_i\beta = x_i\hat{\beta}$ may lead to invalid estimates and confidence intervals. As Raudenbush and Bryk (2002) discuss when the random effects have heavy tails and the hierarchical linear model assumes that δ_i~ *Normal*, the appropriateness of the hypothesis tests and confidence intervals for the fixed effects are affected.

All of the original meta-analyses (with the exception of Raudenbush's analysis of the TEE data set) used either simple FE or simple RE (D-L) models. Thus, due to inherent limitations in the meta-analytic models used, these systematic reviews did not adequately address the important meta-analytic objectives in terms of (a) quantifying heterogeneity, (b) estimating the mean effect sizes and study-specific effects, and (c) predicting an effect in a new study. The validity of future systematic reviews would improve substantially, if meta-analysts gave increased emphasis to the use of random/mixed-effects meta-analytic models with REML or, better yet, to the use of Bayesian hierarchical linear modeling.

A logical continuation of this research project would be to update these original meta-analyses by performing literature searches to identify relevant studies published since the date of the last systematic literature review. Furthermore, what this comparative meta-analysis

project shows and what Viechtbauer (2010) comments on, is a common, though not ideal, practice in meta-analysis: in the presence of heterogeneity, to use simple RE models with subgrouping based upon the levels of study-level characteristics**.** This approach is not optimal because the heterogeneity needs to be estimated separately based on the smaller *k* from the subgroups (Viechtbauer, 2010). To date, many meta-analyses have not used hierarchical linear modeling, so another future research direction would be to reanalyze other important, previously published FE and simple RE meta-analyses using either classical random/mixed-effects (REML) with the Knapp and Hartung adjustment (2003) or, preferably, Bayesian hierarchical linear meta-analytic models, to determine if the meta-analytic inferences change with a more sophisticated modeling approach. DuMouchel and Normand (2000) aptly indicate that Bayesian hierarchical linear modeling methods for meta-analysis are especially suitable for meta-analytic modeling, because such models are able to incorporate FE and random/mixed-effects models into a single Bayesian framework. Given the growing importance of meta-analysis, it is vital that meta-analytic results have the greatest validity: The use of Bayesian hierarchical linear modeling allows for that advantage.

R 2.11.1 with the *metafor* and R2WinBUGS package proved to be excellent resources for conducting meta-analyses and for meeting most of the meta-analytic objectives within the classical (*metafor*) and all of the meta-analytic objectives within the Bayesian (R2WinBUGS) framework. The *metafor* package was very comprehensive in terms of offering functions for analyzing study-level moderators, constructing confidence intervals for the heterogeneity, creating illustrative and well-designed graphical displays, cross-validating data, assessing publication bias, and applying the Knapp and Hartung (2003) adjustment to the standard errors and test statistics of the model coefficients. *Metafor* also provides a number of options for fitting different meta-analytic models as well as options for selecting different estimators for τ^2 (e.g., FE, RE with D-L, REML, maximum-likelihood, and Hunter-Schmidt), although, as this research project emphasizes, the recommended model for meta-analysis

within a classical framework is the random/mixed-effects model with the use of REML estimator for τ^2. In terms of evidence in support of the concurrent validity of the *metafor* package, meta-analytic results obtained in this project via the *metafor* FE and RE(D-L) estimation methods were equivalent to those obtained with the use of the FE and simple RE (D-L) models from the Comprehensive Meta-analysis 2 (CMA 2) software program.

Some suggested functions that could be added to the *metafor* package include functions for computing the Higgins et al. (2009) approximate prediction intervals for a new study and the power of the regression coefficients for the meta-analytic model. Sample R code that was used for these computations can be found in the Appendix. Functions for considering the stochastic dependency among effect sizes are not currently available in *metafor*, however, Viechtbauer (2010) plans to include such functions for handling multivariate analyses and dependent outcome measures in a future version.

Given the complexity and nuances associated with Bayesian model development, the Bayesian Meta-analysis Quality Assurance Checklist (Appendix M) developed for this research project was vital for: (a) ensuring that a robust Bayesian approach was followed, (b) ensuring that the relevant parameters that were necessary for the R program code, (c) reducing errors of forgetting or omitting a relevant aspect of the R program code, and (d) monitoring the items pertaining to Bayesian model convergence. The checklist also provides examples of some relevant parameters that should be monitored in a Bayesian meta-analysis, e.g., the probability that an effect in a new study is greater than 0.

In terms of concurrent validity evidence in support of the R2WinBUGS function, the WinBUGS fitted model results with a DuMouchel prior distribution and the RE ~ normal distribution were identical to those Bayesian results in S-Plus with use of the *hblm* function. An advantage of the *hblm* (DuMouchel, 1995) function with S-Plus is that it produces elegant graphical displays that summarize the essence of a fully Bayesian meta-analysis – a trace plot of τ that depicts the model intercept and study-specific estimates as a function of the posterior

distribution of τ. This graphical display summarizes four out of five of Higgins et al.'s important meta-analytic objectives in a single picture. Meta-analytic trace plots of τ also provide the additional important advantage of representing how meta-analytic estimates change depending upon the different values (e.g., uncertainty) of τ. However, as of the date of this research project, the *hblm* function is only compatible with S-Plus, and it is not compatible with R.

A future application of this research project would be to design the program code in R to create such a meta-analytic Bayesian trace plot with the use of the R2WinBUGS package. In fact, the design of a trace plot should also include an inlaid curve that depicts the estimates for the predicted effect in a new study. Such a trace plot would address the additional important, though underappreciated, meta-analytic objective of illustrating how inferences change depending upon the uncertainty in τ. This trace plot would allow for all five of Higgins et al.'s relevant meta-analytic objectives to be addressed in one informative graphical display, plus the addition of a proposed, sixth important meta-analytic objective: illustrating how meta-analytic inferences change depending upon uncertainty in the estimate of the amount of heterogeneity. Furthermore, in the same vein as Rothstein, Sutton, and Borenstein's (2007) work pertaining to the impact of publication bias, another logical extension of this project would be to evaluate the use and practicality of the classification system (Appendix N) that was developed for this project. This system indicates the impact of the way that meta-analytic fitted models and estimates change, depending upon the uncertainty in the heterogeneity variance with the use of qualitative indicators akin to those used by Rothstein et al. (2007), "minimal", "modest", and "severe".

APPENDIX

APPENDIX A. TEE Meta-analysis: Bayesian summary results.

TEE: Uniform on $\boldsymbol{\tau}$(0,5)

Uniform on τ (0,5)		RE ~ normal distribution			**DIC = -6.47**		
	mean	s.d.	MC err.	2.50%	50%	97.50%	Rhat
β_0	0.412	0.095	.003	0.224	0.408	0.603	1.000
β_1	-0.159	0.039	.001	-0.238	-0.160	-0.079	1.001
τ	0.056	0.045	.002	0.003	0.044	0.171	1.010
$\theta_{.new\ \beta0}$	0.413	0.116	.000	0.188	0.411	0.657	1.001
$\theta_{.new\ \beta1}$	-0.158	0.083	.006	-0.327	-0.159	0.011	1.000
P $\beta_0 < 0$	<.001	0.000	.000	0.000	0.000	0.000	1.000
P $\beta_1 > 0$	<.001	0.000	.000	0.000	0.000	0.000	1.000
P $\theta_{new.\beta0} < 0$	<.001	0.000	.000	0.000	0.000	0.000	1.000
P $\theta_{new.\beta1} > 0$	.033	0.179	.006	0.000	0.000	1.000	1.009
deviance	-10.613	2.802	.090	-16.107	-10.910	-3.661	1.002
Uniform on τ (0,5)		RE ~ t_4-distribution			**DIC= -6.23**		
	mean	s.d.	MC err.	2.5%	50%	97.50%	Rhat
β_0	0.416	0.098	.003	0.211	0.353	0.609	1.006
β_1	-0.161	0.040	.001	-0.240	-0.189	-0.083	1.001
τ	0.043	0.036	.001	0.002	0.015	0.142	1.014
$\theta_{.new\ \beta0}$	0.415	0.112	.004	0.188	0.347	0.639	1.005
$\theta_{.new\ \beta1}$	-0.163	0.073	0.003	-0.317	-0.162	-0.015	1.004
P $\beta_0 < 0$	<.001	0.000	.000	.000	.000	.000	1.000
P $\beta_1 > 0$	<.001	0.032	.001	.000	.000	.000	1.000
P $\theta_{new.\beta0} < 0$	<.001	0.032	.001	.000	.000	.000	1.000
P $\theta_{new.\beta1} > 0$	.017	.129	.004	.000	.000	.000	1.045
deviance	-10.682	2.997	.089	-16.249	-12.338	-3.939	1.002

TEE: DuMouchel

DuMouchel s_0 = .159	RE ~ Normal distribution				**DIC= -7.026**		
	mean	s.d.	MC err.	2.5%	50%	97.50%	Rhat
mu.1	0.411	0.095	.004	0.230	0.411	0.593	1.003
beta	-0.159	0.038	.001	-0.232	-0.160	-0.084	1.001
tau	0.039	0.035	.001	0.001	0.030	0.130	1.013
tausq	0.003	0.005	.000	0.000	0.001	0.017	1.013
theta.new	0.409	0.109	.005	0.190	0.411	0.621	1.005
Pmu<0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Pbeta>0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	<.001	0.000	.000	0.000	0.000	0.000	1.000
deviance	-10.397	2.689	.091	-15.109	-10.840	-4.285	1.003
DuMouchel s_0 = .159	RE ~ t_4-distribution				**DIC= -7.30**		
	mean	s.d.	MC err.	2.5%	50%	97.50%	Rhat
mu.1	0.411	0.091	.003	0.235	0.409	0.593	1.006
beta	-0.160	0.038	.001	-0.234	-0.160	-0.084	1.005
tau	0.034	0.030	.000	0.001	0.027	0.111	1.007
tausq	0.002	0.004	.000	0.000	0.001	0.012	1.007
theta.new	0.410	0.103	.003	0.220	0.406	0.623	1.002
Pmu<0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Pbeta>0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	<.002	0.045	.001	0.000	0.000	0.000	1.292
deviance	-10.705	2.723	.087	-15.940	-11.070	-4.720	1.001

TEE: DuMouchel $s_0/3$

DuMouchel $s_0/3$ = .053	RE ~ Normal				**DIC= -7.31**		
	mean	s.d.	MC err.	2.5%	50%	97.50%	Rhat
mu.1	0.407	0.096	.005	0.221	0.407	0.594	1.024
beta	-0.157	0.039	.002	-0.233	-0.156	-0.081	1.013
tau	0.030	0.030	.001	0.001	0.020	0.106	1.000
tausq	0.002	0.004	.000	0.000	0.000	0.011	1.000
theta.new	0.408	0.109	.006	0.199	0.409	0.614	1.015
Pmu<0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Pbeta>0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	.001	0.032	.001	0.000	0.000	0.000	1.291
deviance	-10.301	2.464	.095	-13.880	-10.840	-4.117	1.003
DuMouchel $s_0/3$ = .053	RE ~ t_4-distribution				**DIC= -7.53**		
	mean	s.d.	MC err.	2.5%	50%	97.50%	Rhat
mu.1	0.411	0.086	.004	0.246	0.404	0.589	1.001
beta	-0.158	0.035	.001	-0.227	-0.157	-0.091	1.001
tau	0.028	0.027	.000	0.001	0.019	0.094	1.001
tausq	0.001	0.003	.000	0.000	0.000	0.009	1.001
theta.new	0.413	0.095	.004	0.229	0.405	0.610	1.003
Pmu<0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Pbeta>0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	<.001	0.000	.000	0.000	0.000	0.000	1.000
deviance	-10.608	2.603	.107	-15.320	-11.040	-4.379	1.002

TEE: DuMouchel 3*s_0

DuMouchel $3s_0$ = .477	RE ~ Normal				**DIC= -6.59**		
	mean	s.d.	MC err.	2.5%	50%	97.50%	Rhat
mu.1	0.415	0.095	.004	0.238	0.414	0.593	1.002
beta	-0.161	0.039	.002	-0.233	-0.161	-0.081	1.003
tau	0.048	0.038	.001	0.002	0.040	0.149	1.001
tausq	0.004	0.006	.000	0.000	0.002	0.022	1.001
theta.new	0.416	0.112	.005	0.199	0.412	0.647	1.005
Pmu<0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Pbeta>0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	<.001	0.000	.000	0.000	0.000	0.000	1.000
deviance	-10.399	2.807	.095	-15.467	-10.750	-4.307	1.008
DuMouchel $3s_0$ = .477	RE ~ t_4-distribution				**DIC= -6.57**		
	mean	s.d.	MC err.	2.5%	50%	97.50%	Rhat
mu.1	0.417	0.098	.005	0.236	0.417	0.616	1.009
beta	-0.161	0.040	.002	-0.238	-0.160	-0.086	1.009
tau	0.041	0.037	.001	0.001	0.031	0.134	1.004
tausq	0.003	0.006	.000	0.000	0.001	0.018	1.004
theta.new	0.418	0.111	.005	0.210	0.417	0.648	1.012
Pmu<0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Pbeta>0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	<.001	0.000	.000	0.000	0.000	0.000	1.000
deviance	-10.582	2.954	.114	-16.053	-10.935	-3.245	1.003

TEE: Gamma on Precision

$1/\tau^2$ ~ Gamma (.01, .01)	RE ~ Normal				**DIC = -4.27**		
	mean	s.d.	MC err.	2.5%	50%	97.50%	Rhat
mu.1	0.421	0.100	.003	0.222	0.421	0.624	1.005
beta	-0.164	0.043	.001	-0.249	-0.164	-0.085	1.003
tau	0.108	0.039	.001	0.054	0.100	0.197	1.000
tausq	0.013	0.010	.000	0.003	0.010	0.039	1.000
theta.new	0.420	0.153	.005	0.110	0.418	0.726	1.000
Pmu<0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Pbeta>0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	.004	0.063	.002	0.000	0.000	0.000	1.105
deviance	-11.042	3.706	.086	-17.909	-11.205	-2.923	1.002
$1/\tau^2$~ Gamma (.01, .01)	RE ~ t_4-distribution				**DIC = -4.07**		
	mean	s.d.	MC err.	2.5%	50%	97.50%	Rhat
mu.1	0.424	0.112	.004	0.206	0.421	0.645	1.004
beta	-0.165	0.046	.006	-0.257	-0.165	-0.075	1.000
tau	0.102	0.037	.001	0.051	0.097	0.188	1.001
tausq	0.012	0.010	.000	0.003	0.009	0.035	1.001
theta.new	0.418	0.153	.005	0.115	0.415	0.708	1.000
Pmu<0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Pbeta>0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	.008	0.089	.013	0.000	0.000	0.000	1.050
deviance	-11.619	4.179	.140	-18.669	-11.910	-1.873	1.004

TEE: Half-normal

Half-normal		RE ~ Normal			**DIC = -5.12**		
	mean	s.d.	MC err.	2.5%	50%	97.50%	Rhat
mu.1	0.419	0.100	.004	0.238	0.414	0.628	1.008
beta	-0.162	0.041	.002	-0.243	-0.161	-0.085	1.006
tau	0.084	0.057	.004	0.005	0.075	0.224	1.191
tausq	0.010	0.014	.000	0.000	0.006	0.050	1.191
theta.new	0.421	0.144	.006	0.136	0.416	0.727	1.016
Pmu<0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Pbeta>0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	.005	0.070	.002	0.000	0.000	0.000	1.018
deviance	-10.790	3.318	.105	-16.809	-11.080	-3.560	1.003
Half-normal		RE ~ t_4-Distribution			**DIC = -5.05**		
	mean	s.d.	MC err.	2.5%	50%	97.50%	Rhat
mu.1	0.418	0.105	.003	0.221	0.417	0.621	1.002
beta	-0.163	0.045	.001	-0.247	-0.161	-0.074	1.001
tau	0.077	0.049	.003	0.012	0.068	0.189	1.017
tausq	0.008	0.012	.000	0.000	0.005	0.036	1.017
theta.new	0.422	0.142	.005	0.140	0.419	0.735	1.004
Pmu<0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Pbeta>0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	.003	0.055	.002	0.000	0.000	0.000	0.999
deviance	-11.109	3.696	.127	-18.218	-11.250	-2.583	1.001

TEE: Uniform on $\boldsymbol{\tau^2}$

Uniform on τ^2 (0,1000)	RE ~ Normal				**DIC = -4.85**		
	mean	s.d.	MC err.	2.5%	50%	97.50%	Rhat
mu.1	0.422	0.103	.003	0.230	0.419	0.627	1.000
beta	-0.163	0.043	.001	-0.251	-0.162	-0.078	1.000
tau	0.093	0.056	.002	0.012	0.082	0.224	1.003
tausq	0.012	0.015	.000	0.000	0.007	0.050	1.003
theta.new	0.420	0.153	.004	0.121	0.409	0.728	1.001
Pmu<0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Pbeta>0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	.009	0.094	.003	0.000	0.000	0.000	1.182
deviance	-10.937	3.569	.120	-17.909	-11.140	-3.865	1.001
Uniform on τ^2 (0,1000)	RE ~ t_4-Distribution				**DIC = -4.99**		
	mean	s.d.	MC err.	2.5%	50%	97.50%	Rhat
mu.1	0.417	0.099	.003	0.224	0.414	0.618	1.004
beta	-0.161	0.042	.001	-0.245	-0.159	-0.079	1.004
tau	0.078	0.049	.002	0.012	0.068	0.193	1.034
tausq	0.008	0.011	.000	0.000	0.005	0.037	1.034
theta.new	0.413	0.139	.004	0.154	0.410	0.702	1.003
Pmu<0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Pbeta>0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	.003	0.055	.002	0.000	0.000	0.000	1.133
deviance	-11.083	3.474	.118	-17.655	-11.210	-3.484	1.004

Appendix B. SMI Meta-analysis: Bayesian summary results.

SMI: Uniform on $\boldsymbol{\tau}$ (0,5)

Uniform on τ (0,5)			RE ~ normal distribution				**DIC = 77.542**
	mean	s.d.	MC error	2.5%	50%	97.5%	Rhat
mu.1	0.966	0.144	.004	0.682	0.962	1.229	1.001
θ_{new}	0.959	0.455	.015	0.090	0.957	1.864	1.004
β_1	-0.456	0.203	.005	-0.857	-0.455	-0.050	1.002
β_2	-0.675	0.190	.006	-1.025	-0.678	-0.288	1.000
β_3	-0.830	0.289	.010	-1.389	-0.830	-0.246	1.005
P $\beta_1 > 0$	.017	0.129	.003	1.000	1.000	1.000	1.026
P $\beta_2 > 0$	<.001	0.000	.000	1.000	1.000	1.000	1.000
P $\beta_3 > 0$	.001	0.032	.001	1.000	1.000	1.000	1.291
P mu.1 < 0	<.001	0.000	.000	0.000	0.000	0.000	1.000
P $\theta_{new} < 0$	.015	0.121	.003	0.000	0.000	0.000	1.019
$\beta_0 + \beta_1$	0.510	0.138	.004	0.238	0.507	0.790	1.003
$\beta_0 + \beta_2$	0.291	0.123	.004	0.057	0.291	0.536	1.001
$\beta_0 + \beta_3$	0.136	0.244	.009	-0.355	0.135	0.603	1.008
P $\beta_0 + \beta_1 < 0$	<.001	0.000	.000	0.000	0.000	0.000	1.000
P $\beta_0 + \beta_2 < 0$	.008	0.089	.003	0.000	0.000	0.000	1.050
P $\beta_0 + \beta_3 < 0$	.283	0.451	.016	0.000	0.000	1.000	1.007
$\theta_{new}.\beta_0 + \beta_1$	0.524	0.452	.012	-0.368	0.519	1.429	1.001
$\theta_{new}.\beta_0 + \beta_2$	0.311	0.441	.014	-0.529	0.287	1.195	1.002
$\theta_{new}.\beta_0 + \beta_3$	0.124	0.486	.012	-0.881	0.125	1.065	1.000
P $\theta_{new}.\beta_0 + \beta_1 < 0$	.110	0.313	.001	0.000	0.000	1.000	1.000
P $\theta_{new}.\beta 0 + \beta 2 < 0$	.241	0.428	.002	0.000	0.000	1.000	1.000
P $\theta_{new}.\beta_0 + \beta_3 < 0$	.387	0.487	.0138	0.000	0.000	1.000	1.000
τ	0.424	0.074	.002	0.294	0.421	0.580	1.000
τ^2	0.185	0.065	.002	0.087	0.177	0.336	1.000
deviance	42.899	10.619	.351	23.915	42.235	66.104	1.000
Uniform on τ (0,5)			RE ~ t_4-distribution				**DIC= 76.978**
	mean	s.d.	MC error	2.5%	50%	97.50%	Rhat
mu.1	0.892	0.184	.006	0.533	0.891	1.250	1.008
θ_{new}	0.883	0.374	.013	0.153	0.893	1.643	1.003
β_1	-0.388	0.220	.007	-0.815	-0.377	0.056	1.003
β_2	-0.619	0.218	.007	-1.055	-0.621	-0.181	1.003
β_3	-0.765	0.273	.008	-1.327	-0.761	-0.245	1.000
P $\beta_1 > 0$	.041	0.198	.006	0.000	1.000	1.000	1.012
P $\beta_2 > 0$	.002	0.045	.001	1.000	1.000	1.000	1.104
P $\beta_3 > 0$	.006	0.077	.002	1.000	1.000	1.000	1.038
P mu.1 < 0	<.001	0.000	.000	0.000	0.000	0.000	1.000
P $\theta_{new} < 0$	.014	0.117	.004	0.000	0.000	0.000	1.009
$\beta_0 + \beta_1$	0.504	0.128	.003	0.254	0.505	0.763	1.001
P $\beta_0 + \beta_1 < 0$	<.001	0.000	.000	0.000	0.000	0.000	1.000
$\beta_0 + \beta_2$	0.273	0.114	.004	0.067	0.271	0.499	1.000
P $\beta_0 + \beta_2 < 0$	.010	0.099	.003	0.000	0.000	0.000	1.018
$\beta_0 + \beta_3$	0.126	0.209	.006	-0.291	0.124	0.543	1.002
P $\beta_0 + \beta_3 < 0$	.267	0.443	.013	0.000	0.000	1.000	1.003
$\theta_{new}.\beta_0 + \beta_1$	0.498	0.355	.011	-0.190	0.503	1.185	1.000

$\theta_{new.}\beta_0 + \beta_2$	0.252	0.352	.012	-0.454	0.259	0.942	1.000
$\theta_{new.}\beta_0 + \beta_3$	0.121	0.392	.012	-0.640	0.117	0.825	1.003
P $\theta_{new.}\beta_0 + \beta_1 < 0$	.081	.273	.007	0.000	0.000	1.000	1.003
P $\theta_{new.}\beta 0 + \beta 2 < 0$	.230	.421	.014	0.000	0.000	1.000	1.003
P $\theta_{new.}\beta_0 + \beta_3 < 0$	.362	.481	.014	0.000	0.000	1.000	1.000
τ	0.323	0.067	.002	0.204	0.317	0.461	1.000
τ^2	0.109	0.046	.001	0.042	0.100	0.213	1.000
deviance	43.49	10.545	.359	23.163	43.150	63.567	1.004

SMI: DuMouchel s_0

DuMouchel s_0 = .300			RE ~ normal distribution				**DIC = 78.105**
	mean	s.d.	MC err.	2.50%	50%	97.50%	Rhat
mu.1	0.972	0.145	.005	0.707	0.968	1.262	1.006
theta.new	0.954	0.421	.011	0.127	0.935	1.817	1.000
β_1	-0.464	0.196	.007	-0.862	-0.460	-0.092	1.002
β_2	-0.672	0.192	.006	-1.038	-0.670	-0.305	1.003
β_3	-0.851	0.274	.008	-1.353	-0.849	-0.297	1.000
P $\beta_1 < 0$	.991	0.094	.003	1.000	1.000	1.000	1.067
P $\beta_2<0$	.999	0.032	.001	1.000	1.000	1.000	1.291
P $\beta_3<0$	.997	0.055	.002	1.000	1.000	1.000	0.999
P mu.1 < 0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	.012	0.109	.003	0.000	0.000	0.000	1.066
τ	0.409	0.073	.002	0.280	0.401	0.574	1.000
τ^2	0.172	0.063	.002	0.078	0.161	0.330	1.000
deviance	43.935	11.032	.372	23.077	43.445	68.259	1.000
DuMouchel s_0 = .300			RE ~ t_4-distribution				**DIC= 77.164**
	mean	s.d.	MC err.	2.5%	50%	97.50%	Rhat
mu.1	0.885	0.182	.005	0.548	0.881	1.271	1.000
theta.new	0.877	0.348	.011	0.178	0.875	1.565	1.002
β_1	-0.380	0.220	.007	-0.802	-0.385	0.032	1.000
β_2	-0.614	0.209	.006	-1.042	-0.610	-0.218	1.000
β_3	-0.771	0.277	.009	-1.309	-0.773	-0.207	1.002
P $\beta_1<0$	.961	0.194	.000	0.000	1.000	1.000	1.008
P $\beta_2<0$	.999	0.032	.002	1.000	1.000	1.000	1.291
P $\beta_3<0$	.997	0.055	.000	1.000	1.000	1.000	1.133
P mu.1 < 0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	.003	0.055	.002	0.000	0.000	0.000	1.133
τ	0.306	0.065	.002	0.185	0.303	0.452	1.002
τ^2	0.098	0.042	.001	0.034	0.092	0.204	1.002
deviance	43.904	11.225	/.321	24.052	42.920	68.202	1.002

SMI: DuMouchel $s_0/3$

DuMouchel $s_0/3$= .100		RE ~ normal distribution				**DIC = 77.643**
	mean	s.d.	MC err.	2.50%	50%	97.50%	Rhat
mu.1	0.968	0.145	.004	0.690	0.967	1.262	1.000
theta.new	0.948	0.435	.013	0.095	0.968	1.793	1.000
β_1	-0.460	0.200	.006	-0.853	-0.459	-0.077	1.000
β_2	-0.667	0.187	.006	-1.055	-0.665	-0.307	1.000
β_3	-0.838	0.283	.009	-1.401	-0.836	-0.274	1.005
P $\beta_1 < 0$	.989	0.104	.003	1.000	1.000	1.000	1.027
P $\beta_2<0$	.999	0.000	.000	1.000	1.000	1.000	1.000
P $\beta_3<0$	.997	0.055	.002	1.000	1.000	1.000	0.999
P mu.1 < 0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	.017	0.129	.004	0.000	0.000	0.000	1.011
τ	0.402	0.072	.002	0.276	0.396	0.563	1.006
τ^2	0.167	0.061	.002	0.076	0.157	0.317	1.006
deviance	44.052	11.004	.382	25.263	43.325	68.218	1.006
DuMouchel $s_0/3$= .100		RE ~ t_4-distribution				**DIC = 76.336**
	mean	s.d.	MC err.	2.5%	50%	97.50%	Rhat
mu.1	0.867	0.174	.006	0.523	0.867	1.208	1.003
theta.new	0.870	0.346	.006	0.201	0.843	1.553	1.001
β_1	-0.373	0.215	.006	-0.792	-0.369	0.047	1.002
β_2	-0.597	0.205	.009	-0.976	-0.596	-0.178	1.001
β_3	-0.759	0.277	.006	-1.307	-0.753	-0.215	1.000
P $\beta_1 < 0$	.962	0.191	.002	0.000	1.000	1.000	1.021
P $\beta_2<0$	.997	0.055	.002	1.000	1.000	1.000	0.999
P $\beta_3<0$	.996	0.063	.002	1.000	1.000	1.000	1.166
P mu.1 < 0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	.007	0.083	.003	0.000	0.000	0.000	1.037
τ	0.305	0.067	.002	0.189	0.302	0.449	1.004
τ^2	0.097	0.044	.002	0.036	0.091	0.202	1.004
deviance	43.559	11.114	.362	23.321	42.920	66.228	1.000

SMI: DuMouchel $3s_0$

DuMouchel $3s_0$= .899		RE ~ normal distribution				**DIC = 77.143**
	mean	s.d.	MC err.	2.50%	50%	97.50%	Rhat
mu.1	0.967	0.150	.005	0.682	0.962	1.263	1.002
theta.new	0.988	0.445	.014	0.082	0.985	1.891	1.001
β_1	-0.461	0.196	.007	-0.842	-0.465	-0.082	1.002
β_2	-0.678	0.194	.006	-1.081	-0.674	-0.310	1.002
β_3	-0.860	0.285	.009	-1.426	-0.862	-0.308	1.002
P $\beta_1 < 0$	.993	0.083	.003	1.000	1.000	1.000	1.064
P $\beta_2<0$	.999	0.000	.000	1.000	1.000	1.000	1.000
P $\beta_3<0$	.999	0.000	.000	1.000	1.000	1.000	1.000
P mu.1 < 0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	.017	0.129	.003	0.000	0.000	0.000	1.006
τ	0.416	0.075	.002	0.283	0.412	0.577	1.001
τ^2	0.178	0.065	.002	0.080	0.170	0.333	1.001
deviance	42.951	11.141	.418	22.910	42.500	66.769	1.000
DuMouchel $3s_0$ = .899		RE ~ t_4-distribution				**DIC = 77.517**
	mean	s.d.	MC err.	2.5%	50%	97.50%	Rhat
mu.1	0.877	0.177	.005	0.518	0.877	1.205	1.001
theta.new	0.875	0.373	.011	0.138	0.881	1.639	1.002
β_1	-0.375	0.212	.007	-0.783	-0.374	0.051	1.001
β_2	-0.603	0.204	.007	-1.014	-0.607	-0.199	1.001
β_3	-0.761	0.268	.007	-1.287	-0.760	-0.236	1.001
P $\beta_1 < 0$	.958	0.200	.006	0.000	1.000	1.000	1.007
P $\beta_2<0$	.997	0.055	.001	1.000	1.000	1.000	1.133
P $\beta_3<0$	.998	0.045	.001	1.000	1.000	1.000	1.104
P mu.1 < 0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	.015	0.121	.004	0.000	0.000	0.000	1.055
τ	0.317	0.068	.002	0.203	0.312	0.465	1.001
τ^2	0.105	0.046	.002	0.041	0.097	0.216	1.001
deviance	43.573	11.228	.339	22.976	43.670	66.673	1.001

SMI: Gamma on Precision

Gamma on precision (.1,.1)	RE ~ normal distribution						**DIC = 76.941**
	mean	s.d.	MC err.	2.50%	50%	97.50%	Rhat
mu.1	0.974	0.144	.005	0.674	0.970	1.276	1.002
theta.new	0.978	0.453	.013	0.056	0.978	1.884	1.003
β_1	-0.457	0.195	.006	-0.839	-0.452	-0.059	1.001
β_2	-0.672	0.191	.007	-1.040	-0.669	-0.295	1.000
β_3	-0.850	0.272	.008	-1.399	-0.855	-0.299	1.001
P $\beta_1 < 0$	.991	0.094	.003	1.000	1.000	1.000	1.017
P $\beta_2<0$	.999	0.000	.000	1.000	1.000	1.000	1.000
P $\beta_3<0$	.999	0.032	.000	1.000	1.000	1.000	1.291
P mu.1 < 0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	.013	0.113	.003	0.000	0.000	0.000	1.036
τ	0.420	0.072	.002	0.287	0.414	0.578	1.002
τ^2	0.182	0.063	.002	0.083	0.171	0.335	1.002
deviance	42.593	10.673	.384	22.683	42.320	65.784	1.000
Gamma on precision (.1,.1)	RE ~ t_4-distribution						**DIC = 76.288**
	mean	s.d.	MC err.	2.5%	50%	97.50%	Rhat
mu.1	0.891	0.190	.005	0.531	0.887	1.257	1.001
theta.new	0.893	0.400	.013	0.120	0.880	1.702	1.001
β_1	-0.388	0.229	.006	-0.842	-0.384	0.085	1.000
β_2	-0.617	0.221	.007	-1.043	-0.616	-0.191	1.001
β_3	-0.773	0.290	.008	-1.343	-0.762	-0.242	1.000
P $\beta_1 < 0$	.953	0.212	.006	0.000	1.000	1.000	1.000
P $\beta_2<0$	.999	0.032	.000	1.000	1.000	1.000	1.291
P $\beta_3<0$	.999	0.032	.000	1.000	1.000	1.000	1.291
P mu.1 < 0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	.009	0.094	.003	0.000	0.000	0.000	1.017
τ	0.334	0.068	.003	0.223	0.328	0.484	1.003
τ^2	0.116	0.048	.002	0.050	0.107	0.234	1.003
deviance	41.893	10.454	.308	22.143	41.360	63.222	1.003

SMI: Half-Normal

Half Normal		RE ~ normal distribution				**DIC =** 77.440
	mean	s.d.	MC error	2.50%	50%	97.50%	Rhat
mu.1	0.967	0.146	.005	0.695	0.965	1.271	1.001
theta.new	0.955	0.455	.015	0.052	0.953	1.885	1.001
β_1	-0.465	0.199	.006	-0.862	-0.466	-0.082	1.000
β_2	-0.663	0.196	.006	-1.061	-0.667	-0.284	1.000
β_3	-0.840	0.281	.009	-1.379	-0.835	-0.317	1.001
P $\beta_1 < 0$	.988	0.109	.003	1.000	1.000	1.000	1.039
P $\beta_2<0$	.999	0.000	.000	1.000	1.000	1.000	1.000
P $\beta_3<0$	.997	0.055	.002	1.000	1.000	1.000	1.294
P mu.1 < 0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	.019	0.136	.004	0.000	0.000	0.000	1.008
τ	0.430	0.074	.002	0.300	0.427	0.581	1.001
τ^2	0.191	0.067	.002	0.090	0.182	0.338	1.001
deviance	42.423	10.265	.318	23.003	42.090	63.959	1.003
Half-Normal		RE ~ t_4-distribution				**DIC =** 77.093
	mean	s.d.	MC error	2.5%	50%	97.50%	Rhat
mu.1	0.892	0.191	.006	0.529	0.894	1.297	1.006
theta.new	0.878	0.404	.013	0.100	0.864	1.719	1.000
β_1	-0.388	0.232	.006	-0.862	-0.387	0.060	1.003
β_2	-0.624	0.227	.000	-1.072	-0.627	-0.185	1.003
β_3	-0.771	0.299	.004	-1.359	-0.766	-0.199	1.001
P $\beta_1 < 0$	.953	0.212	.006	0.000	1.000	1.000	1.008
P $\beta_2<0$	.999	0.032	.000	1.000	1.000	1.000	1.291
P $\beta_3<0$	.989	0.104	.004	1.000	1.000	1.000	1.109
P mu.1 < 0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	.014	0.117	.004	0.000	0.000	0.000	1.038
τ	0.339	0.076	.003	0.209	0.330	0.517	1.001
τ^2	0.121	0.057	.002	0.044	0.109	0.268	1.001
deviance	41.956	10.901	.421	22.754	41.600	65.843	1.001

SMI Uniform on $\boldsymbol{\tau^2}$

Uniform on $\boldsymbol{\tau^2}$		RE ~ normal distribution		**DIC = 77.007**		
	mean	s.d.	MC error	2.50%	50%	97.50%	Rhat
mu.1	0.980	0.151	.005	0.677	0.979	1.275	1.002
theta.new	0.983	0.467	.015	0.109	0.976	1.937	1.004
β_1	-0.463	0.206	.006	-0.859	-0.459	-0.057	1.003
β_2	-0.683	0.201	.007	-1.064	-0.688	-0.277	1.005
β_3	-0.860	0.282	.010	-1.385	-0.855	-0.325	1.001
P $\beta_1 < 0$	.990	0.099	.003	1.000	1.000	1.000	1.018
P $\beta_2<0$	.998	0.045	.001	1.000	1.000	1.000	1.104
P $\beta_3<0$	.999	0.032	.000	1.000	1.000	1.000	1.291
P mu.1 < 0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	.017	0.129	.003	0.000	0.000	0.000	1.021
τ	0.437	0.073	.002	0.310	0.431	0.587	1.002
τ^2	0.196	0.067	.002	0.096	0.186	0.345	1.002
deviance	41.676	10.207	.330	22.761	41.170	62.697	1.001
Uniform on τ^2		RE ~ t_4-distribution		**DIC = 76.945**		
	mean	s.d.	MC error	2.5%	50%	97.50%	Rhat
mu.1	0.880	0.178	.006	0.543	0.883	1.231	1.003
theta.new	0.906	0.401	.014	0.122	0.904	1.665	1.007
β_1	-0.382	0.217	.007	-0.802	-0.385	0.052	1.003
β_2	-0.608	0.212	.007	-1.008	-0.616	-0.189	1.001
β_3	-0.761	0.280	.008	-1.293	-0.752	-0.206	1.001
P $\beta_1 < 0$	.958	0.200	.007	0.000	1.000	1.000	1.001
P $\beta_2<0$	.998	0.045	.001	1.000	1.000	1.000	1.292
P $\beta_3<0$	.994	0.077	.002	1.000	1.000	1.000	1.038
P mu.1 < 0	<.001	0.000	.000	0.000	0.000	0.000	1.000
Ptheta.new<0	.014	0.117	.003	0.000	0.000	0.000	1.016
τ	0.340	0.072	.003	0.215	0.336	0.498	1.001
τ^2	0.121	0.052	.002	0.046	0.113	0.248	1.001
deviance	41.855	10.717	.348	23.092	40.900	64.358	1.001

Appendix C. Zinc for the Cold Meta-analysis: Bayesian summary results

Zinc for the Common Cold: Uniform on τ (0,5)

Uniform on τ (0,5)			RE~ normal distribution				DIC = 4.640
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
β_0	-1.084	0.652	0.01971	-2.448	-1.062	0.251	1.004
θ_{new}	-1.121	1.589	0.0476	-4.549	-1.058	2.158	1.005
P mu.1< 0	.961	0.194	.006	0	1	1	1.003
P $\theta_{new} < 0$	.8	0.4	.011	0	1	1	1.002
τ	1.372	0.63	.021	0.621	1.214	3.006	1.003
τ^2	2.28	2.514	.082	0.386	1.474	9.034	1.003
deviance	-1.203	3.461	0.108	-6.063	-1.913	7.227	1.004
Uniform on τ (0,5)			RE~ t_4 distribution				DIC = 5.091
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
β_0	-0.927	0.614	.012	-2.220	-0.901	0.189	1.001
θ_{new}	-0.946	1.495	.029	-4.044	-0.890	1.805	1.001
P mu.1< 0	.954	0.209	.004	0.000	1.000	1.000	1.001
P $\theta_{new} < 0$	.788	0.409	.008	0.000	1.000	1.000	1.001
τ	1.234	0.659	.013	0.447	1.071	3.068	1.001
τ^2	1.955	2.546	.046	0.200	1.146	9.415	1.001
deviance	-0.955	3.619	.074	-6.017	-1.505	7.615	1.001

Zinc for the Common Cold: DuMouchel s_0

DuMouchel s_0			RE ~ normal distribution				DIC = 5.178
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
β_0	-1.028	0.466	.014	-2.068	-1.013	-0.197	1.000
θ_{new}	-1.089	1.196	.040	-3.543	-1.052	1.190	1.008
P mu.1< 0	.989	0.104	.004	1.000	1.000	1.000	1.015
P $\theta_{new} < 0$	.864	0.343	.011	0.000	1.000	1.000	1.004
τ	1.034	0.406	.012	0.516	0.958	2.047	1.002
τ^2	1.235	1.168	.035	0.267	0.917	4.191	1.002
deviance	-0.734	3.668	.113	-5.820	-1.311	7.839	1.006
DuMouchel s_0			RE ~ t_4 distribution				DIC = 5.543
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
β_0	-0.887	0.453	.016	-1.860	-0.877	-0.047	1.006
θ_{new}	-0.878	1.053	.035	-3.049	-0.879	1.133	-0.879
P mu.1< 0	.979	0.143	.005	1.000	1.000	1.012	1.000
P $\theta_{new} < 0$	.831	0.375	.011	1.000	1.000	1.013	1.000
τ	0.871	0.405	.014	0.378	0.786	1.877	1.003
τ^2	0.923	1.244	.042	0.143	0.618	3.521	1.003
deviance	-0.587	4.092	.132	-5.979	-1.307	9.476	1.011

Zinc for the Common Cold: DuMouchel $s_0/3$

DuMouchel $s_0/3$			RE ~ normal distribution				DIC = 4.733
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
β_0	-1.016	0.439	.016	-1.969	-0.989	-0.141	1.006
θ_{new}	-1.023	1.171	.041	-3.363	-0.947	1.108	1.002
P mu.1< 0	.988	0.109	.004	1.000	1.000	1.000	1.029
P $\theta_{new} < 0$	.851	0.356	.011	0.000	1.000	1.000	1.001
τ	1.008	0.402	.013	0.518	0.929	1.939	1.006
τ^2	1.178	1.372	.043	0.268	0.864	3.759	1.006
deviance	-0.957	3.672	.120	-5.953	-1.718	7.837	1.001
DuMouchel $s_0/3$			RE~ t_4 distribution				DIC = 4.90
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
β_0	-0.882	0.466	.018	-1.890	-0.843	-0.008	1.006
θ_{new}	-0.848	1.022	.032	-2.945	-0.810	1.137	1.003
P mu.1< 0	.977	0.150	.004	1.000	1.000	1.000	1.011
P $\theta_{new} < 0$	.838	0.368	.012	0.000	1.000	1.000	1.010
τ	0.831	0.415	.014	0.337	0.733	1.899	1.007
τ^2	0.862	1.139	.036	0.113	0.537	3.605	1.007
deviance	-0.645	3.767	.138	-5.867	-1.416	8.154	1.006

Zinc for the Common Cold: DuMouchel $3*s_0$

DuMouchel $3*s_0$			RE ~ normal distribution				DIC = 4.942
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
β_0	-1.000	0.501	.017	-1.942	-0.986	0.045	1.001
θ_{new}	-1.001	1.275	.043	-3.485	-0.970	1.510	1.004
P mu.1< 0	.969	0.173	.004	0.000	1.000	1.000	1.007
P $\theta_{new} < 0$	.807	0.395	.011	0.000	1.000	1.000	1.001
τ	1.103	0.452	.015	0.532	1.002	2.313	1.006
τ^2	1.421	1.516	.049	0.283	1.004	5.354	1.006
deviance	-0.874	3.668	.084	-5.833	-1.629	8.229	1.000
DuMouchel $3*s_0$			RE ~ t_4 distribution				DIC = 5.050
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
β_0	-0.906	0.486	.015	-1.951	-0.870	-0.058	1.000
θ_{new}	-0.921	1.165	.036	-3.344	-0.887	1.354	1.000
P mu.1< 0	.982	0.133	.004	1.000	1.000	1.000	1.003
P $\theta_{new} < 0$	.829	0.376	.011	0.000	1.000	1.000	1.000
τ	0.935	0.430	.015	0.378	0.841	1.986	1.000
τ^2	1.059	1.210	.040	0.143	0.707	3.947	1.000
deviance	-0.841	3.778	.120	-6.099	-1.654	7.787	1.000

Zinc for the Common Cold: Gamma on Precision

Inverse Gamma (.01, .01)			RE ~ normal distribution				DIC = 5.165
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
β_0	-0.988	0.528	.016	-2.114	-0.980	0.125	1.007
θ_{new}	-0.958	1.383	.046	-3.906	-0.969	1.830	1.005
P mu.1< 0	.968	0.176	.005	0.000	1.000	1.000	1.020
P $\theta_{new} < 0$	.798	0.401	.013	0.000	1.000	1.000	1.005
τ	1.174	0.556	.015	0.565	1.044	2.586	1.001
τ^2	1.688	2.325	.063	0.319	1.091	6.688	1.001
deviance	-0.803	3.928	.123	-6.010	-1.510	8.916	1.003
Inverse Gamma (.01, .01)			RE ~ t_4 distribution				DIC = 4.817
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
β_0	-0.892	0.503	.016	-1.931	-0.859	0.031	1.001
θ_{new}	-0.908	1.307	.047	-3.573	-0.908	1.992	1.003
P mu.1< 0	.972	0.165	.005	0.000	1.000	1.000	1.007
P $\theta_{new} < 0$	.812	0.391	.015	0.000	1.000	1.000	1.001
τ	1.014	0.513	.017	0.407	0.906	2.258	1.004
τ^2	1.291	1.683	.057	0.165	0.820	5.098	1.004
deviance	-1.036	3.675	.101	-5.983	-1.725	7.932	1.001

Zinc for the Common Cold: Half-Normal

Half-normal			RE~ normal distribution				DIC = 4.715
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
β_0	-1.016	0.589	.017	-2.170	-0.993	0.096	1.003
θ_{new}	-1.076	1.491	.054	-4.153	-1.011	1.735	1.001
P mu.1< 0	.957	0.203	.006	0.000	1.000	1.000	1.008
P $\theta_{new} < 0$	.787	0.409	.012	0.000	1.000	1.000	1.000
τ	1.300	0.410	.013	0.644	1.244	2.245	1.004
τ^2	1.857	1.189	.038	0.414	1.548	5.040	1.004
deviance	-1.126	3.388	.094	-6.116	-1.687	7.154	1.003
Half-normal			RE ~ t_4 distribution				DIC = 4.607
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
β_0	-0.946	0.584	.019	-2.271	-0.911	0.154	1.010
θ_{new}	-0.920	1.310	.041	-3.512	-0.929	1.714	1.001
P mu.1< 0	.955	0.207	.008	0.000	1.000	1.000	1.072
P $\theta_{new} < 0$	.793	0.405	.012	0.000	1.000	1.000	1.001
τ	1.193	0.426	.015	0.518	1.138	2.168	1.000
τ^2	1.605	1.151	.041	0.268	1.295	4.697	1.000
deviance	-1.256	3.593	.120	-6.105	-1.956	7.935	1.001

Zinc for the Common Cold: Uniform distribution on $\boldsymbol{\tau^2}$

τ^2 ~ uniform (0, 1000)			RE ~ normal distribution				DIC = 4.931
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
β_0	-1.033	0.924	.030	-2.575	-1.031	0.535	1.012
θ_{new}	-0.964	2.191	.072	-5.426	-0.969	3.662	1.000
P mu.1< 0	.937	0.243	.007	0.000	1.000	1.000	1.004
P θ_{new} < 0	.721	0.449	.016	0.000	1.000	1.000	1.000
τ	1.772	1.210	.037	0.710	1.455	4.787	1.002
τ^2	4.601	13.686	.411	0.504	2.117	22.915	1.002
deviance	-1.128	3.648	.135	-6.141	-1.869	7.942	1.003
τ^2 ~ uniform (0, 1000)			RE ~ t_4 distribution				DIC = 4.544
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
β_0	-0.972	0.856	.030	-2.610	-0.965	0.694	1.001
θ_{new}	-0.849	2.251	.083	-4.928	-0.877	3.185	1.007
P mu.1< 0	.932	0.252	.009	0.000	1.000	1.000	1.012
P θ_{new} < 0	.746	0.436	.016	0.000	1.000	1.000	1.000
τ	1.690	1.261	.038	0.560	1.400	4.588	1.001
τ^2	4.444	15.354	.461	0.313	1.960	21.054	1.001
deviance	-1.308	3.503	.101	-6.088	-1.920	7.347	1.002

Appendix D. Detracking k =22 Meta-analysis: Bayesian summary results

Detracking (k=22) : Uniform on $\boldsymbol{\tau}$ (0,5)

Uniform on τ (0,5)			RE ~ normal distribution				DIC = -3.358
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
mu.1	0.142	0.179	.005	-0.226	0.145	0.479	1.004
β_1 (Mid-achieving)	1.276	0.502	.013	0.341	1.249	2.318	1.000
$B_0 + \beta_1$	1.418	0.471	.014	0.552	1.416	2.368	1.000
θ_{new}	0.168	0.774	.023	-1.308	0.172	1.731	1.004
θ_{new}.Mid-achieving	1.444	0.937	.033	-0.334	1.404	3.282	1.000
P mu.1 < 0	.204	0.403	. 012	0.000	0.000	1.000	1.007
P β_1 < 0	.005	0.070	.070	0.000	0.000	0.000	1.071
P Mid Group < 0	.001	0.032	.032	0.000	0.000	0.000	1.291
P θ_{new} < 0	.413	0.493	.435	0.000	0.000	1.000	1.003
P θ_{new}.Mid < 0	.054	0.226	.007	0.000	0.000	1.000	1.023
τ	0.753	0.152	.015	0.506	0.732	1.103	1.000
τ^2	0.590	0.260	.009	0.257	0.536	1.218	1.000
deviance	-24.345	7.078	.221	-36.009	-24.915	-8.455	1.000
Uniform on τ (0,5)			RE ~ t_4 distribution				DIC = -4.880
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
mu.1	0.104	0.138	.004	-0.188	0.103	0.396	1.000
β_1 (Mid-achieving)	0.859	0.592	.017	-0.153	0.790	2.095	1.006
$B_0 + \beta_1$	0.962	0.572	.017	0.003	0.898	2.253	1.006
θ_{new}	0.118	0.590	.019	-1.008	0.107	1.241	1.000
θ_{new}.Mid-achieving	0.946	0.799	.025	-0.545	0.890	2.738	1.003
P mu.1 < 0	.200	0.400	.012	0.000	0.000	1.000	1.002
P β_1 < 0	.053	0.224	.007	0.000	0.000	1.000	1.005
P Mid Group < 0	.025	0.156	.005	0.000	0.000	0.000	1.038
P θ_{new} < 0	.410	0.492	.017	0.000	0.000	1.000	1.000
P θ_{new}.Mid < 0	.089	0.285	.009	0.000	0.000	1.000	1.000
τ	0.535	0.133	.004	0.316	0.522	0.857	1.000
τ^2	0.304	0.157	.004	0.100	0.272	0.734	1.000
deviance	-25.56	6.563	.196	-37.158	-25.955	-11.392	1.008

Detracking (k=22): DuMouchel s_0

DuMouchel s_0			RE ~ normal distribution			DIC = -2.497
	mean	s.d.	2.50%	50%	97.50%	Rhat
mu.1	0.141	0.162	-0.178	0.143	0.474	1.002
β_1 (Mid-achieving)	1.308	0.470	0.399	1.296	2.243	1.004
$B_0 + \beta_1$	1.449	0.439	0.643	1.448	2.345	1.014
θ_{new}	0.136	0.745	-1.251	0.104	1.584	1.003
θ_{new}.Mid-achieving	1.421	0.816	-0.170	1.386	3.072	1.006
P mu.1 < 0	.199	0.399	0.000	0.000	1.000	1.000
P $\beta_1 > 0$	.004	0.063	0.000	0.000	0.000	1.029
P Mid Group < 0	.000	0.000	0.000	0.000	0.000	1.000
P $\theta_{new} < 0$	.437	0.496	0.000	0.000	1.000	1.002
P θ_{new}.Mid < 0	.037	0.189	0.000	0.000	1.000	1.033
τ	0.703	0.132	0.485	0.689	1.013	1.000
τ^2	0.511	0.196	0.235	0.474	1.026	1.000
deviance	-23.485	7.441	-36.206	-24.210	-7.271	1.001
DuMouchel s_0			RE ~ t_4 distribution			DIC = -3.657
	mean	s.d.	2.50%	50%	97.50%	Rhat
mu.1	0.112	0.133	-0.145	0.113	0.387	1.003
β_1 (Mid-achieving)	0.797	0.528	-0.045	0.740	2.053	1.001
$B_0 + \beta_1$	0.909	0.508	0.072	0.844	2.055	1.001
θ_{new}	0.136	0.504	-0.796	0.123	1.113	1.004
θ_{new}.Mid-achieving	0.896	0.708	-0.411	0.852	2.402	1.002
P mu.1 < 0	.201	0.401	0.000	0.000	1.000	1.000
P $\beta_1 > 0$	.035	0.184	0.000	0.000	1.000	1.007
P Mid Group < 0	.018	0.133	0.000	0.000	0.000	0.999
P $\theta_{new} < 0$	.392	0.488	0.000	0.000	1.000	1.001
P θ_{new}.Mid < 0	.082	0.274	0.000	0.000	1.000	1.001
τ	0.479	0.117	0.290	0.469	0.739	1.003
τ^2	0.243	0.123	0.084	0.220	0.546	1.003
deviance	-24.740	7.050	-36.640	-25.585	-9.271	1.000

Detracking (k=22): DuMouchel $s_0/3$

DuMouchel $s_0/3$			RE ~ normal distribution			DIC =-2.896
	mean	s.d.	2.50%	50%	97.50%	Rhat
mu.1	0.139	0.176	-0.221	0.136	0.504	1.000
β_1 (Mid-achieving)	1.284	0.489	0.307	1.270	2.255	1.002
$B_0 + \beta_1$	1.424	0.450	0.523	1.420	2.362	1.001
θ_{new}	0.145	0.751	-1.342	0.135	1.657	1.003
θ_{new}.Mid-achieving	1.384	0.844	-0.203	1.360	3.004	1.001
P mu.1 < 0	.194	0.395	0.000	0.000	1.000	1.000
P $\beta_1 > 0$	.004	0.063	0.000	0.000	0.000	1.166
P Mid Group < 0	.002	0.045	0.000	0.000	0.000	1.292
P $\theta_{new} < 0$	.416	0.493	0.000	0.000	1.000	1.000
P θ_{new}.Mid < 0	.050	0.218	0.000	0.000	1.000	1.018
τ	0.699	0.139	0.482	0.681	1.031	1.000
τ^2	0.508	0.210	0.232	0.464	1.061	1.000
deviance	-23.584	7.707	-36.419	-24.505	-6.271	1.000
DuMouchel $s_0/3$			RE ~ t_4 distribution			DIC = -4.734
	mean	s.d.	2.50%	50%	97.50%	Rhat
mu.1	0.106	0.134	-0.151	0.102	0.373	1.003
β_1 (Mid-achieving)	0.804	0.554	-0.104	0.735	2.097	1.007
$B_0 + \beta_1$	0.910	0.534	0.052	0.844	2.212	1.005
θ_{new}	0.086	0.501	-0.926	0.088	1.080	1.002
θ_{new}.Mid-achieving	0.891	0.746	-0.496	0.812	2.554	1.009
P mu.1 < 0	.200	0.400	0.000	0.000	1.000	1.007
P $\beta_1 > 0$	.049	0.216	0.000	0.000	1.000	1.010
P Mid Group < 0	.014	0.117	0.000	0.000	0.000	1.001
P $\theta_{new} < 0$	.414	0.493	0.000	0.000	1.000	1.004
P θ_{new}.Mid < 0	.087	0.282	0.000	0.000	1.000	1.014
τ	0.478	0.116	0.296	0.462	0.747	1.001
τ^2	0.242	0.123	0.088	0.213	0.558	1.001
deviance	-25.255	7.024	-37.086	-26.465	-9.528	1.004

Detracking (k=22): DuMouchel $3s_0$

DuMouchel $3s_0$			RE ~ normal distribution				DIC = -2.608
	mean	s.d.		2.50%	50%	97.50%	Rhat
mu.1	0.143	0.165	.005	-0.182	0.139	0.474	1.000
β_1 (Mid-achieving)	1.271	0.482	.016	0.291	1.271	2.237	1.006
$B_0 + \beta_1$	1.413	0.453	.015	0.487	1.415	2.325	1.006
θ_{new}	0.136	0.724	.023	-1.374	0.158	1.623	1.001
θ_{new}.Mid-achieving	1.411	0.854	.024	-0.288	1.421	3.085	1.000
P mu.1 < 0	.180	0.384	.011	0.000	0.000	1.000	1.000
P $\beta_1 > 0$	.007	0.083	.003	0.000	0.000	0.000	1.037
P Mid Group < 0	.001	0.032	.001	0.000	0.000	0.000	1.291
P $\theta_{new} < 0$	.411	0.492	.017	0.000	0.000	1.000	1.002
P θ_{new}.Mid < 0	.044	0.205	.006	0.000	0.000	1.000	1.011
τ	0.707	0.135	.004	0.484	0.694	0.990	1.001
τ^2	0.518	0.203	.006	0.235	0.482	0.980	1.001
deviance	-23.829	7.310	.202	-36.700	-24.170	-6.997	1.002
DuMouchel $3s_0$			RE ~ t_4 distribution				DIC = -4.225
	mean	s.d.		2.50%	50%	97.50%	Rhat
mu.1	0.104	0.134	.005	-0.160	0.108	0.360	1.002
β_1 (Mid-achieving)	0.826	0.583	.021	-0.154	0.760	2.169	1.003
$B_0 + \beta_1$	0.930	0.568	.020	-0.024	0.856	2.204	1.003
θ_{new}	0.103	0.526	.017	-0.958	0.102	1.073	1.009
θ_{new}.Mid-achieving	0.946	0.764	.083	-0.472	0.906	2.522	1.003
P mu.1 < 0	.213	0.409	.015	0.000	0.000	1.000	1.004
P $\beta_1 > 0$	.047	0.212	.007	0.000	0.000	1.000	1.029
P Mid Group < 0	.029	0.168	.006	0.000	0.000	1.000	1.000
P $\theta_{new} < 0$	.414	0.493	.008	0.000	0.000	1.000	1.004
P θ_{new}.Mid < 0	.090	0.286	.001	0.000	0.000	1.000	1.003
τ	0.487	0.120	.004	0.300	0.467	0.771	1.004
τ^2	0.251	0.133	.004	0.090	0.218	0.595	1.004
deviance	-25.13	6.726	.200	-36.38	-25.690	-10.531	1.003

Detracking (k=22): Inverse Gamma on precision

Inverse Gamma on $1/\tau^2$			RE ~ normal distribution			DIC = -3.414
	mean	s.d.	2.50%	50%	97.50%	Rhat
mu.1	0.134	0.184	-0.233	0.140	0.481	1.011
β_1 (Mid-achieving)	1.319	0.492	0.431	1.302	2.332	1.003
$B_0 + \beta_1$	1.453	0.458	0.570	1.442	2.349	1.002
θ_{new}	0.125	0.764	-1.479	0.144	1.575	1.001
θ_{new}.Mid-achieving	1.464	0.898	-0.390	1.492	3.283	1.000
P mu.1 < 0	.223	0.416	0.000	0.000	1.000	1.006
P $\beta_1 > 0$	.005	0.070	0.000	0.000	0.000	1.116
P Mid Group < 0	.001	0.032	0.000	0.000	0.000	1.291
P $\theta_{new} < 0$	.415	0.493	0.000	0.000	1.000	1.005
P θ_{new}.Mid < 0	.056	0.230	0.000	0.000	1.000	1.000
τ	0.735	0.148	0.491	0.720	1.053	1.004
τ^2	0.562	0.242	0.241	0.519	1.108	1.004
deviance	-24.165	7.257	-36.150	-24.830	-8.146	1.000
Inverse Gamma on $1/\tau^2$			RE ~ t_4 distribution			DIC = -4.682
	mean	s.d.	2.50%	50%	97.50%	Rhat
mu.1	0.098	0.143	-0.180	0.100	0.374	1.001
β_1 (Mid-achieving)	0.831	0.574	-0.156	0.781	2.167	1.002
$B_0 + \beta_1$	0.929	0.569	-0.043	0.879	2.210	1.002
θ_{new}	0.103	0.550	-0.977	0.117	1.140	1.000
θ_{new}.Mid-achieving	0.923	0.797	-0.497	0.859	2.740	1.001
P mu.1 < 0	.235	0.424	0.000	0.000	1.000	1.002
P $\beta_1 > 0$	.044	0.205	0.000	0.000	1.000	1.007
P Mid Group < 0	.027	0.162	0.000	0.000	1.000	1.013
P $\theta_{new} < 0$	.404	0.491	0.000	0.000	1.000	1.000
P θ_{new}.Mid < 0	.101	0.301	0.000	0.000	1.000	1.004
τ	0.508	0.128	0.309	0.492	0.791	1.000
τ^2	0.275	0.151	0.096	0.242	0.626	1.000
deviance	-25.35	6.907	-37.120	-25.815	-11.330	1.001

Detracking (k=22): Half-normal

Half-normal			RE ~ normal distribution				DIC = -3.817
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
mu.1	0.153	0.183	.006	-0.208	0.144	0.523	1.014
β_1 (Mid-achieving)	1.308	0.537	.015	0.175	1.312	2.372	1.002
$B_0 + \beta_1$	1.461	0.501	.014	0.434	1.447	2.421	1.001
θ_{new}	0.145	0.806	.015	-1.446	0.152	1.706	1.004
θ_{new}.Mid-achieving	1.473	0.931	.007	-0.274	1.455	3.296	1.002
P mu.1 < 0	.186	0.389	.014	0.000	0.000	1.000	1.021
P $\beta_1 > 0$	.013	0.113	.003	0.000	0.000	0.000	1.067
P Mid Group < 0	.005	0.070	.002	0.000	0.000	0.000	1.116
P $\theta_{new} < 0$	.431	0.495	.015	0.000	0.000	1.000	1.003
P θ_{new}.Mid < 0	.062	0.241	.007	0.000	0.000	1.000	1.008
τ	0.781	0.155	.005	0.536	0.762	1.141	1.001
τ^2	0.634	0.263	.008	0.287	0.580	1.301	1.001
deviance	-24.756	7.309	.260	-36.810	-25.720	-8.260	1.000
Half-normal			RE ~ t_4 distribution				DIC = -5.211
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
mu.1	0.108	0.148	.005	-0.180	0.108	0.396	1.001
β_1 (Mid-achieving)	0.893	0.645	.018	-0.180	0.811	2.370	1.000
$B_0 + \beta_1$	1.001	0.625	.017	-0.015	0.902	2.433	1.000
θ_{new}	0.112	0.591	.020	-1.068	0.118	1.315	1.004
θ_{new}.Mid-achieving	1.001	0.828	.024	-0.458	0.970	2.835	1.002
P mu.1 < 0	.227	0.419	.012	0.000	0.000	1.000	1.000
P $\beta_1 > 0$	.051	0.220	.005	0.000	0.000	1.000	1.006
P Mid Group < 0	.029	0.168	.005	0.000	0.000	1.000	1.028
P $\theta_{new} < 0$	.403	0.491	.014	0.000	0.000	1.000	1.001
P θ_{new}.Mid < 0	.101	0.301	.011	0.000	0.000	1.000	1.000
τ	0.577	0.149	.005	0.347	0.556	0.909	1.000
τ^2	0.355	0.194	.006	0.121	0.309	0.826	1.000
deviance	-25.72	6.942	.170	-37.000	-26.270	-10.293	1.005

Detracking (k=22) : Uniform on $\boldsymbol{\tau^2}$

Uniform on τ^2			RE ~ normal distribution			DIC = -4.091
	mean	s.d.	2.50%	50%	97.50%	Rhat
mu.1	0.148	0.176	-0.181	0.141	0.492	1.004
β_1 (Mid-achieving)	1.281	0.520	0.292	1.285	2.312	1.001
$B_0 + \beta_1$	1.429	0.486	0.468	1.429	2.429	1.005
θ_{new}	0.152	0.804	-1.497	0.139	1.704	1.002
θ_{new}.Mid-achieving	1.421	0.921	-0.388	1.407	3.196	1.000
P mu.1 < 0	.197	0.398	0.000	0.000	1.000	1.009
P $\beta_1 > 0$	.009	0.094	0.000	0.000	0.000	1.051
P Mid Group < 0	.003	0.055	0.000	0.000	0.000	1.294
P $\theta_{new} < 0$	.417	0.493	0.000	0.000	1.000	1.000
P θ_{new}.Mid < 0	.060	0.237	0.000	0.000	1.000	1.000
τ	0.781	0.159	0.535	0.759	1.147	1.004
τ^2	0.635	0.270	0.286	0.577	1.315	1.004
deviance	-24.833	7.040	-36.229	-25.765	-8.790	1.003
Uniform on τ^2			RE ~ t_4 distribution			DIC = -4.026
	mean	s.d.	2.50%	50%	97.50%	Rhat
mu.1	0.098	0.148	-0.200	0.090	0.411	1.001
β_1 (Mid-achieving)	0.914	0.624	-0.148	0.851	2.360	1.004
$B_0 + \beta_1$	1.011	0.607	0.014	0.928	2.409	1.003
θ_{new}	0.129	0.600	-1.030	0.111	1.376	1.002
θ_{new}.Mid-achieving	1.017	0.849	-0.543	0.961	2.876	1.000
P mu.1	0.228	0.419	0.000	0.000	1.000	1.001
P β_1	0.049	0.216	0.000	0.000	1.000	1.000
P Mid Group	0.022	0.147	0.000	0.000	0.000	1.012
P θ_{new}	0.413	0.493	0.000	0.000	1.000	1.002
P θ_{new}.Mid	0.099	0.299	0.000	0.000	1.000	1.000
τ	0.573	0.149	0.338	0.553	0.925	1.004
τ^2	0.350	0.192	0.114	0.306	0.855	1.004
deviance	-25.15	7.178	-36.750	-25.935	-8.892	1.001

Appendix E. Detracking k = 20 Meta-analysis: Bayesian Summary Results

Detracking (k=20): Uniform on $\boldsymbol{\tau}$ (0,5)

Uniform on τ			RE ~ normal distribution				DIC = -10.332
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
mu.1	0.358	0.151	.006	0.066	0.358	0.660	1.001
β_1 (High)	-0.478	0.240	.008	-0.975	-0.488	0.012	1.001
β_2 (All)	-0.555	0.273	.007	-1.094	-0.548	-0.045	1.000
$B_0 + \beta_1$	-0.120	0.187	.007	-0.483	-0.120	0.239	1.005
$B_0 + \beta_2$	-0.197	0.221	.006	-0.673	-0.192	0.259	1.002
θ_{new}	0.331	0.441	.013	-0.505	0.330	1.210	1.001
θ_{new}.High	-0.110	0.478	.014	-1.088	-0.124	0.904	1.003
θ_{new}.All	-0.196	0.483	.015	-1.194	-0.192	0.742	1.000
P mu.1< 0	.012	0.109	.003	0.000	0.000	0.000	1.009
P $\beta_1 < 0$	.972	0.165	.005	0.000	1.000	1.000	1.011
P $\beta_2 < 0$	.979	0.143	.005	1.000	1.000	1.000	1.023
P High Group < 0	.754	0.431	.011	0.000	1.000	1.000	1.003
P All Group < 0	.835	0.371	.012	0.000	1.000	1.000	1.001
P $\theta_{new} < 0$	.208	0.406	.013	0.000	0.000	1.000	1.000
P θ_{new}.High Group<0	.594	0.491	.015	0.000	1.000	1.000	1.000
P θ_{new}.All Group<0	.657	0.475	.013	0.000	1.000	1.000	1.000
τ	0.409	0.085	.003	0.273	0.399	0.611	1.000
τ^2	0.174	0.075	.002	0.075	0.159	0.374	1.000
deviance	-27.936	6.343	.217	-38.240	-28.685	-13.592	1.000
Uniform on τ			RE ~ t_4 distribution				DIC = -11.298
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
mu.1	0.335	0.150	.005	0.051	0.333	0.632	1.005
β_1 (High)	-0.439	0.242	.007	-0.941	-0.432	0.018	1.002
β_2 (All)	-0.521	0.316	.008	-1.147	-0.506	0.097	1.002
$B_0 + \beta_1$	-0.104	0.192	.006	-0.513	-0.091	0.249	1.000
$B_0 + \beta_2$	-0.186	0.274	.008	-0.752	-0.169	0.349	1.000
θ_{new}	0.327	0.409	.012	-0.489	0.321	1.152	1.005
$\theta_{new.\beta1}$	-0.441	0.431	.011	-1.292	-0.436	0.360	1.003
$\theta_{new.\beta2}$	-0.514	0.494	.009	-1.486	-0.511	0.441	1.000
θ_{new}.High	-0.104	0.422	.015	-0.902	-0.116	0.725	1.001
θ_{new}.All	-0.163	0.470	.015	-1.059	-0.164	0.765	1.002
P mu.1 < 0	.008	0.089	.003	0.000	0.000	0.000	1.123
P $\beta_1 < 0$	.972	0.165	.005	0.000	1.000	1.000	1.026
P $\beta_2 < 0$	.956	0.205	.004	0.000	1.000	1.000	1.039
P High Group < 0	.692	0.462	.015	0.000	1.000	1.000	1.000
P All Group < 0	.754	0.431	.014	0.000	1.000	1.000	1.000
P $\theta_{new} < 0$	.197	0.398	.011	0.000	0.000	1.000	1.002
P $\theta_{new.\beta1} < 0$	.850	0.357	.013	0.000	1.000	1.000	1.004
P $\theta_{new.\beta2} < 0$	0.862	0.345	.015	0.000	1.000	1.000	1.000
P θ_{new}.High Group<0	.610	0.488	.015	0.000	1.000	1.000	1.004
P θ_{new}.All Group<0	.648	0.478	.016	0.000	1.000	1.000	1.000
τ	0.361	0.092	.003	0.225	0.347	0.584	1.004
τ^2	0.139	0.078	.002	0.051	0.120	0.341	1.004
deviance	-28.586	6.213	.196	-38.730	-29.295	-14.302	1.001

Detracking (k=20): DuMouchel s_0

DuMouchel s_0			RE ~ normal distribution				DIC = -10.797
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
mu.1	0.359	0.137	.005	0.101	0.356	0.630	1.002
β_1 (High)	-0.480	0.220	.008	-0.903	-0.480	-0.075	1.002
β_2 (All)	-0.538	0.246	.008	-1.022	-0.533	-0.057	1.002
$B_0 + \beta_1$	-0.122	0.169	.007	-0.449	-0.124	0.208	1.000
$B_0 + \beta_2$	-0.179	0.202	.005	-0.588	-0.181	0.206	1.001
θ_{new}	0.347	0.425	.015	-0.496	0.355	1.150	1.006
θ_{new}.High	-0.114	0.430	.014	-0.918	-0.113	0.667	1.002
θ_{new}.All	-0.189	0.442	.014	-1.032	-0.192	0.718	1.001
P mu.1< 0	.005	0.070	.005	0.000	0.000	0.000	1.071
P $\beta_1 < 0$	.986	0.117	.003	1.000	1.000	1.000	1.031
P $\beta_2 <0$	.986	0.117	.004	1.000	1.000	1.000	1.065
P High Group < 0	.768	0.422	.013	0.000	1.000	1.000	1.001
P All Group < 0	.824	0.381	.013	0.000	1.000	1.000	1.000
P $\theta_{new} < 0$	.192	0.394	.014	0.000	0.000	1.000	1.005
P θ_{new}.High Group<0	.604	0.489	.017	0.000	1.000	1.000	1.001
P θ_{new}.All Group<0	.689	0.463	.014	0.000	1.000	1.000	1.001
τ	0.385	0.081	.003	0.259	0.374	0.570	1.000
τ^2	0.155	0.070	.002	0.067	0.140	0.325	1.000
deviance	-27.936	6.141	.183	-38.125	-28.560	-14.381	1.001
DuMouchel s_0			RE ~ t_4 distribution				DIC = -11.199
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
mu.1	0.336	0.146	.004	0.064	0.332	0.636	1.000
β_1 (High)	-0.433	0.236	.006	-0.907	-0.429	0.019	1.000
β_2 (All)	-0.530	0.299	.007	-1.162	-0.525	0.038	1.001
$B_0 + \beta_1$	-0.098	0.191	.008	-0.464	-0.099	0.257	1.000
$B_0 + \beta_2$	-0.194	0.262	.009	-0.741	-0.190	0.321	1.001
θ_{new}	0.321	0.394	.011	-0.440	0.335	1.160	1.000
θ_{new}.High	-0.106	0.434	.015	-1.009	-0.097	0.757	1.002
θ_{new}.All	-0.202	0.451	.013	-1.098	-0.203	0.701	1.002
P mu.1< 0	.008	0.089	.002	0.000	0.000	0.000	1.029
P $\beta_1 < 0$	.973	0.162	.005	0.000	1.000	1.000	1.007
P $\beta_2 <0$	.963	0.189	.004	0.000	1.000	1.000	1.029
P High Group < 0	.698	0.460	.016	0.000	1.000	1.000	1.000
P All Group < 0	.773	0.419	.017	0.000	1.000	1.000	1.000
P $\theta_{new} < 0$	.194	0.395	.011	0.000	0.000	1.000	1.000
P θ_{new}.High Group<0	.594	0.491	.013	0.000	1.000	1.000	1.003
P θ_{new}.All Group<0	.689	0.463	.015	0.000	1.000	1.000	1.000
τ	0.359	0.091	.002	0.216	0.347	0.578	1.002
τ^2	0.137	0.074	.002	0.047	0.121	0.334	1.002
deviance	-28.354	5.888	.236	-38.229	-29.005	-15.099	1.003

Detracking (k=20): DuMouchel $s_0/3$

DuMouchel $s_0/3$			RE ~ normal distribution				DIC = -10.955
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
mu.1	0.362	0.140	.005	0.075	0.359	0.649	1.002
β_1 (High)	-0.485	0.216	.005	-0.936	-0.484	-0.093	1.000
β_2 (All)	-0.559	0.237	.006	-1.057	-0.546	-0.108	1.002
$B_0 + \beta_1$	-0.123	0.170	.007	-0.461	-0.119	0.233	1.000
$B_0 + \beta_2$	-0.198	0.197	.005	-0.596	-0.201	0.193	1.002
θ_{new}	0.360	0.420	.014	-0.496	0.358	1.186	1.007
θ_{new}.High	-0.113	0.435	.014	-0.993	-0.122	0.747	1.002
θ_{new}.All	-0.218	0.454	.014	-1.090	-0.219	0.701	1.001
P mu.1< 0	.005	0.070	.002	0.000	0.000	0.000	1.116
P $\beta_1 < 0$	.990	0.099	.003	1.000	1.000	1.000	1.032
P $\beta_2 < 0$	.992	0.089	.003	1.000	1.000	1.000	1.123
P High Group < 0	.788	0.409	.015	0.000	1.000	1.000	1.000
P All Group < 0	.836	0.370	.011	0.000	1.000	1.000	1.003
P $\theta_{new} < 0$	.160	0.366	.012	0.000	0.000	1.000	1.008
P θ_{new}.High Group<0	.619	0.486	.012	0.000	1.000	1.000	1.000
P θ_{new}.All Group<0	.692	0.462	.015	0.000	1.000	1.000	1.003
τ	0.380	0.080	.002	0.260	0.367	0.562	1.002
τ^2	0.151	0.068	.002	0.068	0.135	0.316	1.002
deviance	-28.070	6.196	.215	-38.349	-28.900	-14.051	1.002
DuMouchel $s_0/3$			RE ~ t_4 distribution				DIC = -10.368
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
mu.1	0.332	0.134	.004	0.093	0.328	0.610	1.000
β_1 (High)	-0.422	0.223	.007	-0.889	-0.418	0.006	1.005
β_2 (All)	-0.512	0.277	.007	-1.073	-0.500	0.004	1.002
$B_0 + \beta_1$	-0.090	0.177	.006	-0.442	-0.082	0.251	1.005
$B_0 + \beta_2$	-0.180	0.245	.007	-0.650	-0.179	0.284	1.001
θ_{new}	0.329	0.355	.011	-0.387	0.331	1.015	1.001
θ_{new}.High	-0.083	0.372	.013	-0.814	-0.083	0.658	1.001
θ_{new}.All	-0.189	0.415	.014	-0.995	-0.195	0.615	1.002
P mu.1< 0	.007	0.083	.003	0.000	0.000	0.000	1.037
P $\beta_1 < 0$	.973	0.162	.006	0.000	1.000	1.000	1.106
P $\beta_2 < 0$	.973	0.162	.004	0.000	1.000	1.000	1.005
P High Group < 0	.690	0.463	.014	0.000	1.000	1.000	1.001
P All Group < 0	.761	0.426	.013	0.000	1.000	1.000	1.000
P $\theta_{new} < 0$	.170	0.376	.011	0.000	0.000	1.000	1.002
P θ_{new}.High Group<0	.589	0.492	.016	0.000	1.000	1.000	1.003
P θ_{new}.All Group<0	.663	0.473	.015	0.000	1.000	1.000	1.002
τ	0.317	0.074	.002	0.195	0.310	0.488	1.002
τ^2	0.106	0.052	.002	0.038	0.096	0.238	1.002
deviance	-27.849	6.129	.183	-37.577	-28.520	-14.682	1.003

Detracking (k=20): DuMouchel $3s_0$

DuMouchel $3s_0$			RE ~ normal distribution				DIC = -10.596
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
mu.1	0.360	0.145	.005	0.072	0.365	0.641	1.007
β_1 (High)	-0.482	0.229	.007	-0.934	-0.492	-0.037	1.010
β_2 (All)	-0.571	0.252	.008	-1.046	-0.569	-0.098	1.003
$B_0 + \beta_1$	-0.122	0.175	.006	-0.460	-0.125	0.244	1.007
$B_0 + \beta_2$	-0.211	0.204	.006	-0.611	-0.212	0.193	1.000
θ_{new}	0.378	0.434	.014	-0.465	0.381	1.222	1.004
θ_{new}.High	-0.105	0.436	.013	-1.005	-0.106	0.791	1.006
θ_{new}.All	-0.218	0.449	.013	-1.086	-0.209	0.697	1.000
P mu.1< 0	.009	0.094	.003	0.000	0.000	0.000	1.051
P $\beta_1 < 0$	.982	0.133	.004	1.000	1.000	1.000	1.068
P $\beta_2 <0$	.987	0.113	.004	1.000	1.000	1.000	1.082
P High Group < 0	.760	0.427	.013	0.000	1.000	1.000	1.006
P All Group < 0	.860	0.347	.010	0.000	1.000	1.000	1.000
P $\theta_{new} < 0$	.181	0.385	.013	0.000	0.000	1.000	1.009
P θ_{new}.High Group<0	.602	0.490	.013	0.000	1.000	1.000	1.004
P θ_{new}.All Group<0	.684	0.465	.013	0.000	1.000	1.000	1.000
τ	0.392	0.084	.002	0.271	0.378	0.590	1.002
τ^2	0.161	0.075	.002	0.073	0.143	0.348	1.002
deviance	-27.907	6.470	.190	-38.518	-28.645	-14.143	1.002
DuMouchel $3s_0$			RE ~ t_4 distribution				DIC = -10.290
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
mu.1	0.329	0.143	.004	0.050	0.327	0.609	1.000
β_1 (High)	-0.427	0.234	.007	-0.908	-0.423	0.006	1.000
β_2 (All)	-0.512	0.291	.009	-1.163	-0.499	0.023	1.001
$B_0 + \beta_1$	-0.098	0.185	.006	-0.482	-0.094	0.248	1.000
$B_0 + \beta_2$	-0.183	0.256	.007	-0.726	-0.169	0.317	1.001
θ_{new}	0.348	0.384	.015	-0.368	0.362	1.104	1.002
θ_{new}.High	-0.104	0.400	.013	-0.993	-0.083	0.675	1.001
θ_{new}.All	-0.183	0.453	.013	-1.070	-0.178	0.727	1.000
P mu.1< 0	.009	0.094	.003	0.000	0.000	0.000	1.017
P $\beta_1 < 0$	.972	0.165	.005	0.000	1.000	1.000	1.017
P $\beta_2 <0$	.968	0.176	.005	0.000	1.000	1.000	1.008
P High Group < 0	.715	0.452	.012	0.000	1.000	1.000	1.000
P All Group < 0	.782	0.413	.013	0.000	1.000	1.000	1.000
P $\theta_{new} < 0$	.175	0.380	.012	0.000	0.000	1.000	1.001
P θ_{new}.High Group<0	.607	0.489	.017	0.000	1.000	1.000	1.005
P θ_{new}.All Group<0	.651	0.477	.013	0.000	1.000	1.000	1.000
τ	0.332	0.081	.002	0.201	0.323	0.510	1.002
τ^2	0.117	0.060	.002	0.040	0.104	0.260	1.002
deviance	-27.926	6.457	.211	-38.500	-28.600	-14.301	1.002

Detracking (k=20): Inverse Gamma on precision

Inverse Gamma			RE ~ normal distribution				DIC = -10.003
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
mu.1	0.352	0.148	.005	0.051	0.349	0.659	1.004
β_1 (High)	-0.482	0.234	.006	-0.928	-0.476	-0.021	1.001
β_2 (All)	-0.540	0.253	.009	-1.036	-0.543	-0.034	1.001
$B_0 + \beta_1$	-0.130	0.179	.005	-0.475	-0.125	0.213	1.008
$B_0 + \beta_2$	-0.187	0.209	.007	-0.606	-0.188	0.212	1.000
θ_{new}	0.361	0.432	.016	-0.432	0.338	1.209	1.007
θ_{new}.High	-0.126	0.456	.013	-1.036	-0.112	0.813	1.003
θ_{new}.All	-0.198	0.465	.013	-1.128	-0.209	0.720	1.002
P mu.1< 0	.012	0.109	.003	0.000	0.000	0.000	1.083
P $\beta_1 < 0$	.980	0.140	.004	1.000	1.000	1.000	1.044
P $\beta_2 <0$	.985	0.121	.004	1.000	1.000	1.000	1.018
P High Group < 0	.779	0.415	.014	0.000	1.000	1.000	1.016
P All Group < 0	.827	0.378	.013	0.000	1.000	1.000	1.001
P $\theta_{new} < 0$	.196	0.397	.013	0.000	0.000	1.000	1.006
P θ_{new}.High Group<0	.617	0.486	.013	0.000	1.000	1.000	1.004
P θ_{new}.All Group<0	.672	0.470	.014	0.000	1.000	1.000	1.007
τ	0.402	0.086	.003	0.264	0.388	0.586	1.000
τ^2	0.169	0.076	.003	0.070	0.150	0.343	1.000
deviance	-27.698	6.537	.211	-37.610	-28.620	-12.460	1.000
Inverse Gamma			RE ~ t_4 distribution				DIC = -10.611
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
mu.1	0.326	0.147	.005	0.039	0.327	0.614	1.000
β_1 (High)	-0.417	0.244	.007	-0.915	-0.406	0.024	1.002
β_2 (All)	-0.512	0.309	.010	-1.126	-0.506	0.096	1.007
$B_0 + \beta_1$	-0.091	0.190	.006	-0.475	-0.093	0.268	1.000
$B_0 + \beta_2$	-0.186	0.266	.010	-0.727	-0.180	0.337	1.006
θ_{new}	0.331	0.392	.013	-0.497	0.323	1.103	1.004
θ_{new}.High	-0.094	0.404	.013	-0.896	-0.112	0.724	1.001
θ_{new}.All	-0.184	0.443	.015	-1.027	-0.167	0.668	1.005
P mu.1< 0	.009	0.094	.003	0.000	0.000	0.000	1.051
P $\beta_1 < 0$	.964	0.186	.006	0.000	1.000	1.000	1.002
P $\beta_2 <0$	.954	0.209	.007	0.000	1.000	1.000	1.062
P High Group < 0	.681	0.466	.013	0.000	1.000	1.000	1.000
P All Group < 0	.756	0.429	.014	0.000	1.000	1.000	1.003
P $\theta_{new} < 0$	.172	0.377	.013	0.000	0.000	1.000	1.001
P θ_{new}.High Group<0	.612	0.488	.013	0.000	1.000	1.000	1.002
P θ_{new}.All Group<0	.660	0.474	.015	0.000	1.000	1.000	1.000
τ	0.343	0.085	.003	0.219	0.331	0.537	1.001
τ^2	0.125	0.069	.003	0.048	0.110	0.289	1.001
deviance	-28.178	6.141	.180	-38.527	-28.800	-15.552	1.000

Detracking (k=20): Half-normal

Half-normal			RE ~ normal distribution				DIC = -10.478
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
mu.1	0.367	0.158	.004	0.049	0.366	0.683	1.002
β_1 (High)	-0.497	0.245	.007	-1.000	-0.492	-0.010	1.001
β_2 (All)	-0.549	0.272	.008	-1.109	-0.558	0.022	1.005
$B_0 + \beta_1$	-0.130	0.191	.006	-0.510	-0.133	0.243	1.004
$B_0 + \beta_2$	-0.182	0.223	.007	-0.628	-0.176	0.274	1.003
θ_{new}	0.335	0.483	.013	-0.613	0.306	1.332	1.003
θ_{new}.High	-0.132	0.467	.015	-1.078	-0.136	0.787	1.000
θ_{new}.All	-0.214	0.507	.016	-1.191	-0.210	0.776	1.003
P mu.1< 0	.010	0.099	.003	0.000	0.000	0.000	1.004
P $\beta_1 < 0$	.975	0.156	.005	1.000	1.000	1.000	1.004
P $\beta_2 <0$	.973	0.162	.006	0.000	1.000	1.000	1.018
P High Group < 0	.767	0.423	.013	0.000	1.000	1.000	1.006
P All Group < 0	.811	0.391	.013	0.000	1.000	1.000	1.001
P $\theta_{new} < 0$	.230	0.421	.012	0.000	0.000	1.000	1.001
P θ_{new}.High Group<0	.616	0.487	.012	0.000	1.000	1.000	1.000
P θ_{new}.All Group<0	.687	0.464	.015	0.000	1.000	1.000	1.000
τ	0.434	0.101	.003	0.280	0.418	0.685	1.003
τ^2	0.198	0.099	.003	0.078	0.175	0.469	1.003
deviance	-28.190	6.387	.238	-38.709	-28.655	-13.812	1.001
Half-normal			RE ~ t_4 distribution				DIC = -10.074
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
mu.1	0.335	0.151	.005	0.029	0.332	0.653	1.004
β_1 (High)	-0.435	0.246	.008	-0.923	-0.427	0.008	1.002
β_2 (All)	-0.521	0.310	.010	-1.154	-0.521	0.081	1.002
$B_0 + \beta_1$	-0.099	0.200	.006	-0.498	-0.100	0.292	1.001
$B_0 + \beta_2$	-0.186	0.273	.009	-0.753	-0.183	0.326	1.002
θ_{new}	0.328	0.428	.012	-0.515	0.330	1.186	1.001
θ_{new}.High	-0.098	0.457	.015	-1.008	-0.102	0.798	1.003
θ_{new}.All	-0.187	0.471	.016	-1.115	-0.194	0.754	1.003
P mu.1< 0	.013	0.113	.003	0.000	0.000	0.000	1.036
P $\beta_1 < 0$	.967	0.179	.006	0.000	1.000	1.000	1.009
P $\beta_2 <0$	.960	0.196	.006	0.000	1.000	1.000	1.016
P High Group < 0	.695	0.461	.014	0.000	1.000	1.000	1.001
P All Group < 0	.748	0.435	.015	0.000	1.000	1.000	1.001
P $\theta_{new} < 0$	.198	0.398	.013	0.000	0.000	1.000	1.002
P θ_{new}.High Group<0	.606	0.489	.017	0.000	1.000	1.000	1.008
P θ_{new}.All Group<0	.669	0.471	.015	0.000	1.000	1.000	1.001
τ	0.383	0.100	.003	0.233	0.371	0.625	1.001
τ^2	0.157	0.087	.003	0.054	0.137	0.390	1.001
deviance	-28.001	6.148	.207	-38.310	-28.585	-14.595	1.002

Detracking (k=20) : Uniform on $\boldsymbol{\tau^2}$

Uniform on τ^2			RE ~ normal distribution				DIC = -10.438
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
mu.1	0.364	0.157	.005	0.049	0.361	0.674	1.001
β_1 (High)	-0.483	0.248	.008	-0.982	-0.484	0.024	1.001
β_2 (All)	-0.565	0.277	.009	-1.102	-0.564	-0.017	1.002
B_0 + β_1	-0.119	0.188	.006	-0.485	-0.117	0.259	1.002
B_0 + β_2	-0.202	0.224	.008	-0.628	-0.208	0.273	1.002
θ_{new}	0.379	0.460	.016	-0.481	0.380	1.320	0.999
θ_{new}.High	-0.129	0.480	.012	-1.134	-0.104	0.760	1.002
θ_{new}.All	-0.192	0.500	.016	-1.113	-0.214	0.794	1.000
P mu.1< 0	.007	0.083	.003	0.000	0.000	0.000	1.037
P β_1 < 0	.973	0.162	.005	0.000	1.000	1.000	1.001
P β_2 <0	.976	0.153	.005	1.000	1.000	1.000	1.007
P High Group < 0	.741	0.439	.013	0.000	1.000	1.000	1.001
P All Group < 0	.821	0.383	.011	0.000	1.000	1.000	1.004
P θ_{new} < 0	.196	0.397	.011	0.000	0.000	1.000	1.001
P θ_{new}.High Group<0	.609	0.488	.014	0.000	1.000	1.000	1.002
P θ_{new}.All Group<0	.676	0.468	.013	0.000	1.000	1.000	1.000
τ	0.428	0.095	.003	0.286	0.412	0.667	1.000
τ^2	0.192	0.091	.003	0.082	0.170	0.445	1.000
deviance	-28.193	6.138	.200	-38.699	-28.835	-14.485	1.006
Uniform on τ^2			RE ~ t_4 distribution				DIC = -10.563
	mean	s.d.	MCerror	2.50%	50%	97.50%	Rhat
mu.1	0.338	0.153	.004	0.049	0.332	0.672	1.000
β_1 (High)	-0.438	0.251	.006	-0.964	-0.424	0.038	1.006
β_2 (All)	-0.539	0.312	.010	-1.166	-0.532	0.020	1.003
B_0 + β_1	-0.100	0.194	.005	-0.492	-0.103	0.267	1.006
B_0 + β_2	-0.202	0.272	.009	-0.759	-0.195	0.292	1.006
θ_{new}	0.351	0.432	.014	-0.499	0.332	1.235	1.000
θ_{new}.High	-0.114	0.429	.013	-0.975	-0.103	0.704	1.001
θ_{new}.All	-0.227	0.479	.013	-1.122	-0.227	0.704	1.010
P mu.1< 0	.015	0.121	.004	0.000	0.000	0.000	1.044
P β_1 < 0	.964	0.186	.006	0.000	1.000	1.000	1.013
P β_2 <0	.964	0.186	.006	0.000	1.000	1.000	1.021
P High Group < 0	.720	0.449	.014	0.000	1.000	1.000	1.011
P All Group < 0	.771	0.420	.013	0.000	1.000	1.000	1.001
P θ_{new} < 0	.202	0.401	.013	0.000	0.000	1.000	1.001
P θ_{new}.High Group<0	.610	0.488	.015	0.000	1.000	1.000	1.001
P θ_{new}.All Group<0	.690	0.463	.014	0.000	1.000	1.000	1.008
τ	0.381	0.090	.003	0.237	0.370	0.583	1.003
τ^2	0.153	0.075	.003	0.056	0.137	0.340	1.003
deviance	-28.346	6.227	.212	-38.840	-28.795	-14.472	1.000

Appendix F. Chondroitin (*k*=18) Meta-analysis: Bayesian summary results

Chondroitin (*k*=18): Uniform on $\boldsymbol{\tau}$ (0,5)

Uniform on τ (0,5)		RE ~ *Normal* distribution		DIC = 3.190		
Node	Mean	s.d.	2.5%	50%	97.5%	Rhat
β_0	-0.322	0.210	-0.755	-0.315	0.092	1.004
β_5	-0.512	0.245	-1.025	-0.513	-0.031	1.002
B_6	0.008	0.004	0.001	0.008	0.015	1.008
τ	0.440	0.102	0.278	0.427	0.672	1.001
τ^2	0.204	0.099	0.077	0.182	0.452	1.001
$\beta_0 + \beta_5$	-0.833	0.130	-1.076	-0.837	-0.596	1.002
θ_{new}	-0.302	0.495	-1.234	-0.315	0.687	1.002
$\theta_{new.\ \beta0 + \beta5}$	-0.839	0.483	-1.810	-0.829	0.150	1.001
$\theta_{new.\ \beta5}$	-0.495	0.512	-1.486	-0.511	0.563	1.000
$\theta_{new.\ \beta6}$	-0.011	0.442	-0.856	-0.014	0.892	1.002
p $\beta_5 < 0$	.979	0.143	1.000	1.000	1.000	1.009
p $\beta_6 < 0$	.016	0.125	0.000	0.000	0.000	1.006
p $\beta_0 < 0$	.943	0.232	0.000	1.000	1.000	1.014
p $\beta_0 + \beta_5 < 0$	.999	0.000	1.000	1.000	1.000	1.000
p $\theta_{new} < 0$	.739	0.440	0.000	1.000	1.000	1.004
p $\theta_{new.B5} < 0$	.833	0.373	0.000	1.000	1.000	1.000
p $\theta_{new.B6} < 0$	.496	0.500	0.000	0.000	1.000	1.000
p $\theta_{new.\beta0 + \beta5} < 0$	.956	0.205	0.000	1.000	1.000	1.007
deviance	-12.688	6.042	-22.698	-13.290	0.587	1.001
Uniform on τ (0,5)		RE ~ t_4 distribution		DIC = 2.8		
Node	Mean	s.d.	2.5%	50%	97.5%	Rhat
β_0	-0.311	0.208	-0.719	-0.321	0.086	1.000
β_5	-0.452	0.251	-0.961	-0.448	0.033	1.000
B_6	0.007	0.004	0.000	0.007	0.015	1.001
τ	0.358	0.110	0.204	0.341	0.619	1.002
τ^2	0.140	0.098	0.041	0.117	0.383	1.002
$\beta_0 + \beta_5$	-0.763	0.138	-1.034	-0.759	-0.490	1.002
θ_{new}	-0.311	0.457	-1.161	-0.306	0.602	1.000
$\theta_{new.\ \beta0 + \beta5}$	-0.769	0.385	-1.536	-0.759	-0.018	1.001
$\theta_{new.\ \beta5}$	-0.444	0.460	-1.401	-0.462	0.498	1.002
$\theta_{new.\ \beta6}$	0.002	0.371	-0.813	0.002	0.701	1.001
p $\beta_5 < 0$	.969	0.173	0.000	1.000	1.000	1.006
p $\beta_6 < 0$	.027	0.162	0.000	0.000	1.000	1.001
p $\beta_0 < 0$	.941	0.236	0.000	1.000	1.000	1.006
p $\beta_0 + \beta_5 < 0$	.999	0.000	1.000	1.000	1.000	1.000
p $\theta_{new} < 0$	.792	0.406	0.000	1.000	1.000	1.000
p $\theta_{new.B5} < 0$	.838	0.368	0.000	1.000	1.000	1.001
p $\theta_{new.B6} < 0$	.498	0.500	0.000	0.000	1.000	1.000
p $\theta_{new.\beta0 + \beta5} < 0$	.978	0.147	1.000	1.000	1.000	1.031
deviance	-13.039	6.266	-22.977	-13.755	1.178	1.001

Chondroitin (*k*=18): DuMouchel s_0

DuMouchel s_0		RE ~ Normal distribution		DIC = 2.779		
Node	Mean	s.d.	2.5%	50%	97.5%	Rhat
β_0	-0.316	0.192	-0.707	-0.312	0.055	1.005
β_5	-0.505	0.230	-0.972	-0.503	-0.048	1.006
B_6	0.008	0.003	0.001	0.008	0.015	1.001
τ	0.407	0.096	0.255	0.392	0.620	1.004
τ^2	0.175	0.087	0.065	0.154	0.384	1.004
$\beta_0 + \beta_5$	-0.821	0.127	-1.071	-0.818	-0.578	1.009
θ_{new}	-0.314	0.471	-1.238	-0.309	0.624	1.002
$\theta_{new.\ \beta0 + \beta5}$	-0.825	0.431	-1.714	-0.816	0.033	1.001
$\theta_{new.\ \beta5}$	-0.513	0.471	-1.488	-0.506	0.495	1.004
$\theta_{new.\ \beta6}$	0.038	0.424	-0.706	0.019	0.971	1.007
p $\beta_5 < 0$	.983	0.129	1.000	1.000	1.000	1.050
p $\beta_6 < 0$	.014	0.117	0.000	0.000	0.000	1.016
p $\beta_0 < 0$	.950	0.218	0.000	1.000	1.000	1.020
p $\beta_0 + \beta_5 < 0$	.999	0.000	1.000	1.000	1.000	1.000
p $\theta_{new} < 0$	.766	0.423	0.000	1.000	1.000	1.001
p $\theta_{new.B5} < 0$	.877	0.328	0.000	1.000	1.000	1.010
p $\theta_{new.B6} < 0$	.485	0.500	0.000	0.000	1.000	1.005
p $\theta_{new.\beta0 + \beta5} < 0$	.973	0.162	0.000	1.000	1.000	1.042
deviance	-12.628	5.871	-22.409	-13.045	-0.047	1.000
DuMouchel s_0		RE ~ t_4 distribution		DIC = 3.237		
Node	Mean	s.d.	2.5%	50%	97.5%	Rhat
β_0	-0.315	0.192	-0.702	-0.312	0.069	1.002
β_5	-0.443	0.230	-0.919	-0.433	0.009	1.003
B_6	0.007	0.003	0.001	0.007	0.014	1.002
τ	0.321	0.090	0.176	0.312	0.517	1.000
τ^2	0.111	0.068	0.031	0.097	0.267	1.000
$\beta_0 + \beta_5$	-0.758	0.127	-1.026	-0.753	-0.513	1.000
θ_{new}	-0.298	0.381	-1.019	-0.302	0.486	1.002
$\theta_{new.\ \beta0 + \beta5}$	-0.764	0.356	-1.488	-0.758	-0.112	1.001
$\theta_{new.\ \beta5}$	-0.430	0.417	-1.240	-0.441	0.426	1.000
$\theta_{new.\ \beta6}$	0.008	0.349	-0.720	0.011	0.673	1.005
p $\beta_5 < 0$	.973	0.162	0.000	1.000	1.000	1.038
p $\beta_6 < 0$	.017	0.129	0.000	0.000	0.000	1.050
p $\beta_0 < 0$	.951	0.216	0.000	1.000	1.000	1.000
p $\beta_0 + \beta_5 < 0$	.999	0.000	1.000	1.000	1.000	1.000
p $\theta_{new} < 0$	.798	0.401	0.000	1.000	1.000	1.000
p $\theta_{new.B5} < 0$	.858	0.349	0.000	1.000	1.000	1.000
p $\theta_{new.B6} < 0$	.482	0.500	0.000	0.000	1.000	1.004
p $\theta_{new.\beta0 + \beta5} < 0$	.989	0.104	1.000	1.000	1.000	1.050
deviance	-12.471	6.215	-22.379	-13.210	1.767	1.001

Chondroitin (*k*=18): DuMouchel $s_0/3$

DuMouchel $s_0/3$		RE ~ Normal distribution		DIC = 2.586		
Node	Mean	s.d.	2.5%	50%	97.5%	Rhat
β_0	-0.311	0.199	-0.713	-0.308	0.068	1.003
β_5	-0.512	0.235	-0.990	-0.507	-0.072	1.005
B_6	0.008	0.003	0.001	0.008	0.015	1.002
τ	0.402	0.097	0.249	0.386	0.642	1.001
τ^2	0.171	0.090	0.062	0.149	0.412	1.001
$\beta_0 + \beta_5$	-0.823	0.128	-1.063	-0.828	-0.567	1.002
θ_{new}	-0.300	0.460	-1.164	-0.313	0.614	1.009
$\theta_{new.\ \beta0 + \beta5}$	-0.832	0.438	-1.689	-0.815	0.012	1.000
$\theta_{new.\ \beta5}$	-0.536	0.460	-1.452	-0.527	0.319	1.007
$\theta_{new.\ \beta6}$	0.008	0.415	-0.813	0.032	0.798	1.001
p $\beta_5 < 0$	.983	0.129	1.000	1.000	1.000	1.114
p $\beta_6 < 0$	.005	0.070	0.000	0.000	0.000	1.071
p $\beta_0 < 0$	.953	0.212	0.000	1.000	1.000	1.000
p $\beta_0 + \beta_5 < 0$	1.000	0.000	1.000	1.000	1.000	1.000
p $\theta_{new} < 0$	.753	0.431	0.000	1.000	1.000	1.001
p $\theta_{new.B5} < 0$	.892	0.310	0.000	1.000	1.000	1.016
p $\theta_{new.B6} < 0$	.464	0.499	0.000	0.000	1.000	1.001
p $\theta_{new.\beta0 + \beta5} < 0$	.971	0.168	0.000	1.000	1.000	1.010
deviance	-12.691	6.118	-22.576	-13.220	1.495	1.003
DuMouchel $s_0/3$		RE ~ t_4 distribution		DIC = 3.312		
Node	Mean	s.d.	2.5%	50%	97.5%	Rhat
β_0	-0.315	0.181	-0.660	-0.316	0.061	1.001
β_5	-0.440	0.218	-0.935	-0.434	-0.046	1.002
B_6	0.007	0.003	0.001	0.007	0.014	1.000
τ	0.311	0.088	0.179	0.298	0.517	1.001
τ^2	0.104	0.066	0.032	0.089	0.267	1.001
$\beta_0 + \beta_5$	-0.755	0.129	-1.015	-0.757	-0.498	1.000
θ_{new}	-0.321	0.366	-1.006	-0.342	0.425	1.000
$\theta_{new.\ \beta0 + \beta5}$	-0.748	0.360	-1.488	-0.746	-0.111	1.000
$\theta_{new.\ \beta5}$	-0.446	0.390	-1.290	-0.436	0.305	1.001
$\theta_{new.\ \beta6}$	0.012	0.321	-0.637	0.010	0.630	1.001
p $\beta_5 < 0$	.987	0.113	1.000	1.000	1.000	1.036
p $\beta_6 < 0$	.006	0.077	0.000	0.000	0.000	1.038
p $\beta_0 < 0$	.953	0.212	0.000	1.000	1.000	1.003
p $\beta_0 + \beta_5 < 0$	.999	0.000	1.000	1.000	1.000	1.000
p $\theta_{new} < 0$	.814	0.389	0.000	1.000	1.000	1.001
p $\theta_{new.B5} < 0$	.894	0.308	0.000	1.000	1.000	1.000
p $\theta_{new.B6} < 0$	.482	0.500	0.000	0.000	1.000	1.000
p $\theta_{new.\beta0 + \beta5} < 0$	.983	0.129	1.000	1.000	1.000	1.000
deviance	-12.332	6.205	-22.599	-12.910	1.341	1.000

Chondroitin (*k*=18): DuMouchel $3s_0$

DuMouchel $3s_0$		RE ~ Normal distribution		DIC = 2.786		
Node	Mean	s.d.	2.5%	50%	97.5%	Rhat
β_0	-0.313	0.203	-0.694	-0.313	0.098	1.000
β_5	-0.515	0.240	-1.006	-0.511	-0.034	1.001
B_6	0.008	0.004	0.001	0.008	0.015	1.004
τ	0.415	0.099	0.268	0.402	0.644	1.005
τ^2	0.182	0.092	0.072	0.162	0.415	1.005
$\beta_0 + \beta_5$	-0.829	0.128	-1.102	-0.830	-0.582	1.000
θ_{new}	-0.309	0.476	-1.272	-0.296	0.620	1.001
$\theta_{new.\ \beta0 + \beta5}$	-0.822	0.469	-1.732	-0.818	0.105	1.000
$\theta_{new.\ \beta5}$	-0.537	0.485	-1.620	-0.532	0.401	1.000
$\theta_{new.\ \beta6}$	0.035	0.400	-0.735	0.031	0.853	1.002
p $\beta_5 < 0$	.982	0.133	1.000	1.000	1.000	1.031
p $\beta_6 < 0$	.019	0.136	0.000	0.000	0.000	1.004
p $\beta_0 < 0$	.947	0.224	0.000	1.000	1.000	1.000
p $\beta_0 + \beta_5 < 0$	.999	0.000	1.000	1.000	1.000	1.000
p $\theta_{new} < 0$	.746	0.436	0.000	1.000	1.000	1.002
p $\theta_{new.B5} < 0$	.881	0.324	0.000	1.000	1.000	1.003
p $\theta_{new.B6} < 0$	.471	0.499	0.000	0.000	1.000	1.003
p $\theta_{new.\beta0 + \beta5} < 0$	.959	0.198	0.000	1.000	1.000	1.000
deviance	-12.652	6.192	-22.679	-13.325	0.954	1.002
DuMouchel $3s_0$		RE ~ t_4 distribution		DIC = 2.728		
Node	Mean	s.d.	2.5%	50%	97.5%	Rhat
β_0	-0.317	0.191	-0.687	-0.313	0.050	1.000
β_5	-0.448	0.232	-0.918	-0.447	-0.016	1.000
B_6	0.007	0.003	0.001	0.007	0.014	1.000
τ	0.335	0.094	0.185	0.323	0.543	1.001
τ^2	0.121	0.073	0.034	0.105	0.295	1.001
$\beta_0 + \beta_5$	-0.764	0.130	-1.024	-0.759	-0.532	1.000
θ_{new}	-0.329	0.388	-1.126	-0.327	0.427	1.001
$\theta_{new.\ \beta0 + \beta5}$	-0.755	0.359	-1.493	-0.754	-0.060	1.004
$\theta_{new.\ \beta5}$	-0.445	0.404	-1.268	-0.451	0.327	1.000
$\theta_{new.\ \beta6}$	0.015	0.344	-0.721	0.011	0.665	1.000
p $\beta_5 < 0$	.976	0.153	1.000	1.000	1.000	1.017
p $\beta_6 < 0$	.015	0.121	0.000	0.000	0.000	1.043
p $\beta_0 < 0$	.957	0.203	0.000	1.000	1.000	1.000
p $\beta_0 + \beta_5 < 0$	.999	0.000	1.000	1.000	1.000	1.000
p $\theta_{new} < 0$	.810	0.392	0.000	1.000	1.000	1.012
p $\theta_{new.B5} < 0$	.873	0.333	0.000	1.000	1.000	1.000
p $\theta_{new.B6} < 0$	.483	0.500	0.000	0.000	1.000	1.000
p $\theta_{new.\beta0 + \beta5} < 0$	.978	0.147	1.000	1.000	1.000	1.015
deviance	-12.858	5.934	-22.788	-13.525	0.737	1.005

Chondroitin (*k*=18): Gamma on Precision

Gamma on Precision		RE ~ Normal distribution		DIC = 3.1		
Node	Mean	s.d.	2.5%	50%	97.5%	Rhat
β_0	-0.318	0.200	-0.704	-0.321	0.125	1.002
β_5	-0.503	0.240	-0.975	-0.507	-0.050	1.001
B_6	0.008	0.003	0.001	0.008	0.015	1.001
τ	0.420	0.099	0.263	0.405	0.645	1.000
τ^2	0.186	0.093	0.069	0.164	0.416	1.000
$\beta_0 + \beta_5$	-0.821	0.133	-1.101	-0.824	-0.549	1.000
θ_{new}	-0.334	0.490	-1.305	-0.321	0.625	1.015
$\theta_{new.\ \beta0 + \beta5}$	-0.846	0.442	-1.656	-0.858	0.004	1.001
$\theta_{new.\ \beta5}$	-0.500	0.517	-1.553	-0.505	0.583	1.003
$\theta_{new.\ \beta6}$	0.003	0.440	-0.855	-0.004	0.893	1.001
p $\beta_5 < 0$	.980	0.140	1.000	1.000	1.000	1.004
p $\beta_6 < 0$	.013	0.113	0.000	0.000	0.000	1.010
p $\beta_0 < 0$	.950	0.218	0.000	1.000	1.000	1.018
p $\beta_0 + \beta_5 < 0$	.999	0.000	1.000	1.000	1.000	1.000
p $\theta_{new} < 0$	.764	0.425	0.000	1.000	1.000	1.006
p $\theta_{new.B5} < 0$	.863	0.344	0.000	1.000	1.000	1.000
p $\theta_{new.B6} < 0$	.503	0.500	0.000	1.000	1.000	1.000
p $\theta_{new.\beta0 + \beta5} < 0$	.974	0.159	0.025	1.000	1.000	1.009
deviance	-12.554	6.227	-22.529	-13.350	1.362	1.008
Gamma on Precision		RE ~ t_4 distribution		DIC = 2.867		
Node	Mean	s.d.	2.5%	50%	97.5%	Rhat
β_0	-0.315	0.189	-0.688	-0.312	0.065	1.002
β_5	-0.450	0.236	-0.918	-0.442	0.001	1.003
B_6	0.007	0.003	0.001	0.008	0.014	1.002
τ	0.340	0.098	0.191	0.330	0.567	1.004
τ^2	0.125	0.079	0.036	0.109	0.321	1.004
$\beta_0 + \beta_5$	-0.765	0.134	-1.071	-0.757	-0.517	1.003
θ_{new}	-0.298	0.404	-1.061	-0.300	0.527	1.003
$\theta_{new.\ \beta0 + \beta5}$	-0.767	0.370	-1.539	-0.770	-0.026	1.002
$\theta_{new.\ \beta5}$	-0.450	0.420	-1.273	-0.443	0.329	1.005
$\theta_{new.\ \beta6}$	0.010	0.344	-0.684	0.000	0.711	1.005
p $\beta_5 < 0$	.974	0.159	0.025	1.000	1.000	1.004
p $\beta_6 < 0$	.013	0.113	0.000	0.000	0.000	1.036
p $\beta_0 < 0$	.956	0.205	0.000	1.000	1.000	1.004
p $\beta_0 + \beta_5 < 0$	.999	0.000	1.000	1.000	1.000	1.000
p $\theta_{new} < 0$	.782	0.413	0.000	1.000	1.000	1.000
p $\theta_{new.B5} < 0$	.859	0.348	0.000	1.000	1.000	1.003
p $\theta_{new.B6} < 0$	.500	0.500	0.000	0.500	1.000	1.000
p $\theta_{new.\beta0 + \beta5} < 0$	.977	0.150	1.000	1.000	1.000	1.039
deviance	-12.715	6.179	-22.599	-13.320	0.355	1.000

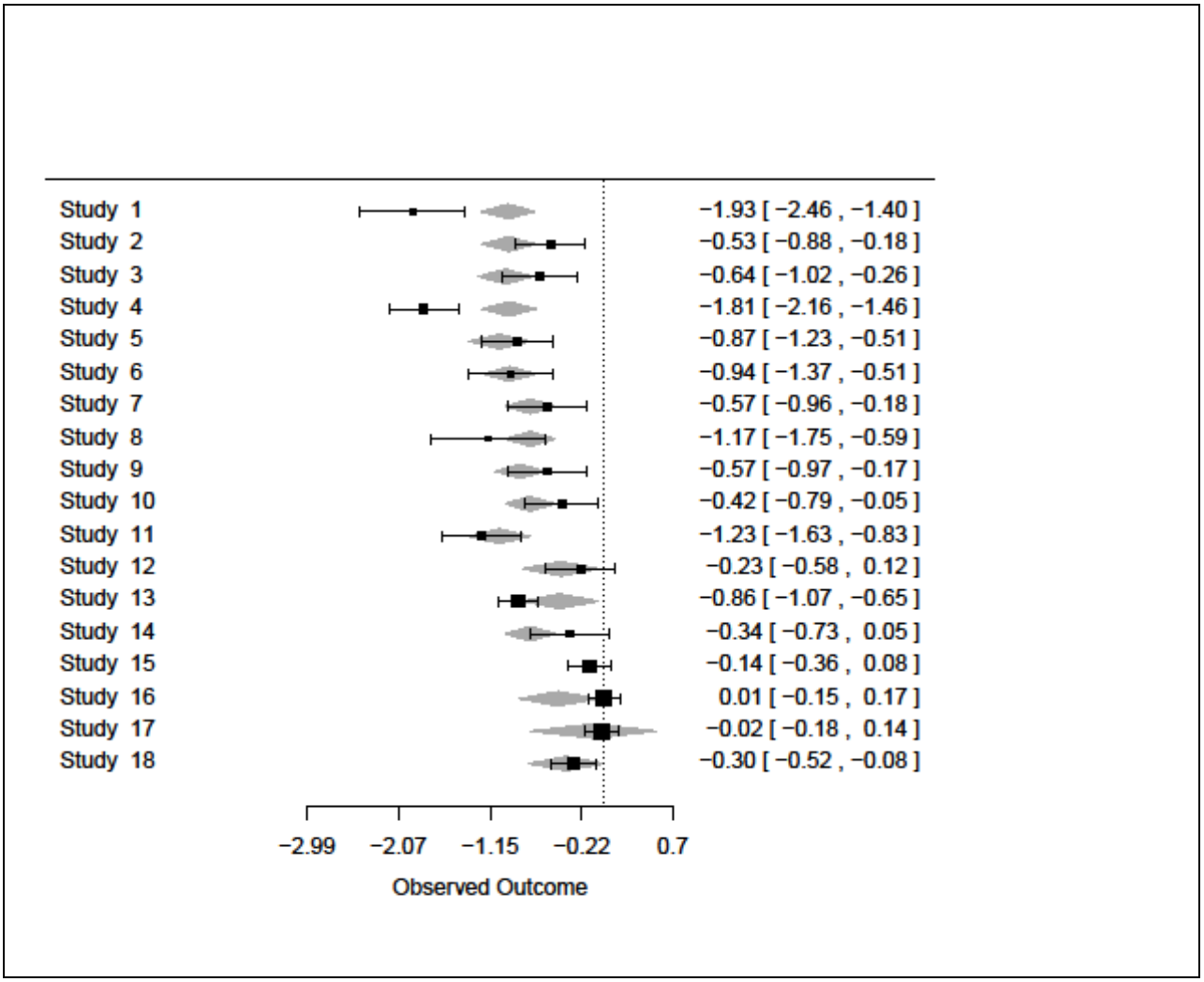
Chondroitin (*k*=18): Half-normal

Half-normal		RE ~ Normal distribution		DIC = 3.294		
Node	Mean	s.d.	2.5%	50%	97.5%	Rhat
β_0	-0.306	0.229	-0.737	-0.318	0.161	1.002
β_5	-0.525	0.279	-1.122	-0.510	0.022	1.001
B_6	0.008	0.004	-0.001	0.008	0.015	1.000
τ	0.469	0.111	0.294	0.458	0.725	1.000
τ^2	0.232	0.115	0.087	0.210	0.526	1.000
$\beta_0 + \beta_5$	-0.831	0.147	-1.107	-0.827	-0.545	1.000
θ_{new}	-0.302	0.551	-1.395	-0.287	0.819	1.002
$\theta_{new.\ \beta0 + \beta5}$	-0.847	0.522	-1.913	-0.835	0.231	1.002
$\theta_{new.\ \beta5}$	-0.522	0.558	-1.612	-0.522	0.637	1.000
$\theta_{new.\ \beta6}$	0.039	0.472	-0.832	0.030	0.965	1.004
p $\beta_5 < 0$	.972	0.165	0.000	1.000	1.000	1.003
p $\beta_6 < 0$	.033	0.179	0.000	0.000	1.000	0.999
p $\beta_0 < 0$	.907	0.290	0.000	1.000	1.000	1.008
p $\beta_0 + \beta_5 < 0$	.999	0.000	1.000	1.000	1.000	1.000
p $\theta_{new} < 0$	.726	0.446	0.000	1.000	1.000	1.001
p $\theta_{new.B5} < 0$	.836	0.370	0.000	1.000	1.000	1.002
p $\theta_{new.B6} < 0$	.474	0.500	0.000	0.000	1.000	1.002
p $\theta_{new.\beta0 + \beta5} < 0$	.949	0.220	0.000	1.000	1.000	1.012
deviance	-12.877	5.951	-22.810	-13.515	1.598	1.002
Half-normal		RE ~ t_4 distribution		DIC = 2.976		
Node	Mean	s.d.	2.5%	50%	97.5%	Rhat
β_0	-0.310	0.216	-0.732	-0.310	0.118	1.004
β_5	-0.462	0.253	-0.963	-0.460	0.039	1.002
B_6	0.007	0.004	0.000	0.007	0.014	1.006
τ	0.389	0.115	0.217	0.370	0.655	1.002
τ^2	0.165	0.110	0.047	0.137	0.429	1.002
$\beta_0 + \beta_5$	-0.772	0.144	-1.062	-0.767	-0.506	1.002
θ_{new}	-0.310	0.474	-1.250	-0.321	0.607	1.002
$\theta_{new.\ \beta0 + \beta5}$	-0.766	0.416	-1.602	-0.769	0.057	1.001
$\theta_{new.\ \beta5}$	-0.460	0.469	-1.424	-0.461	0.389	1.000
$\theta_{new.\ \beta6}$	0.019	0.405	-0.761	0.026	0.845	1.004
p $\beta_5 < 0$	.967	0.179	0.000	1.000	1.000	1.009
p $\beta_6 < 0$	.027	0.162	0.000	0.000	1.000	1.032
p $\beta_0 < 0$	.934	0.248	0.000	1.000	1.000	1.014
p $\beta_0 + \beta_5 < 0$	.999	0.000	1.000	1.000	1.000	1.000
p $\theta_{new} < 0$	.756	0.429	0.000	1.000	1.000	1.003
p $\theta_{new.B5} < 0$	.843	0.364	0.000	1.000	1.000	1.002
p $\theta_{new.B6} < 0$	.478	0.500	0.000	0.000	1.000	1.002
p $\theta_{new.\beta0 + \beta5} < 0$	.964	0.186	0.000	1.000	1.000	1.000
deviance	-13.092	6.003	-22.858	-13.715	-0.419	1.000

Chondroitin(*k*=18): Uniform on $\boldsymbol{\tau^2}$

Uniform on τ^2		RE ~ Normal distribution		DIC = 2.460		
Node	Mean	s.d.	2.5%	50%	97.5%	Rhat
β_0	-0.325	0.227	-0.761	-0.328	0.092	1.002
β_5	-0.499	0.266	-1.025	-0.501	0.029	1.003
B_6	0.008	0.004	0.000	0.008	0.016	1.004
τ	0.474	0.122	0.295	0.453	0.748	1.000
τ^2	0.239	0.135	0.087	0.205	0.559	1.000
$\beta_0 + \beta_5$	-0.824	0.147	-1.109	-0.820	-0.540	1.000
θ_{new}	-0.348	0.534	-1.390	-0.350	0.707	1.000
$\theta_{new.\ \beta0 + \beta5}$	-0.806	0.522	-1.863	-0.805	0.250	1.004
$\theta_{new.\ \beta5}$	-0.507	0.561	-1.648	-0.504	0.576	1.001
$\theta_{new.\ \beta6}$	-0.004	0.479	-0.993	0.012	0.945	1.003
p $\beta_5 < 0$	.968	0.176	0.000	1.000	1.000	1.023
p $\beta_6 < 0$	.021	0.143	0.000	0.000	0.000	1.022
p $\beta_0 < 0$	.934	0.248	0.000	1.000	1.000	1.007
p $\beta_0 + \beta_5 < 0$	.999	0.000	1.000	1.000	1.000	1.000
p $\theta_{new} < 0$	.747	0.435	0.000	1.000	1.000	1.006
p $\theta_{new.B5} < 0$	.833	0.373	0.000	1.000	1.000	1.000
p $\theta_{new.B6} < 0$	.496	0.500	0.000	0.000	1.000	1.000
p $\theta_{new.\beta0 + \beta5} < 0$	.956	0.205	0.000	1.000	1.000	1.007
deviance	-13.375	5.719	-23.018	-13.785	-0.878	1.001
Uniform on τ^2		RE ~ t_4 distribution		DIC = 2.823		
Node	Mean	s.d.	2.5%	50%	97.5%	Rhat
β_0	-0.323	0.224	-0.795	-0.316	0.108	1.007
β_5	-0.453	0.265	-0.951	-0.455	0.095	1.007
B_6	0.007	0.004	0.000	0.007	0.015	1.001
τ	0.396	0.122	0.217	0.381	0.666	1.003
τ^2	0.171	0.117	0.047	0.145	0.444	1.003
$\beta_0 + \beta_5$	-0.777	0.146	-1.082	-0.773	-0.501	1.001
θ_{new}	-0.330	0.463	-1.318	-0.327	0.598	1.004
$\theta_{new.\ \beta0 + \beta5}$	-0.765	0.425	-1.640	-0.776	0.075	1.002
$\theta_{new.\ \beta5}$	-0.454	0.503	-1.549	-0.451	0.534	1.003
$\theta_{new.\ \beta6}$	0.000	0.403	-0.782	0.001	0.786	1.001
p $\beta_5 < 0$	.952	0.214	0.000	1.000	1.000	1.000
p $\beta_6 < 0$	.031	0.173	0.000	0.000	1.000	1.007
p $\beta_0 < 0$	.941	0.236	0.000	1.000	1.000	1.007
p $\beta_0 + \beta_5 < 0$	.999	0.000	1.000	1.000	1.000	1.000
p $\theta_{new} < 0$	.780	0.414	0.000	1.000	1.000	1.006
p $\theta_{new.B5} < 0$	.829	0.376	0.000	1.000	1.000	1.000
p $\theta_{new.B6} < 0$	.499	0.500	0.000	0.000	1.000	1.001
p $\theta_{new.\beta0 + \beta5} < 0$	.958	0.200	0.000	1.000	1.000	1.031
deviance	-13.146	6.055	-23.039	-13.795	1.074	1.001

Chondroitin(k=18): Classical Forest Plot, Mixed-Effects Model with REML



Appendix G. Chondroitin (*k*=20) Meta-analysis: Bayesian summary results

Chondroitin (*k*=20): Uniform on τ (0,5)

Uniform on τ (0,5)			RE ~ *Normal*			DIC = 7.7
Node	Mean	s.d.	2.5%	50%	97.5%	Rhat
β_0	-0.041	0.266	-0.587	-0.037	0.474	1.000
β_1	-0.768	0.289	-1.304	-0.782	-0.188	1.000
β_2	-0.839	0.447	-1.706	-0.848	0.047	1.002
τ	0.469	0.103	0.299	0.456	0.708	1.000
τ^2	0.231	0.106	0.089	0.208	0.501	1.000
$\beta_0 + \beta_1$	-0.809	0.132	-1.056	-0.812	-0.546	1.003
$\beta_0 + \beta_2$	-0.880	0.523	-1.892	-0.860	0.154	1.003
θ_{new}	-0.032	0.546	-1.116	-0.048	1.043	1.004
$\theta_{new.\ \beta0 + \beta1}$	-0.825	0.496	-1.786	-0.836	0.192	1.006
$\theta_{new.\ \beta0 + \beta2}$	-0.861	0.709	-2.154	-0.866	0.544	1.000
$\theta_{new.\ \beta1}$	-0.754	0.559	-1.854	-0.751	0.290	1.001
$\theta_{new.\ \beta2}$	-0.878	0.663	-2.146	-0.899	0.462	1.000
p $\beta_1 < 0$	.993	0.083	1.000	1.000	1.000	1.009
p $\beta_2 < 0$	.969	0.173	0.000	1.000	1.000	1.015
p $\beta_0 < 0$	.553	0.497	0.000	1.000	1.000	1.000
p $\beta_0 + \beta_1 < 0$	.999	0.000	1.000	1.000	1.000	1.000
p $\beta_0 + \beta_2 < 0$	.956	0.205	0.000	1.000	1.000	1.004
p $\theta_{new\ \beta0} < 0$	.533	0.499	0.000	1.000	1.000	1.006
p $\theta_{new.\ \beta1} < 0$	.913	0.282	0.000	1.000	1.000	1.006
p $\theta_{new.\beta2} < 0$	.898	0.303	0.000	1.000	1.000	1.000
p $\theta_{new.\beta0 + \beta1} < 0$	.958	0.200	0.000	1.000	1.000	1.007
p $\theta_{new.\beta0 + \beta2} < 0$	.888	0.315	0.000	1.000	1.000	1.006
deviance	-10.081	6.648	-20.929	-10.840	4.208	1.006

Uniform on τ (0,5)			RE ~ t_4 distribution			DIC = 7.5
Node	Mean	s.d.	2.5%	50%	97.5%	Rhat
β_0	-0.046	0.248	-0.575	-0.046	0.454	1.001
β_1	-0.695	0.281	-1.267	-0.699	-0.134	1.000
β_2	-0.941	0.489	-1.847	-0.967	0.129	1.004
τ	0.371	0.109	0.201	0.353	0.614	1.000
τ^2	0.149	0.093	0.040	0.125	0.377	1.000
$\beta_0 + \beta_1$	-0.741	0.123	-0.982	-0.738	-0.509	1.006
$\beta_0 + \beta_2$	-0.987	0.561	-2.072	-0.992	0.176	1.003
θ_{new}	-0.041	0.458	-0.977	-0.022	0.859	1.000
$\theta_{new.\ \beta0 + \beta1}$	-0.720	0.405	-1.499	-0.711	0.044	1.007
$\theta_{new.\ \beta0 + \beta2}$	-0.977	0.669	-2.189	-0.992	0.359	1.000
$\theta_{new.\ \beta1}$	-0.701	0.470	-1.669	-0.697	0.270	1.006
$\theta_{new.\ \beta2}$	-0.954	0.620	-2.132	-0.943	0.285	1.007
p $\beta_1 < 0$	0.989	0.104	1.000	1.000	1.000	1.100
p $\beta_2 < 0$	0.963	0.189	0.000	1.000	1.000	1.026
p $\beta_0 < 0$	0.579	0.494	0.000	1.000	1.000	1.000
p $\beta_0 + \beta_1 < 0$	1.000	0.000	1.000	1.000	1.000	1.000
p $\beta_0 + \beta_2 < 0$	0.963	0.189	0.000	1.000	1.000	1.048
p $\theta_{new\ \beta0} < 0$	0.525	0.500	0.000	1.000	1.000	1.001
p $\theta_{new.\ \beta1} < 0$	0.933	0.250	0.000	1.000	1.000	1.003
p $\theta_{new.\beta2} < 0$	0.945	0.228	0.000	1.000	1.000	1.042
p $\theta_{new.\beta0 + \beta1} < 0$	0.967	0.179	0.000	1.000	1.000	1.009
p $\theta_{new.\beta0 + \beta2} < 0$	0.929	0.257	0.000	1.000	1.000	1.001
deviance	-10.132	6.484	-20.710	-10.905	4.224	1.003

Appendix H. R program with the R2WinBUGS Package

```
        ####    Bayesian Analysis ####
SMI.Data <- read.table("E:/Programs and Data/Data Files/SMI.csv",sep=",", header=T)
attach(SMI.Data)

names(SMI.Data)
nstudies <- nrow(SMI.Data)
nstudies

library(R2WinBUGS)

         ########              1a. Uni Tau (0,5)  RE ~ normal
model.file <- file.path("E:/Reanalyses/SMI", "SMIUniTau05.txt")
##This specifies where the WinBugs code is located for uniform on tau (0,5) prior distribution##
file.show(model.file)

         ########  The data and variables listed below are for the fitted model ##########
data <- c("nstudies", "Relaxation","Multimodal","Organizational","yi", "sei")

parameters <- c("mu", "mu.1", "theta.new","theta.new.Rel","theta.new.Multi","theta.new.Org",
"beta1", "beta2","beta10","Rel.Mean","p.Rel.Mean","Multi.Mean","p.Multi.Mean",
"Org.Mean","p.Org.Mean","pbeta1.0","pbeta2.0","pbeta10.0","pmu0","pthetanew.0",
"ptheta.new.Rel","ptheta.new.Multi","ptheta.new.Org","tau", "tausq")

UniTau <- bugs(data, inits=NULL, parameters, model.file,
  n.chains=3, n.iter=50000, digits=3,
  debug=TRUE, bugs.directory="c:/Program Files/WinBUGS14/")

print(UniTau, digits.summary = 3)

        ####  This specifies the filepath to write the results ###
FILEPATH <- "E:/Reanalyses/SMI/"
To.File <- function(FILE){sprintf("%s%s", FILEPATH, FILE)}

write.table(UniTau$summary,
  To.File("Uni Tau Normal SMI.csv"), sep=",")

      ### The analysis can be repeated with other Bayesian prior distributions and with the
      ###  RE ~ t4 distribution by changing the code in the model file.   ####
```

Appendix I. WinBUGS program.

```
model
{
   for (i in 1:nstudies){
       prec[i] <- 1/(sei[i]*sei[i])
       theta[i] <- mu.1 +
       beta1*Relaxation[i] +
        beta2*Multimodal[i] +
        beta10*Organizational[i]

     ### The next 2 lines will vary depending upon whether the random effect ~ t4 distribution or
a
     ###  normal distribution.
     #mu[i] ~ dnorm((theta[i],prec.1)
       mu[i] ~ dt(theta[i],prec.1,4)

       yi[i] ~ dnorm(mu[i],prec[i])
   }
     mu.1 ~ dnorm(0,.001)
     beta1 ~ dnorm(0,.001)
     beta2 ~ dnorm(0,.001)
     beta10 ~ dnorm(0,.001)

         Rel.Mean <- mu.1 + beta1
         p.Rel.Mean <- equals(max(Rel.Mean,0),0)

         Multi.Mean <- mu.1 + beta2
         p.Multi.Mean <- equals(max(Multi.Mean,0),0)

         Org.Mean <- mu.1 + beta10
         p.Org.Mean <- equals(max(Org.Mean,0),0)

         pbeta1.0 <- equals(max(beta1,0),0)
         pbeta2.0 <- equals(max(beta2,0),0)
         pbeta10.0 <- equals(max(beta10,0),0)

### Listed below is the code for different prior distributions for the heterogeneity variance###
###  Code for Uniform on tau (0,5) ####
    tau ~ dunif(0,5)
     tausq <- tau*tau
     prec.1 <- 1/tausq
################################################

#### Code for Uniform on tau-squared (0,1000)
     tausq ~ dunif(0,1000)
     tau <- sqrt(tausq)
     prec.1 <- 1/tausq
################################################
```

```
####Code for Half-normal#########################
    hn ~ dnorm(0,.1)
    tausq <- abs(hn)
# Here, tau^2 has upper 95% value at 6.2.
  tau <- sqrt(tausq)
  prec.1 <- 1/tausq
##############################################

####Code for Inverse Gamma##################
   prec.1 ~ dgamma(.01,.01)
  tau <- 1/sqrt(prec.1)
  tausq <- tau*tau
##############################################

### Code for DuMouchel s0 ###################
    s0 <- …. # Compute the s0 value
    k ~ dunif(0,1)
   tau <- (s0*(1-k))/k
   tausq <- tau*tau
   prec.1 <- 1/tausq
##############################################

#### Code for DuMouchel s0/3 #################
    s0div3 <- ….  # use s0/3 value here
    k ~ dunif(0,1)
    tau <- (s0div3*(1-k))/k
    tausq <- tau*tau
    prec.1 <- 1/tausq
##############################################

#### Code for DuMouchel 3s0 #################
s0Mult3 <- …. #use 3*s0 value here
k ~ dunif(0,1)
tau <- (s0Mult3*(1-k))/k
tausq <- tau*tau
 prec.1 <- 1/tausq
##############################################

##############################################
          theta.new ~ dnorm(mu.1, prec.1)
          theta.new.Rel ~ dnorm(Rel.Mean, prec.1)
          theta.new.Multi ~ dnorm(Multi.Mean, prec.1)
          theta.new.Org ~ dnorm(Org.Mean, prec.1)
           pmu0 <- equals(max(mu.1,0),0)
          pthetanew.0 <- equals(max(theta.new,0),0)
          ptheta.new.Rel <- equals(max(theta.new.Rel,0),0)
          ptheta.new.Multi <- equals(max(theta.new.Multi,0),0)
          ptheta.new.Org <- equals(max(theta.new.Org,0),0)
}
}
```

Appendix J. S-plus 2000 program with the hblm function.

```
"C:\\hblm.dmp"
Chondroitin.Data <- read.table(
     "E:/Programs and Data/Data Files/CHONDROITIN oral k = 18 studies.csv",
   sep = ",", header = T)
attach(Chondroitin.Data)
names(Chondroitin.Data)

Chondroitin.fr <- data.frame(yi, sei, Recip.sqrt.N, Patient.Blind, Conceal,
     Placebo, ITT, Weeks, Analgesic, Funding)
attach(Chondroitin.fr)

### Via backwards stepwise model selection, only Weeks and Analgesic significant
hblm.model <- hblm(yi ~ Weeks + Analgesic, sei)
hblm.model
##### Note: weeks was not centered about its mean in this data set

plot(hblm.model)
plot(hblm.model$trace)
cvresults <- crossval(hblm.model)
cvresults
write.table(cvresults, file =
     "E:\\Reanalyses\\Chondroitin Analysis\\hblm k=18 Cross Val Results.csv",
     sep = ",")
```

Appendix K. R program with the Metafor Package.

```
ZincforCold.Data <- read.table(
"E:/Programs and Data/Data Files/ZincforCold.csv",  header=T, sep=",")
attach(ZincforCold.Data)
names(ZincforCold.Data)
ZincforCold.Data

library("metafor")
help(metafor)

FILEPATH <- "E:/Reanalyses/Zinc/"
To.File <- function(FILE){sprintf("%s%s", FILEPATH, FILE)}

##################              FE Model ################################################
FE.simple <-rma(yi=yi, vi=vi, method="FE", digits=3)
summary (FE.simple)
#############################################################################################
##

##############  Simple RE Model with DerSimonian & Laird Estimation Method #####
DL.simple <-rma(yi=yi, vi=vi, method="DL", digits=3)
summary (DL.simple)
confint(DL.simple)
#############################################################################################
##

###################   Simple RE Model with REML Estimation ########################
REML.simple <-rma(yi=yi, vi=vi, method="REML", digits=3)
summary(REML.simple)
cint(REML.simple)
        #####  Higgins Prediction Interval Simple REML Model ###
names(REML.simple)
REML.simple$b
mu <- REML.simple$b
REML.simple$k
K <- REML.simple$k
REML.simple$tau2
tau <-sqrt(REML.simple$tau2)
tau
REML.simple$se
semu <- REML.simple$se
alpha <- .05
HigginsErrorSimple <- qt(alpha/2, K-2)*sqrt(tau^2 + semu^2)
HigginsPI.simple <- c(mu + HigginsErrorSimple, mu - HigginsErrorSimple)
HigginsPI.simple



        #####   Publication Bias ###
funnel(REML.simple)
```

```
regtest(REML.simple)
fsn(yi, vi, type="Orwin")
trimfill(REML.simple)
trimfill(REML.simple, side="right")
trimfill(REML.simple, side="left")
        #####   Leave-one-Out  ###
influence(REML.simple)
influence(REML.simple)[[1]]
influence(REML.simple)[[2]]
write.table(influence(REML.simple)[[1]],
  To.File("Influence Zinc.csv"), sep=",", row.names=F)
write.table(influence(REML.simple)[[2]],
  To.File("Influence Zinc Beta.csv"), sep=",", row.names=F)

        #####  Best Linear Unbiased Predictors, Min and Max ###
blup(REML.simple)
predict(REML.simple)
BLUP.REML <- blup(REML.simple)$pred
c(min(BLUP.REML),max(BLUP.REML))
        #####  Plots ###
plot(REML.simple)
forest(REML.simple, slab=Study.Name,
        xlab="Observed standardized difference in means and 95% CI")
qqnorm(REML.simple, main="")


#################   Simple Model REML with  Knapp and Hartung Adjustment ########
REML.simple.KH <-rma(yi=yi, vi=vi, method="REML", knha=TRUE, digits=3)
summary(REML.simple.KH)
confint(REML.simple.KH)
        ###Higgins PI for Knapp and Hartung###
names(REML.simple.KH)
REML.simple.KH$b
mu <- REML.simple.KH$b
REML.simple.KH$k
K <- REML.simple.KH$k
REML.simple.KH$tau2
tau <-sqrt(REML.simple.KH$tau2)
tau
REML.simple.KH$se
semu <- REML.simple.KH$se
alpha <- .05
HigginsErrorSimple <- qt(alpha/2, K-2)*sqrt(tau^2 + semu^2)
HigginsPI.simple.KH <- c(mu + HigginsErrorSimple, mu - HigginsErrorSimple)
HigginsPI.simple.KH
        ### Best Linear Unbiased Predictors, Min and Max ##
BLUP.REML.KH <- blup(REML.simple.KH)$pred
BLUP.REML.KH
c(min(BLUP.REML.KH),max(BLUP.REML.KH))
#############################################################################
#
```

```
#########  Mixed Model Backward Selection Selection of Covariates ##
Mods <- cbind(Syrup, Adult)
REML.mixed <- rma(yi=yi, vi=vi, mods=Mods, method="REML", digits=3)
summary(REML.mixed)
confint(REML.mixed)
blup(REML.mixed)
BLUP.REML.mixed<- blup(REML.mixed)$pred
BLUP.REML.mixed
c(min(BLUP.REML.mixed),max(BLUP.REML.mixed))

REML.mixed$b
mu <- REML.mixed$b
mu
REML.mixed$k
K <- REML.mixed$k
REML.mixed$tau2
tau <-sqrt(REML.mixed$tau2)
tau
REML.mixed$se
semu <- REML.mixed$se
semu
alpha <- .05
HigginsErrorSimple <- qt(alpha/2, K-2)*sqrt(tau^2 + semu^2)
HigginsPI.REML.mixed <- cbind(mu + HigginsErrorSimple, mu - HigginsErrorSimple)
HigginsPI.REML.mixed

        ####  Mixed Model Selection of Covariates - Removal of Syrup, p = .585
Mods <- cbind(
#Syrup, Removed not significant, p = .585 (from full model)
Adult)
REML.mixed <- rma(yi=yi, vi=vi, mods=Mods, method="REML", digits=3)
summary(REML.mixed)
        #### Adult was not significant, p = .103, model reduces to a simple RE model
##########################################################################################
##

############ RE Model (REML) with with tau-squared set to 95% CI Lower Value ########
Mods <- cbind(Syrup, Adult)
REML.mixed.low <- rma(yi=yi, vi=vi, mods=Mods, method="REML", tau2 = .338, digits=3)
summary(REML.mixed.low)
confint(REML.mixed.low)
        ###  Higgins Approximate Prediction Interval ###
mu <- REML.mixed.low$b
K <- REML.mixed.low$k
REML.mixed.low$tau2
tau <-sqrt(REML.mixed.low$tau2)
tau
REML.mixed.low$se
semu <- REML.mixed.low$se
alpha <- .05
```

```
HigginsError.REML.mixed.low <- qt(alpha/2, K-2)*sqrt(tau^2 + semu^2)
HigginsPI.REML.mixed.low <- cbind(mu + HigginsError.REML.mixed.low, mu -
HigginsError.REML.mixed.low)
HigginsPI.REML.mixed.low
BLUP.REML.mixed.low<- blup(REML.mixed.low)$pred
BLUP.REML.mixed.low
c(min(BLUP.REML.mixed.low),max(BLUP.REML.mixed.low))
################################################################################
##

############ RE Model (REML) with with tau-squared set to 95% CI High ###############
Mods <- cbind(Syrup, Adult)
REML.mixed.high <- rma(yi=yi, vi=vi, mods=Mods, method="REML", tau2 = 6.697, digits=3)
summary(REML.mixed.high)
confint(REML.mixed.high)
          ###  Higgins Approximate Prediction Interval ###
mu <- REML.mixed.high$b
K <- REML.mixed.high$k
REML.mixed.high$tau2
tau <-sqrt(REML.mixed.high$tau2)
tau
REML.mixed.high$se
semu <- REML.mixed.high$se
alpha <- .05
HigginsError.REML.mixed.high <- qt(alpha/2, K-2)*sqrt(tau^2 + semu^2)
HigginsPI.REML.mixed.high <- cbind(mu + HigginsError.REML.mixed.high, mu -
HigginsError.REML.mixed.high)
HigginsPI.REML.mixed.high

BLUP.REML.mixed.high<- blup(REML.mixed.high)$pred
BLUP.REML.mixed.high
c(min(BLUP.REML.mixed.high),max(BLUP.REML.mixed.high))
################################################################################
##

######  Computing Power for Regression Coefficients from Full Model  ###################
Sigma.Star <- REML.mixed$vb
Sigma.Star
alpha <- .05
C.alpha2 <- 1.96
min.imp.value <- .33  #value set to .33

Column.Names <- c("Intercept", colnames(Mods))
Column.Names
for(K in 1:length(Column.Names)){
  sigma <- Sigma.Star[K,K]
  POWER <- 1 - pnorm(C.alpha2 - min.imp.value/sqrt(sigma)) +
              pnorm(-C.alpha2 - min.imp.value/sqrt(sigma))
  cat(sprintf("Power for %s is %.04f \n", Column.Names[K], POWER))
}
####################       Correlation Table
```

```
########################################
cor(Mods)
cor(Mods,yi)
write.table(cor(Mods), file=To.File("Correlation Table Zinc.csv"), sep=",")
```

Appendix L. Equations

REML log-likelihood is

$$L_F(\tau^2; Y) = constant - {}^1\!/_2 \sum_{i=1}^{k} \ln\left(v_i^{*-1}\right) - {}^1\!/_2 \ln\left(\sum_{i=1}^{k} v_i^{*-1} X_i X_i^Y\right) - {}^1\!/_2 \sum_{i=1}^{k} v_i^{*-1} \left(Y_i - X_i^Y \beta\right)^2.$$

REML estimation equation for the between-study variance is

$$\tau^2 = {}^{\sum_{i=1}^{k} v_i^{*-2} \left[\left(Y_i - X_i^Y \beta\right)^2 - v_i\right] + tr\left[\left(\sum_{i=1}^{k} v_i^{*-1} X_i X_i^Y\right)^{-1} \left(\sum_{i=1}^{k} v_i^{*-2} X_i X_i^Y\right)\right]}\Big/{\sum_{i=1}^{k} v_i^{*-2}}$$

REML equation for τ^2 (in the balanced case)

$$\tau^2 = \frac{\sum_{i=1}^{k} \left(Y_i - X_i^Y \beta\right)^2 - v}{k-p-1}.$$

Bayesian marginal posterior distribution of μ, which is represented as μ^*(note that this equation does not include covariates and uses a diffuse prior for μ):

$$\mu^* = E(\mu|y) = \int \mu^*(\tau)\pi(\tau|y)d\tau \ .$$

This is calculated by multiplying the curve height of the posterior expectation of μ, given τ, by the histogram height of the posterior distribution of τ and integrating (DuMouchel, 1994).

Appendix M. Sample Bayesian Meta-analysis Checklist.

Sample: Bayesian Quality Assurance Meta-analysis Checklist

Uniform on τ fitted model: $\boldsymbol{X_i\beta} = \widehat{\boldsymbol{\beta}}_{\boldsymbol{0}} + \widehat{\boldsymbol{\beta}}_{\boldsymbol{1}}(\boldsymbol{X_{i1}}) + \widehat{\boldsymbol{\beta}}_{\boldsymbol{2}}(\boldsymbol{X_{i2}})$ with ☒ $\delta_i \sim Normal$ ☐ $\delta_i \sim t_4$

Prior Distribution for τ or τ^2:

☐ Uniform on τ ☐ Uniform on τ^2 ☐ Half-normal ☐ Inverse Gamma ☒ DuM[a] s_0 ☐ DuM $s_0/3$ ☐ DuM $3s_0$

Random effect modeled from: Prior Distribution for **β**: ☒ Normal (0,1000)

☐ Normal distribution ☒ t_4-distribution

Cross-validation conducted?: ☐No ☒Yes. Outlier Detected?: ☒No ☐ Yes

Model specification:

☒ Run 3 chains ☒ 50,000 iterations ☒ Discard first half of iterations

Do meta-analytic conclusions differ from the baseline fitted hblm? ☐No ☐ Yes

The discrepancy[b] between baseline fitted model vs. new model is: ☐Min. ☐ Modest ☐Severe

If severe discrepancy, the new fitted model is:

Parameters to monitor: ☒ **DIC** = _____

Node	Parameter set?	MC error less than 5% of sd?	Is Rhat in close proximity to 1?	Comments
β_0	☒	☒	☒	
β_1	☒	☒	☒	
β_2	☒	☒	☒	
τ	☒	☒	☒	
$\beta_0 + \beta_1$	☒	☒	☒	
$\beta_0 + \beta_2$	☒	☒	☒	
$\theta_{new}\ \beta_0$	☒	☒	☒	
$\theta_{new}\ \beta_1$	☒	☒	☒	
$\theta_{new}\ \beta_2$	☒	☒	☒	
$\theta_{new}\ \beta_0 + \beta_1$	☒	☒	☒	
$\theta_{new}\ \beta_0 + \beta_2$	☒	☒	☒	
$p\ \beta_0 < 0$	☒	☒	☒	
$p\ \beta_1 < 0$	☒	☒	☒	
$p\ \beta_2 < 0$	☒	☒	☒	
$p\ \beta_0+\beta_1 < 0$	☒	☒	☒	
$p\ \beta_0+\beta_2 < 0$	☒	☒	☒	
$p\ \theta_{new}\ \beta_0 < 0$	☒	☒	☒	
$p\ \theta_{new}\ \beta_0 + \beta_1 < 0$	☒	☒	☒	
$p\ \theta_{new}\ \beta_1 < 0$	☒	☒	☒	
$p\ \theta_{new}\ \beta_2 < 0$	☒	☒	☒	
$p\ \theta_{new}\ \beta_0 + \beta_2 < 0$	☒	☒	☒	
Deviance	☒	☒	☒	
θ_i^*	☒	☒	☒	

Note. [a]*DuM = DuMouchel.* [b]Discrepancy refers to the amount of discrepancy between the baseline best-fitting meta-analysis model and the alternative Bayesian meta-analysis model.

Appendix N. Classification System for Describing the Impact of Heterogeneity

Classification Table with Qualitative Indicators for Describing the way that Meta-analysis Inferences/Conclusions Change Depending on Uncertainity in the Amount of Heterogeneity

Do the fitted baseline and fitted alternative models remain identical? (i.e., Are the same study-level covariates selected for inclusion? Does the significance of the intercept and study-level covariates remain the same among models?)	Are the parameter estimates for the model coeffiecients similar in size among the baseline and alternative model?	Impact on Meta-analysis Inferences/Conclusions
Yes	Yes	Minimal
Yes	No, differ to a modest degree	Modest
No	No, differ substantially	Severe

REFERENCES

Abrams, K. & Sanso, B. (1998). Approximate Bayesian inference for random effects meta-analysis. *Statistics in Medicine, 17*(2), 201-218.

American Recovery and Reinvestment Act of 2009 (2009). H.R.1. S.1, 111th U.S. Congress, First Session.

Borenstein, M., Hedges, L., Higgins, J., Rothstein, H. (2005). Comprehensive meta-analysis (version 2) [Computer software]. Englewood, NJ: Biostat.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. Chichester, England: John Wiley & Sons.

Borman, G., & Grigg, J. (2009). Visual and narrative interpretation. In H. Cooper, L.Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 497-519). New York, NY: Russell Sage Foundation.

Burnham, K. & Anderson, D. (1998) *Model Selection and Inference: a Practical Information-theoretic Approach*. New York: Springer.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials 7*(3), 177-188.

DerSimonian, R., & Kacker, R. (2007). Random-effects model for meta-analysis of clinical trials: An update. *Contemporary Clinical Trials, 28*(2), 105-114.

DuMouchel, W. (1994). *Hierarchical Bayes linear models for meta-analysis* . Technical Report 27. Research Triangle Park, NC: National Institute of Statistical Sciences.

DuMouchel, W. (1995). Documentation for hblm. hblm.doc.ps. Retrieved from ftp://ftp.research.att.com/dist/bayes-meta/.

DuMouchel, W. (1995). Hblm.dmp [Software]. Retrieved from ftp://ftp.research.att.com/dist/bayes-meta/.

DuMouchel, W., & Normand, S. (2000). Computer modeling and graphical strategies for meta-analysis. In D. Stangl & D. Berry (Eds.), *Meta-analysis in medicine and health policy* (127-178). New York: Marcel Dekker.

Duval, S. & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*, 89–98.

Egger, M., Davey Smith, G., Schneider, M. & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*(7109), 629–634.

Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models.* New York: Cambridge Univ Press.

Glasziou, P., & Sanders, S. (2002). Investigating causes of heterogeneity in systematic reviews. *Statistics in Medicine, 21*(11), 1503-1511.

Greenhouse, J., & Iyengar, S. (2009). Sensitivity analysis and diagnostics. In H. Cooper, L.Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 417-433). New York, NY: Russell Sage Foundation.

Hardy, R. J., & Thompson, S. G. (1998). Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine, 17*(8), 841-856.

Hedges, L. & Pigott, T. (2004). The power of statistical tests of moderators in meta-analysis. *Psychological Methods, 9*, (4), 426–445.

Higgins, J., & Thompson, S. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539-1558.

Higgins, J., Thompson, S., Deeks, J., & Altman, D. (2003). Measuring inconsistency in meta-analysis. *British Medical Journal, 327*, 557-560.

Higgins, J. & Thompson, S. (2004). Controlling the risk of spurious findings from meta-regression . *Statistics in Medicine, 23*(11), 1663-1682.

Higgins, J., Thompson, S., & Spiegelhalter, D. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society, 172*(1), 137-159.

Ioannidis, J.P., Patsopoulos, N.A., & Rothstein, H.R. (2008). Reasons or excuses for avoiding meta-analyses in forest plots. *British Medical Journal,336*, 1413-1415.

Jackson, D. (2006). The power of the standard test for the presence of heterogeneity in meta-analysis. *Statistics in Medicine, 25*(15), 2688-2699.

Kisamore, J. L., & Brannick, M. T. (2008). An illustration of the consequences of meta-analysis model choice. *Organizational Research Methods, 11*(1), 35-53.

Knapp, G. & Hartung, J. (2003) Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine, 22*(17)*,* 2693–2710.

Konstantopoulos, S., & Hedges, L. (2009). *Analyzing effect sizes: Fixed-effects models*. In H. Cooper, L. Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 295-315). New York, NY: Russell Sage Foundation.

Lambert, P.C., Sutton, A.J., Burton, P.R., Abrams, K.R., & Jones, D.R. (2005). How vague is vague?: A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine, 24,* 2401-2428.

Lau, J., Ioannidis, J. P., & Schmid, C. H. (1998). Summing up evidence: One answer is not always enough. *Lancet, 351*(9096), 123-127.

Lee, K.J., & Thompson, S.G. (2007). Flexible parametric models for random-effects distributions. *Statistics in Medicine, 27, 428-434.*

Lewis, S., & Clarke, M. (2001). Forest plots: Trying to see the wood and the trees. *British Medical Journal, 322*(7300), 1479-1480.

Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Lynch, S. & Western, B. (2004). Bayesian Posterior Predictive Checks for Comples Models. *Sociological Methods and Research, 32,*(3), 301-335.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). *WinBUGS* -- a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing, 10*, 325-337.

Matt, G. C. T. (2009). Threats to the validity of generalized inferences. In H. Cooper, L. Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 537-560). New York, NY: Russell Sage Foundation.

Morris, C. N. (1983). Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association, 78*(381), 47-55.

National Research Council. (1992). *Combining information: Statistical issues and opportunities for research*. Washington, D.C.: National Academy of Sciences Press.

Orwin, R. G. (1983) A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics, 8,* 157–159.

Overton, R.C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods, 3*, 354-379.

Patsopoulos, N. A., Analatos, A. A., & Ioannidis, J. P. A. (2005). Relative citation impact of various study designs in the health sciences. *Journal of the American Medical Association, 293*(19), 2362-2366.

Pauler, D., & Wakefield, J. (2000). Modeling and implementation issues in Bayesian meta-analysis. In D. Stangl & D. Barry (Eds.), *Meta-analysis in medicine and health policy* (pp. 205-230). New York, NY: Marcel Dekker.

Raudenbush, S. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of the expectancy induction: A synthesis of findings from 18 experiments. *Journal of Educational Psychology, 76*(1), 85-97.

Raudenbush, S., & Bryk, A. (1985). Empirical Bayes meta-analysis. *Journal of Educational Statistics, 10*(2), 75-98.

Raudenbush, S. W., & Bryk, A. (2002). Bayesian inference for hierarchical models. *Hierarchical linear models: Applications and data analysis methods* (2nd ed., pp. 399-435). Thousand Oaks, CA: Sage.

Raudenbush, S. (2009). Analyzing effect sizes: Random effect models. In H. Cooper, L. Hedges & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 295-315). New York, NY: Russell Sage Foundation.

Reichenbach, S., Sterchi, R, Scherer, M., Trelle, S., Burgi, E., Burgi, U.,…Juni, P. (2007). Meta-analysis: Chondroitin for osteoarthritis of the knee or hip. *Annals of Internal Medicine, 146,* 580-590.

Review Manager (RevMan) [Computer program]. Version 5.0. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2008.

Richardson, K. & Rothstein, H. (2008). Effects of occupational stress management interventions: a meta-analysis. *Journal of Occupational Heatlh Psychology, 13*(1), 69-93.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641.

Rothstein, H., Sutton, A., & Borenstein, M. (2005). *Publication Bias in Meta-Analysis: Prevention, Assessment, and Adjustments.* John Wiley & Sons, Chichester, England.

Rui, N. (2009). Four decades of research on the effects of detracking reform: Where do we stand? A systematic review of the evidence. *Journal of Evidence-Based Medicine*, *2*(3), 164-183.

Schmid, C. H. (2001). Using Bayesian inference to perform meta-analysis. *Evaluation & The Health Professions, 24*(2), 165-189.

Schmidt, F. (2008). Meta-analysis: A constantly evolving research tool. *Organizational Research Methods, 11*(1), 96-113.

Schmidt, F., Le, H., & Oh, I. (2009). Correcting for the distorting effects of study artifacts in meta-analysis. In H. Cooper, L. Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., 317-333.). New York, NY: Russell Sage Foundation.

Schmidt, F., Oh, I., & Hayes, T. (2009). Fixed- versus random effects models in meta-analysis: Model properties and an empirical comparison of the differences in results. *British Journal of Mathematical and Statistical Psychology, 62*, 97-128.

Seltzer, M. (1993). Sensitivity analysis for the fixed-effects in the hierarchical model: A Gibbs sampling approach. *Journal of Educational Statistics, 18*(3), 207-235.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company.

Shadish, W. & Haddock (2009). Combining estimates of effect size. In H. Cooper, L. Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp.). New York, NY: Russell Sage Foundation.

Smith, T. C., Spiegelhalter, D. J., & Thomas, A. (1995). Bayesian approaches to random-effects meta-analysis: A comparative study. *Statistics In Medicine, 14*(24), 2685-2699.

Spiegelhalter D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society Series B, 64*(4), 583-640.

Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). *WinBUGS User Manual, Version 1.4.* Cambridge, Medical Research Council Biostatistics Unit.

Spiegelhalter, D. J., Abrams, K., & Myles, J. (2004). *Bayesian approaches to clinical trials and health-care evaluation*. Chichester, England: John Wiley & Sons.

Stangl, D. & Berry, D. (2000). Meta-analysis: Past and present challenges. In D. Stangl & D. Berry (Eds.), *Meta-analysis in medicine and health policy.* New York, NY: Marcel Dekker.

Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software, 12,* (3), 1-17.

Sutton, A., Abrams, K., Jones, D., Sheldon, T., & Song, F. (2000). *Methods for meta-analysis in medical research*. New York, NY: John Wiley & Sons.

Sutton, A., & Abrams, K. (2001). Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research, 10*, 277-303.

Sutton, A., & Higgins, J. (2008). Recent developments in meta-analysis. *Statistics in Medicine, 27*, 625-650.

Switzer, F., Paese, P., & Drasgow, F. (1992). Boostrap estimates of standard errors in validity generalization. *Journal of Applied Pscyhology, 77,* 123-129.

Viechtbauer W (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics, 30.*(3), 261-293.

Viechtbauer, W. (2007). Confidence intervals for the amount of heterogeneity in meta-analysis. *Statistics in Medicine, 26,* 37-52.

Viechtbauer, W. (2009). Metafor Package. R package version 0.5-7.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. Journal of Statistical Software, 36(3), 1-48.