Title
Robust Shrinkage Estimation of Effect Sizes for Bayesian Meta-Analysis Models

Author
Kim, Junok

Publication Date
2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Robust Shrinkage Estimation of Effect Sizes for Bayesian Meta-Analysis Models

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Education

by

Junok Kim

2023

ABSTRACT OF THE DISSERTATION

Robust Shrinkage Estimation of Effect Sizes for Bayesian Meta-Analysis Models

by

Junok Kim

Doctor of Philosophy in Education

University of California, Los Angeles, 2023

Professor Michael H. Seltzer, Chair

This study introduces a fully Bayesian (FB) approach based on heavy-tailed distributional assumptions to overcome the over-shrinkage issue that often occurs under the conventional random-effects models for meta-analysis. In a meta-analysis with outlying study results, Empirical Bayes (EB) estimates of outliers are often shrunk to an average effect size by an excessive amount. This over-shrinkage of outliers can be problematic especially when attempting to answer substantive questions concerning how large the largest effect size in a given sample of studies might be.

In order to address this issue, I employ an FB approach specifying t-distributional assumptions for random effects (Bayes-t model) based on a normal-gamma formulation of the t-distribution (Seltzer, 1993; Seltzer et al., 1996). Specifically, the shrinkage estimate of an outlier from the Bayes-t model is compared to those from more standard approaches such as the EB method and an FB model based on a normal-distributional assumption for random effects

(Bayes-normal model). For the implementation of FB random-effects models, a Gibbs sampling algorithm is used to obtain marginal distributions of parameters of interest from a complex joint distribution of all parameters in the model.

The findings from the empirical data-analysis and simulation study results highlight the advantage of employing a t-distributional assumption for random effects fully Bayesian (FB) meta-analysis models to yield robust shrinkage estimates of outliers. Specifically, the empirical data-analysis results show that the least amount of shrinkage occurs under the Bayes-t model with small degrees of freedom, compared to the EB and Bayes-normal models. The results from a targeted simulation study are also consistent with the data-analysis results. The shrinkage estimate of outlier from the Bayes-t model have good properties, such as less bias, smaller mean squared errors, and actual coverage closer to the nominal coverage of 95%, compared to those from the conventional methods. Further, the Bayes risk of the shrinkage estimates of outliers is the smallest under the Bayes-t model, while the risk averaged over the entire set of studies in a meta-analysis sample is similar across the estimation methods.

The over-shrinkage problem stemming from EB estimation and the possibility that employing heavy-tailed distributional assumptions might alleviate the issue have been discussed in some of the early literature on multilevel models and MCMC estimation. Motivated by these works, this study demonstrates that the Bayes-t approach is capable of providing improved shrinkage estimates of true effect sizes of outliers. Future study should address the sensitivity of results to the prior specification and the possible presence of publication bias, and examine whether the current findings extend to more complex settings, i.e. meta-analytic datasets with dependent effect sizes within each study.

The dissertation of Junok Kim is approved.

Li Cai

Minjeong Jeon

Mark P. Hansen

Michael H. Seltzer, Committee Chair

University of California, Los Angeles

2023

To my family

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I would like to express my gratitude to Dr. Michael Seltzer, my advisor and committee chair, for guiding and supporting me throughout my entire doctoral studies. Mike has introduced me to the wonderful world of Bayesian statistics and has shown me what rigorous research is. I truly appreciate his insights and expertise he shared with me during my dissertation process. I am also deeply grateful to the members of my dissertation committee: Drs. Li Cai, Minjeong Jeon, and Mark Hansen for sharing their valuable feedback and suggestions to refine my dissertation. I am also thankful to Dr. Noreen Webb for her guidance, understanding, and support while I was working a teaching assistant for her Generalizability Theory class during the pandemic.

My graduate school experience greatly benefitted from the faculty members and peers in Social Research Methodology (SRM). First of all, I would like to thank Dr. Mike Rose, who was a great writer, teacher, mentor, and friend to us all at UCLA. I am also thankful to Drs. Christina Christie and Jose-Felipe Martinez for creating such a supportive and cheerful community, and Dr. Matthew Madison for his guidance and support while I was working for his research project. I also want to thank my cohort, Drs. Mariana Barragan-Torres, Sijia Hwang, Meredith Langi, Yonsoo Suh, and Shujin Zhong, for their emotional support and intellectual stimulus. I would also like to specifically thank the former students who are always willing to share their experience whenever I need advices: Drs. Ellie Yun, Seungwon Chung, Jayashri Srinivasan, and Julie Liao. Yun Kim also deserves special thanks for her help from Los Angeles during the days I had been working remotely from Chicago and Miami.

I am also appreciative of my supervisors and colleagues at CRESST. I would like to thank Drs. Kilchan Choi and Greg K. W. K. Chung for all the help and guidance while I had been working for their wonderful research projects. Huge thanks go out to Drs. Nami Shin and Eunhee Keum for always making the time to ask my questions.

I would also like to extend my thanks to the other colleagues who helped me during my dissertation process: Dr. Kashia Rosenau for giving me the opportunity to fully engage in the entire steps of meta-analysis, and Ingrid Tien for sharing her amazing meta-analytic dataset. Their expertise and insights greatly expanded my understanding of meta-analysis at the substantive standpoint. I also want to thank Jinkyung Jang, my old friend and an outstanding historian, for reviewing my writings on countless occasions and always giving me helpful feedback to improve my dissertation drafts.

Most importantly, I owe my greatest thanks to my husband Dr. Younghak Kwon, who has been always positive and encouraging during my doctoral journey. I could not have finished my degree without his support. Last but not the least, I would like to thank my parents and sister for their unwavering love, patience and encouragement.

VITA

EDUCATION

2013	Bachelor of Art, Education
	Seoul National University, South Korea
2016	Master of Art, Education
	Seoul National University, South Korea

WORK

2016-2021	Graduate Student Researcher: Research Assistant
	National Center for Research on Evaluation, Standards, and Student Testing
	(CRESST), University of California, Los Angeles
2018-2019	Graduate Student Researcher: Research Assistant
	Department of Education, University of California, Los Angeles
2019-2022	Graduate Student Researcher: Special Reader and Teaching Assistant
	Department of Education, University of California, Los Angeles

CHAPTER 1

Introduction

Meta-analysis plays a key role in many disciplines, including education, psychology, public health, and medicine. The purpose of meta-analysis is to utilize the findings from multiple studies of a treatment or policy of interest in order to obtain an estimate of the average effect size and an estimate of the extent to which the study outcomes vary across studies, and to investigate how differences in certain study characteristics relate to differences in effect-size estimates.

Also of importance in conducting a meta-analysis is the question: Among the studies in one's sample, how large might the largest true effect of the program or policy of interest be? To answer this question, obtaining sound estimates of outlying studies is important, as it enables a close examination of outlying study results that can bring to light key elements or factors for a particularly successful implementation of a program of interest (Rubin, 1981).

For meta-analysis, a random-effects model is widely used to estimate key parameters in the model, including what are termed Empirical Bayes (EB) estimates, or shrinkage estimates, of true effect sizes for individual studies (Raudenbush & Bryk, 2002). Despite many advantages, EB estimation can result in substantial over-shrinkage, i.e., substantial bias, in the case of outlying effect sizes. This is clearly problematic when one's interest centers on how large the largest true effect of the program or policy of interest might be.

This study aims to illustrate the problem associated with EB estimates and assess the statistical properties of an alternative estimation approach for meta-analysis models that helps overcome the shrinkage problems. While random-effects models for meta-analysis typically employ normal distributional assumptions, i.e., the distribution of effect-size parameters across studies is assumed to be normal, I show how the use of t-distributional assumptions with small degrees of freedom alleviates the over-shrinkage problem, and provides more sound estimates of outlying effect sizes.

1.1. Random-effects models for meta-analysis

A random-effects model for standard meta-analysis is often represented as a two-level multilevel model consisting of a within-study model at Level-1 and a between-study model at Level-2. In the within-study model, for each of *k* studies, the observed effect size reported in study *i*, d_i, is viewed as a function of the true effect size for study *i*, δ_i , plus sampling error, e_i, which is assumed to follow a normal distribution with mean zero and known variance, s_i^2:

$$d_i = \delta_i + e_i, \qquad e_i \sim N(0, s_i^2). \tag{1.1}$$

In meta-analysis the sampling variance s_i^2 is often set equal to the squared standard error of d_i.

In the between-study model, the true effect sizes are viewed as being normally distributed around a grand mean μ with variance τ^2:

$$\delta_i = \mu + u_i, \qquad u_j \sim N(0, \tau^2). \tag{1.2}$$

The level-2 residual u_i represents the deviation of the true value of the effect size for study *i* from the grand mean μ, and the u_i's are assumed normally distributed with mean 0 and variance τ^2. Note that from a Bayesian perspective, the between-study model in Equation 1.2 can be viewed as a prior distribution for the δ_i's. As will be discussed later, the between-study model can be expanded to include potential predictors of the variation in study effect sizes.

Using a program such as HLM 8 (Raudenbush & Congdon, 2021), one can obtain the restricted maximum likelihood (REML) estimate of τ^2 (i.e., $\hat{\tau}^2$), and also a precision-weighted estimate of the average effect size (i.e., $\hat{\mu}$), where a weighted average of the observed effect sizes (i.e., the d_i's) is obtained using weights of the following form: $1/(\tau^2 + s_i^2)$. Note that those observed effect sizes with smaller error variances s_i^2 will receive more weight than those with larger s_i^2's when computing $\hat{\mu}$.

1.2. Over-shrinkage problem of outliers

In the case of the within-study and between-study models outlined above, there are two sources of information regarding the magnitude of the true effect size, δ_i, for a given study in the sample of *k* studies. The first source of information is based strictly on study *i*'s data, d_i. The second source of information in this example is the estimate of the grand mean based on the sample of *k* studies, $\hat{\mu}$.

Lindley and Smith (1972) and Box and Tiao, (1973/1992) show that in the case of normal data / normal prior models such as the random-effects model outlined in Equations 1.1 and 1.2,

with μ and τ^2 as known (i.e., with the prior distribution as known), the optimal estimator is:

$$\delta_i^* = \lambda_i * d_i + (1 - \lambda_i) * \mu, \quad (1.3)$$

where $\lambda_i = \tau^2/(\tau^2 + s_i^2)$.

In the case of ML-based programs, e.g., HLM 8 (Raudenbush & Congdon, 2021), a REML estimate of τ^2 (i.e., $\hat{\tau}^2$) is substituted for τ^2 in the above equations, and the estimate of the grand mean, (i.e., $\hat{\mu}$), described above, replaces μ, in efforts to obtain estimates of δ_i^*'s for the *k* studies in one's sample. The resulting estimates of the δ_i^*'s are often termed EB estimates.

As a simple illustration to help convey the logic of the above estimator, suppose that $\hat{\tau}^2$ and s_i^2 are equal in magnitude. This will result in λ_i equaling a value of 0.5 for study *i*. Thus in computing δ_i^*, equal weight of 0.5 will be placed on the data, d_i, and on the estimate of the grand mean, $\hat{\mu}$. On the other hand, if s_i^2 is very small relative to $\hat{\tau}^2$ as in the situation where d_i is estimated with a relatively high degree of precision, then λ_i will approach a value of 1, resulting in a weight close to a value of 1 for d_i and a value close to 0 for $\hat{\mu}$ that will used in computing δ_i^*. In contrast, if s_i^2 is very large in relation to $\hat{\tau}^2$ (i.e., d_i is estimated with a relatively low degree of precision), then a small weight will be placed on d_i and a great deal of weight will placed on $\hat{\mu}$. This shows that the observed effect-size estimate, d_i, is always pulled or shrunk to some degree toward the grand mean under EB approach. As such, EB estimates are sometimes termed shrinkage estimates.

Note that in meta-analysis models that include a study-specific predictor in the between-study model, e.g., a measure of the duration of the program of interest for each study, each study will be shrunk toward a conditional mean given its predictor value, instead of being shrunk

toward an estimate of the grand mean. I will also examine the meta-regression model with a single predictor in later chapters.

In discussing models based on EB estimation, the term "borrowing strength" is often used. That is when the error variance connected with an effect size is fairly large relative to $\hat{\tau}^2$, a large proportion of weight will be placed on the grand mean, i.e., a source of information based on all of the studies in the sample. When the error variance is small relative to $\hat{\tau}^2$, only a small amount of weight will be placed on the grand mean, and a large amount of weight will then be placed on the effect-size estimate of study *i* based strictly on the data from that study.

In meta-analysis, the EB estimates of true effect sizes, $\delta_i's$, will on average have smaller mean squared errors in a sample when compared to the set of observed data, the d_i's, in a given sample. However, there is no guarantee that the EB estimate for each study will be close to its true effect size, as discussed in Efron and Morris (1971, 1972). This will especially be the case with respect to outlying effect sizes, for which there tends to be substantial shrinkage toward the grand mean.

For illustration of the over-shrinkage problem for outliers, consider the previous example where τ^2 and s_i^2 are equal. In this case, $\lambda_i = \tau^2/(\tau^2 + s_i^2) = 0.5$, and $1 - \lambda_i = 0.5$. Suppose that the grand mean μ is equal to the standardized mean difference of 0.2. As the standardized mean differences are in standard deviation units, the average effects size of 0.2 corresponds to the difference of one-fifth of a standard deviation. Note that in the social sciences, an effect size of 0.2 represents a small effect, an effect size of 0.5 is interpreted as a moderate effect, and an effect size over 0.8 indicates a large effect (Cohen, 1988/2013). This results in a shrinkage

estimate as follows:

$$\delta_i^* = 0.5 * d_i + 0.5 * 0.2.$$

First let's suppose the observed effect size of study *i* is 0.3, which is one-tenth of an effect size larger than the overall mean. The shrinkage estimate of study *i*'s true effect size will be:

$$\delta_i^* = 0.5 * 0.3 + 0.5 * 0.2 = 0.25,$$

which is slightly smaller than the observed effect size of 0.3 and little larger than the grand mean of 0.2.

Now suppose that for another study in the sample, the observed effect size is 1.2, which is substantially larger than the grand mean of 0.2. The weights again are 0.5 as above. The resulting shrinkage estimate in this case is:

$$\delta_i^* = 0.5 * 1.2 + 0.5 * 0.2 = 0.7.$$

So while the observed effect size is 1.2, the EB estimate or shrinkage estimate is 0.7. That is a marked reduction of half an effect size. In addition, the interpretation of the size of effect size will be changed accordingly at a substantive point. Thought the study reports a large effect size of 1.2, the corresponding shrinkage estimate of 0.7 indicates that the intervention has a moderate effect.

In both examples, a weight of 0.5 was placed on the grand mean. In the first example, a weight of 0.5 was placed on the observed effect size of 0.3, which is slightly larger than the grand mean of 0.25, i.e., it is close to the grand mean. In the second case, a weight of 0.5 was placed on the observed effect size of 1.2, which is substantially larger than the grand mean. What

we see is that in computing the corresponding EB estimate, the effect size of 1.2 was cut in half, down to a value of 0.6. This shows that the effect sizes that are substantially larger than the grand mean will be reduced substantially in forming EB estimates. Similarly, in the case of effect sizes that are substantially smaller than the grand mean, the concern is that they will be pulled toward the grand mean by an excessive amount.

A standard meta-analysis model based on an EB approach often fails to provide sensible estimates of outlying study results because of the over-shrinkage problem illustrated above. This calls for an alternative approach which will provide more robust results for extreme cases.

1.3. Robust approach assuming a t-distribution for random effects

A promising approach to alleviate the over-shrinkage issue is employing a heavy-tailed distribution for random effects (Dempster, 1983; Seltzer et al., 1996; West, 1984). For example, a random-effects model based on a t-distributional assumption provides an improved shrinkage estimates of outliers, which are often shrunk toward the overall mean by a large amount under the assumption of normally distributed random effects (West, 1984). As will be described in the following chapter, an FB model based on a t-distributional assumption with small degrees of freedom for the random effects yields weights, which are termed q_i's, for each study based on the distance between the observed effect size for a study and the average effect size of all studies in the sample. When d_i is far from the grand mean, for example, the weight for such a study will tend to be close to a value of 0; and when d_i is close to the grand mean, the weight for such a study will tend to be close to a value of 1.

In the Bayes-t approach, the between study variance component τ^2 is essentially divided by a given study's weight q_i, yielding the parameter variance for study i, τ^2/q_i. For an outlier, q_i will tend to be close to 0, and this in turn will inflate the parameter variance τ^2/q_i for this study. For a study whose observed effect size d_i is close to the grand mean, the parameter variance τ^2 will be divided by a value approximately equal to a value 1, and hence the parameter variance for such a study will be very close to τ^2.

Returning to the issue of forming a compromise estimator δ_i^* based on d_i and the grand mean μ (Equation 1.3), in the Bayes-t formulation with small df, as will be seen, d_i, the magnitude of the observed effect size for a study, is now multiplied by: $\lambda_i = \frac{\tau^2/q_i}{\tau^2/q_i + s_i^2}$. Thus for an outlier λ_i will approach a value in the vicinity of 1, and $1 - \lambda_i$ will approach a value of 0. This will lead to substantially less shrinkage compared to the EB approach. In this way, employing t-distributional assumptions in the between-study model provides protection for the outlier against the over-shrinkage toward the grand mean (or toward a conditional mean when predictors are included in the between-study model). The robust estimates of outliers will be particularly valuable if one wishes to investigate and learn from outliers, instead of treating an outlier merely as a statistical nuisance.

To this end, in this dissertation, I demonstrate that employing t-distributional assumptions with small degrees of freedom for effect-size parameters within an FB framework helps overcome the over-shrinkage problem for outlying effect sizes. Though the FB multilevel models assuming t-distributed random effects have been widely used in small-sample settings, the primary interests often lie on obtaining robust estimates of fixed effects and their precisions (Seltzer, 1993; Seltzer et al., 1996, 2002; Smith et al., 1995; Thompson & Becker, 2020a). An early work by West (1984) addresses the over-shrinkage issue, however, his study is based on

the assumption that the parameter variance τ^2 is known, which is problematic when the number of studies in a meta-analysis, for example, is small and the uncertainty around τ^2 is substantial. Given the limitations of the previous literature, this study aims to examine the properties of shrinkage estimates of outliers under the t-distributional assumption for random effects in FB meta-analysis models.

For the implementation of this approach, I build on the work of Seltzer (1993) and Seltzer et al. (1996), who focus on Bayesian analysis employing t-distributional assumptions for random effects in standard multilevel settings. Specifically, Markov Chain Monte Carlo (MCMC) will be used to implement meta-analysis models based on heavy-tailed distribution for random effects. In this way, I will be able to obtain the marginal posterior distribution of the effect-size parameter for each study in the dataset. In contrast to the models based on ML, the uncertainty regarding the magnitude of parameters, e.g., parameter variance τ^2, will be reflected in the marginal posterior distributions of the effect-size parameters. Of particular interest will be the marginal posterior distributions of outlying effect sizes, which will provide the information concerning the probability that the true effect size for the outlying study exceeds a value of particular interest from a substantive standpoint.

1.4. Outline of dissertation

The organization of the dissertation is as follows. In Chapter 2, I lay out the steps and logic of the MCMC algorithm I will be using to estimate meta-analysis models in which t-distributional assumptions are specified for the between-study model. In Chapter 3, I will then

illustrate the use and value of FB approach based on t-distributional assumptions through analyses of the empirical dataset from a meta-analysis focusing on the effects of school-based writing-to-learn interventions, conducted by Bangert-Drowns et al. (2004). In Chapter 4, the properties of the shrinkage estimates from the EB estimation method and FB models assuming a normal distribution as well as two t-distributions with different degrees of freedom are examined and compared through a targeted simulation study. In Chapter 5, I will summarize the primary findings and will discuss the implications and limitations of the current study.

CHAPTER 2

Fully Bayesian Random Effects Models for Meta-Analysis

This chapter presents a Gibbs sampling algorithm to estimate random-effects models for meta-analysis within a fully Bayesian (FB) framework. In the following sections, I first provide an overview of the Gibbs sampler and describe the steps to obtain marginal posterior distributions of parameters of interest in FB models. Specifically, I introduce a standard FB approach based on a normality assumption for the random effects (e.g., the true effect sizes), and then focus on an alternative FB model assuming a heavy-tailed distribution for the random effects in a meta-analysis. For simplicity, I mostly focus on the unconditional model without predictors.

2.1. Gibbs sampling algorithm to implement fully Bayesian random-effects models

In this study, I employ an FB approach that has been used to yield robust estimates of parameters of interest in meta-analyses with a small number of studies (Larose & Dey, 1997; Seltzer, 1990; Smith et al., 1995). Within an FB framework, meta-analysis random-effects models can be estimated based on t-distributional assumptions with small degrees of freedom for the between-study model. Further, the marginal posterior distributions of parameters of interest from FB models reflect uncertainty in all other unknowns in the models. FB methods require

specifying prior distributions for all of the unknowns in one's model, which quantifies the prior knowledge or belief regarding the unknowns. Combined with the likelihood function for the observed data, the full posterior distribution is proportional to the joint distribution of all prior distributions in the model and the likelihood. From the joint posterior, the marginal posterior distribution of a parameter of interest can be obtained by integrating over all other unknowns. However, the required integrations are intractable in most cases, except in certain cases with, for example, a very small number of unknowns. As a viable alternative, one can use Markov-Chain Monte Carlo (MCMC) to obtain marginal posterior distributions of parameters of interest in high-dimensional settings, and in which distributional assumptions other than normality assumptions are specified. As noted in the previous chapter, I will demonstrate and explain how employing t-distributional assumptions with small degrees of freedom in meta-analysis models can alleviate over-shrinkage problems connected with the EB approach based on normality assumptions for the random effects.

To help illustrate the logic of the MCMC approach for FB models, suppose that one wants to estimate three parameters, θ_1, θ_2, and θ_3 given the data, $\mathbf{y}$. With the likelihood of $p(\mathbf{y}|\theta_1, \theta_2, \theta_3)$ and the prior distributions for $p(\theta_1)$, $p(\theta_2)$ and $p(\theta_3)$, the joint posterior distribution is:

$$\mathrm{p}(\theta_1, \theta_2, \theta_3|\mathbf{y}) \propto \mathrm{p}(\mathbf{y}|\theta_1, \theta_2, \theta_3)\mathrm{p}(\theta_1)\mathrm{p}(\theta_2)\mathrm{p}(\theta_3). \quad (2.1)$$

Obtaining the marginal posterior distributions, $p(\theta_1|\mathbf{y})$, $p(\theta_2|\mathbf{y})$, and $p(\theta_3|\mathbf{y})$ directly from the full joint posterior, or carrying out the required integrations, is typically not feasible, especially for complex models such as random-effect models in which t-distributional assumptions are specified for random effects.

To overcome this difficulty, one can adapt simulation methods to obtain accurate simulations of marginal posterior distributions of interest. Note that MCMC methods such as Gibbs sampling or Metropolis-Hastings have been widely applied to estimate high-dimensional, complex models (Gelman et al., 2013; Gilks et al., 1995). As will be seen, the Gibbs sampler is fairly straightforward to implement in the case of FB random-effects models for meta-analysis since the conditional distributions of unknowns in the model have known forms (e.g., normal, gamma, etc.) (Gelfand et al., 1990; Tanner & Wong, 1987).

Returning to the example above involving three unknowns, a Gibbs sampling algorithm proceeds as follows:

(1) To initiate a Gibbs sampling algorithm, compute starting values for the unknown variables, i.e. $\theta^{(0)} = \left\{\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}\right\}$.

(2) A value is drawn or sampled, (i.e., $\theta_1^{(1)}$) from the full conditional distribution of θ_1 given the current value of all other variables and the data, $p\left(\theta_1 \middle| \mathbf{y}, \theta_2 = \theta_2^{(0)}, \theta_3 = \theta_3^{(0)}\right)$.

(3) Using the updated value of θ_1, $\theta_2^{(1)}$ is sampled from $p\left(\theta_2 \middle| \mathbf{y}, \theta_3 = \theta_3^{(0)}, \theta_1 = \theta_1^{(1)}\right)$.

(4) Using the recent updates, $\theta_3^{(1)}$ is sampled from $p\left(\theta_3 \middle| \mathbf{y}, \theta_1 = \theta_1^{(1)}, \theta_2 = \theta_2^{(1)}\right)$.

(5) Now we have a complete vector for the first cycle, $\theta^{(1)} = \left\{\theta_1^{(1)}, \theta_2^{(1)}, \theta_3^{(1)}\right\}$.

(6) Repeating (2) - (5) for c times until convergence, and with a large number of additional iterations (i.e., m iterations), we essentially have m draws from the joint posterior distribution $p(\theta_1, \theta_2, \theta_3|\mathbf{y})$:

$$\left\{\theta_1^{(c+1)}, \theta_2^{(c+1)}, \theta_3^{(c+1)}\right\}$$

$$\left\{\theta_1^{(c+2)}, \theta_2^{(c+2)}, \theta_3^{(c+2)}\right\}$$

.

.

.

$$\left\{\theta_1^{(c+m)}, \theta_2^{(c+m)}, \theta_3^{(c+m)}\right\}.$$

If one is interested in the posterior distribution of θ_1, a histogram can be constructed using the m values generated for θ_1, i.e., a plot of the set of values $\theta_1^{(c+1)}$ to $\theta_1^{(c+m)}$, which provides an accurate approximation of the marginal posterior distribution of $\theta^{(1)}$. The mode of the plot would also provide an accurate estimate of the mode of the marginal posterior of θ_1. In addition, one could calculate the proportion of values generated for θ_1 that lie above a value of substantive interest, which would provide an accurate estimate of the probability that θ_1 lies above that value based on the marginal posterior distribution of θ_1. Furthermore, one can compute the value below which 2.5 percent of the distribution lies, and the value above which 97.5% lies, which would provide a range between which 95% of the distribution lies; we might view this as a 95% Credible Interval (CI).

Note that if the marginal posterior distribution of θ_2 is of interest, one could go through the same steps as above using the m set of values generated for θ_2 to construct a plot of the marginal

posterior distribution of θ_2, identify the mode of the distribution, construct a 95% CI, calculate the mean and standard deviation of the marginal posterior and the like. In short, we can obtain highly accurate approximations of the marginal posterior for each unknown in the model.

2.2. Fully Bayesian meta-analysis models assuming normally distributed random effects

In this section, I focus on an FB meta-analysis model without any predictors, assuming a normal distribution for the random effects in the between-study model. A Gibbs sampling algorithm to obtain the marginal posterior distributions of parameters of interest is described accordingly. For the illustration, all prior distributions are assumed to be uniform, e.g., $p(\tau^2) \propto \mathrm{c}$ where c is a constant.

In a within-study model in a meta-analysis, a standardized effect-size estimate, d_i, for the i^{th} study, $i = 1, \cdots, k$, is assumed to have a normal distribution with a mean equal to the true effect size for study *i* (i.e., δ_i), and an error variance of s_i^2,

$$d_i|\delta_i, s_i^2 \sim N(\delta_i, s_i^2). \tag{2.2}$$

The observed s_i^2 values in a meta-analysis are typically treated as known values, and the corresponding density function for the observed effect size estimates **d** in the within-study model above is only conditioned on $\boldsymbol{\delta}$, i.e., the vector of true effect sizes, which are unknown:

$$p(\boldsymbol{d}|\boldsymbol{\delta}) \propto \prod_{i=1}^{k} \left(\frac{1}{s_i^2}\right)^{\frac{1}{2}} \exp\left[-\frac{1}{2s_i^2}(d_i - \delta_i)^2\right]. \tag{2.3}$$

In the between-study model, the effect size parameter δ_i is a function of the average effect size, μ, and the random effects, the u_i's, that capture the differences between the the true effect

sizes δ_i's. The δ_i's are assumed to be normally distributed with mean of zero and effect-size variance τ^2:

$$\delta_i|\mu, \tau^2 \sim N(\mu, \tau^2). \tag{2.4}$$

The density function for $\boldsymbol{\delta}$ is:

$$p(\boldsymbol{\delta}|\mu, \tau^2) \propto \prod_{i=1}^{k} \left(\frac{1}{\tau^2}\right)^{\frac{1}{2}} \exp\left[-\frac{1}{2\tau^2}(\delta_i - \mu)^2\right]. \tag{2.5}$$

In meta-analysis models, the unknown parameters to be estimated are study-specific effect sizes δ_i, the grand mean μ, and the heterogeneity parameter τ^2. To implement a Gibbs sampler, the joint posterior distribution of these parameters given the data is determined first as follows:

$$p(\boldsymbol{\delta}, \mu, \tau^2|\mathbf{d}) \propto p(\mathbf{d}|\boldsymbol{\delta})\, p(\boldsymbol{\delta}|\mu, \tau^2)\, p(\mu)\, p(\tau^2). \tag{2.6}$$

Note that in a meta-regression model with predictors, which will be used in later analysis, δ_i is assumed to be normally distributed with a conditional mean $\mathbf{X}_i\boldsymbol{\beta}$ and a conditional variance τ^2 where $\mathbf{X}_i$ is $1 \times p$ study-level covariates, and $\boldsymbol{\beta}$ is $p \times 1$ regression coefficients. As such, the average effect size μ will be replaced with $\mathbf{X}_i\boldsymbol{\beta}$ in Equations 2.4, 2.5 and 2.6. In a conditional model, the unknown paramters to be estimated are δ_i, $\boldsymbol{\beta}$ and τ^2.

As mentioned previously, to obtain the marginal posterior distribution of a parameter of interest in complex models with many unknowns as in the above joint posterior, it is generally not possible to integrate over all of the unknowns. However, alternatively, the Gibbs sampler provides a viable approach to obtain the marginal posterior distributions of unknowns of interest. For each parameter in complex joint posterior distributions, it is possible to sample from the

conditional posterior distribution of each unknown, given the current values of all other unknowns. In many cases, the conditional posterior distributions that one needs to sample from have a known distributional form (e.g., normal, gamma, etc.), which makes it straightforward to generate sample values.

In this section, each step of a Gibbs sampling algorithm for FB meta-analysis models assuming a normal distribution for random effects is described as follows.

Step 0: Starting values are defined. In order to initialize the Gibbs sampling algorithm, one can use REML estimates of the grand mean μ and paramter variance τ^2 as starting values for these parameters: $\mu^{(0)} = \mu_{\text{reml}}$ and $\tau^{2(0)} = \tau^2_{\text{reml}}$. Note that in meta-regression models, the REML estimates of fixed effects $\boldsymbol{\beta}$ are used as starting values such that $\boldsymbol{\beta}^{(0)} = \boldsymbol{\beta}_{\text{reml}}$ instead of $\mu^{(0)} = \mu_{\text{reml}}$.

Step 1: Sample $\boldsymbol{\delta}$ from $p(\boldsymbol{\delta}|\mathbf{d}, \mu, \tau^2)$.

After dropping parts of the joint posterior distribution that reduce to constants after conditioning on the current values of all other unknowns (i.e., μ and τ^2) and the data, the conditional distribution of δ_i is normal as follows:

$$\delta_i|\mu, \tau^2, \mathrm{d_i} \sim N\left(\lambda_i * d_i \ + \ (1-\lambda_i) * \mu, \left(\frac{1}{s_i^2} + \frac{1}{\tau^2}\right)^{-1}\right), \tag{2.7}$$

where $\lambda_i = \frac{\tau^2}{s_i^2+\tau^2}$.

As in a standard two-level multilevel model with normally distributed data and a normal prior, the conditional posterior of δ_i given the current values of μ and τ^2 and given the data, $\mathrm{d_i}$, is normal with a mean that is the weighted sum of d_i and the average effect size, μ. As the form

of the weight λ_i implies, more weight will be placed on the observed study result d_i rather than the grand mean μ when the ratio of parameter variance to the total variance is large, i.e., close to a value of 1. Note that in the conditional model with the predictors in the between-study model, the conditional distribution of δ_i in Equation 2.7 is the same except that the average effect size μ would be replaced with the conditional mean for study *i*, $X_i\boldsymbol{\beta}$**.**

Step 2: Sample μ from $p(\mu|\mathbf{d}, \tau^2, \boldsymbol{\delta})$.

The form of the conditional posterior distribution for the average effect size μ given the observed effect sizes in the sample of studies, the current values for the true values of the effect sizes $\boldsymbol{\delta}$ and effect-size variance τ^2, is a normal distribution as follows:

$$\mu|\mathbf{d}, \tau^2, \boldsymbol{\delta} \sim N\left(\frac{1}{\mathrm{k}}\sum_{i=1}^{k} \delta_i, \left(\frac{\mathrm{k}}{\tau^2}\right)^{-1}\right). \tag{2.8}$$

The conditional mean and variance of μ corresponds to the results from a regression model in which the current values of $\boldsymbol{\delta}$ are regressed on the intercept. In a meta-regression model with predictors, the fixed-effects $\boldsymbol{\beta}$ will be sampled from the conditional posterior distribution of $\boldsymbol{\beta}$ given $\boldsymbol{\delta}$ and τ^2, which is a normal distribution wih mean of $\left(\sum_{i=1}^{k} \mathbf{X}_i'\mathbf{X}_i\right)^{-1}\left(\sum_{i=1}^{k} \mathbf{X}_i\delta_i\right)'$ and variance of $\left(\frac{1}{\tau^2}\sum_{i=1}^{k} \mathbf{X}_i'\mathbf{X}_i\right)^{-1}$.

Step 3: Sample τ^2 from $p(\tau^2|\boldsymbol{d}, \boldsymbol{\delta}, \mu)$.

The conditional distribution of τ^2 is an inverse gamma distribution with shape parameter α, and scale parameter β:

$$\tau^2|\boldsymbol{d}, \boldsymbol{\delta}, \mu \sim \text{Inv} - \text{Gamma}\left(\alpha = \frac{k}{2} - 1, \beta = \frac{\sum_{i=1}^{k} (\delta_i - \mu)^2}{2}\right). \quad (2.9)$$

Note that k is the number of studies in the sample. The mode of τ^2 is $\tau^2_{\text{mode}} = \frac{\beta}{\alpha+1} = \frac{\sum_{i=1}^{k} (\delta_i - \mu)^2}{\text{k}}$, which is determined by the extent to which the study-specific effects δ_i deviate from the overall mean μ. Under the assumption of normally-distributed random effects, the τ^2_{mode} would be sensitive to outlying effect sizes when the sample size k is small. In the conditional model, the overall average μ in Equation 2.9 will be replaced with the conditional mean $X_i\boldsymbol{\beta}$.

Note that if $\tau^2 \sim \text{Inv} - \text{Gamma}(\alpha, \beta)$, then $1/\tau^2 \sim \text{Gamma}(\alpha, 1/\beta)$. Often the latter is used in Bayesian analysis because it is easier to sample from a gamma distribution rather than from its inverse.

As shown in the case of the posterior distribution of δ_i in Step 1, the weight parameter that determines the amount of shrinkage is the variance ratio $\lambda_i = \frac{\tau^2}{s_i^2 + \tau^2}$. Under the standard assumption of normally-distributed random-effects, the shrinkage of individual effect-size estimates, d_i , toward the grand mean, μ, can be severe especially for outlying effect sizes. A large amount of bias in the estimate of the true effect size for an outlying study can be problematic if research interests focus on how large the largest true effect size might be for the program or policy of interest, and under what conditions this might occur. As such, an alternative

approach is requred to reduce the bias in estimates of true effect sizes in the case of outlying observed effect sizes. One potentially valuable approach is to replace the standard normal distributional assumption for random effects with an alternative distributional assumption with heavier tails at both ends of the distribution, specifically, a t-distribution with small degrees of freedom, as discussed in the previous chapter.

2.3. Fully Bayesian meta-analysis models assuming t-distributed random effects

In meta-regression analyses, the errors of the between-sttudy model are often assumed to be normally distributed. Under the normality assumption, however, the shrinkage of outlying effect-size estimates toward the grand mean (or toward a conditional mean based on a study's level-2 covariate values), can be substantial compared to non-outliers. Recall the example in the first chapter that focused on a scenario where the average effect size was 0.2, and one of the studies in the sample reported an effect-size estimate that was considerbly larger, i.e., 1.2. With equal weight placed on the data and the grand mean (i.e., λ_i = 0.5, and 1 - λ_i = 0.5), the resulting shrinkage estimate was 0.7 as follws:

$$\delta_i^* = 0.5 * 1.2 + 0.5 * 0.2 = 0.7.$$

While the observed effect size is 1.2, the EB estimate (or shrinkage estimate) is 0.7. As mentioned previously, that is a substantial reduction of half an effect size.

In order to obtain robust paramter estimates in the presence of outliers, the distribution of random effects can be assumed to follow a heavy-tailed distribution, i.e., a t-distribution with the

degrees of freemdom ν set to a small value, e.g., 3, which is capable of accommodating extreme values in either tail.

An FB meta-anlaysis model with t-distributed random effects can be specified by reparameterizing the error distributionin in the between-study model. Consider a random variable T distributed according to a standard t-distribution with ν degrees of freedom, i.e. $T \sim t(0, \tau, \nu)$ where a scale parameter τ is equal to 1. The standard t-distribution T, can be obtained by taking the quotient of independent random variables, Z, which follows a standard normal disribution, i.e., $Z \sim N(0,1)$ and Q, which is follows a chi-squared distribution with ν degrees of freedom, $Q \sim \chi^2_\nu$:

$$T = \frac{Z}{\sqrt{Q/\nu}}. \tag{2.10}$$

A general t-distribution U with scale τ is:

$$U = \tau * T = \tau * \frac{Z}{\sqrt{Q/\nu}}\ . \tag{2.11}$$

In the same way, a random effect u_i is assumed to follow a scaled t-distribution, $u_i \sim t_\nu(0, \tau)$ and can be expressed as follows:

$$u_i = \tau * \frac{z_i}{\sqrt{q_i}}\ , \tag{2.12}$$

where $z_i \sim N(0,1)$ and $q_i \sim \chi^2_\nu/\nu$.

With t-distributed errors, the between-study model then becomes

$$\delta_i = \mu + u_i = \mu + \tau \frac{z_i}{\sqrt{q_i}} . \tag{2.13}$$

This between-study model with t-distirbuted random effects can be re-expressed as a normal distribution conditioned on μ, τ^2, q_i :

$$\delta_i|\mu, \tau^2, q_i \sim N(\mu, \tau^2/q_i) \tag{2.14}$$

where $q_i \sim \chi_\nu^2/\nu$.

Thus, in the density function for a between-study model, δ_i, as before, is viewed as following a normal distribution with a mean of μ, which is the average effect size across the studies. However, an effect-size variance is now specified as τ^2/q_i, where q_i is a parameter assumed to be chi-squared distributed with ν degrees of freedom, and divided by ν , i.e., $q_i \sim \chi_\nu^2/\nu$ (or equivalently, $q_i \sim \Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ which is a gamma distribution):

$$p(\boldsymbol{\delta}|\mu, \tau^2, \boldsymbol{q}) \propto \prod_{i=1}^{k} \left(\frac{1}{\tau^2/q_i}\right)^{\frac{1}{2}} \exp\left[-\frac{1}{2\tau^2/q_i}(\delta_i - \mu)^2\right]. \tag{2.15}$$

This is the normal-gamma mixture of the t-distribution with degrees of freedom ν, location parameter μ, and scale parameter τ^2, i.e., $t_\nu(\mu, \tau^2)$ (Gelman et al., 2013; Lange et al., 1989; Seltzer et al., 1996; Spiegelhalter et al., 2003). Note also that as the degrees of freedom ν decreases, the tails of the random-effects distributions become heavier, with more capability to accommodate outliers.

The joint posterior distribution of all prior distributions and the likelihood is

$$p(\boldsymbol{\delta}, \mu, \tau^2, \boldsymbol{q}|\boldsymbol{d}) \propto p(\boldsymbol{d}|\boldsymbol{\delta})\, p(\boldsymbol{\delta}|\mu, \tau^2, \boldsymbol{q})\, p(\boldsymbol{q})\, p(\mu)\, p(\tau^2). \tag{2.16}$$

The grand mean μ in Equations 2.14, 2.15, and 2.16 is replaced with a conditoinal mean $X_i\boldsymbol{\beta}$ in meta-regression models. The unknown paramters to be estimated would include $\boldsymbol{\beta}$ as well as $\boldsymbol{\delta}$, τ^2 and $\boldsymbol{q}$.

As in meta-analysis models assuming normally-distributed random effects, a full conditional distribution for each variable is derived from the joint posterior distribution. Most of the conditional distributions are similar to the previous example, however, a gamma-variate, q_i, is included in the model assuming t-distributed random effects. In my analyses as will be seen, in addition to employing normality assumptions in the between-study model, I employ t-distributional assumptions with 3 and 7 degrees of freedom. The Gibbs sampling algorithm for FB models based on heavy-tailed distributions is presented as follows.

Step 0: As starting values, the REML estimates are obtained from the data, i.e. $\mu^{(0)} = \mu$ and $\tau^{2(0)} = \tau^2_{reml}$. For the weight parameter **q**, equal weights to all studies as $\mathbf{q^0} = \left(q_1^{(0)} = \cdots = q_k^{(0)} = 1\right)$ can be specified as initial values. As in the previous section, the starting values of $\boldsymbol{\beta}^{(0)} = \boldsymbol{\beta}_{\text{reml}}$ are used for the fixed-effects in meta-regression models.

Step 1: Sample $\boldsymbol{\delta}$ from $p(\delta_i|d_i, \mu, \tau^2)$.

The conditional distribution of the effect-size paramter δ_i is as follows:

$$\delta_i|d_i, \mu, \tau^2, q_i \sim N\left(\lambda_i * d_i + (1-\lambda_i) * \mu, \left(\frac{1}{s_i^2} + \frac{1}{\tau^2/q_i}\right)^{-1}\right). \qquad (2.17)$$

This is comparable to the normal case except that λ_i is now $\frac{\tau^2/q_i}{s_i^2+\tau^2/q_i}$. As a consequence, the amount of shrinkage is determined by τ^2/q_i as well as the degrees of freedom v_i. In meta-

regresion models with predictors, the grand mean μ in Equation 2.17 will be replaced with a conditional mean $X_i\boldsymbol{\beta}$.

Step 2: Sample μ from $p(\mu|\mathbf{d}, \mu, \tau^2, \mathbf{q})$.

The full distribution of μ is:

$$\mu|\boldsymbol{d}, \boldsymbol{\delta}, \tau^2, \boldsymbol{q} \sim N\left(\left(\sum_{i=1}^{k}\frac{1}{\tau^2/q_i}\right)^{-1}\sum_{i=1}^{k}(\tau^2/q_i)^{-1}\delta_i, \left(\sum_{i=1}^{k}\frac{1}{\tau^2/q_i}\right)^{-1}\right). \quad (2.18)$$

In Equation 2.18, each δ_i is weighted by the inverse of τ^2/q_i when computing the average effect size μ. As will be detailed in Step 4, a small value of q_i is assigned to an outlier, resulting in a large paramter variance τ^2/q_i. In turn, $(\tau^2/q_i)^{-1}$ will be very small, downweighting the estimate of true effect size of outlier δ_i in computing μ. In this way, the Bayes-t model reduces the influence of outliers on the estimates of fixed effects.

Note that in the conditional model, the fixed effects $\boldsymbol{\beta}$ will be sampled from the conditional posterior distribution with a mean of $\sum_{i=1}^{k}(\tau^2/q_i)^{-1}\delta_i\mathbf{X}_i\left(\sum_{i=1}^{k}\frac{\mathbf{X_i}'\mathbf{X_i}}{\tau^2/q_i}\right)^{-1}$ and a variance of $\left(\sum_{i=1}^{k}\frac{\mathbf{X_i}'\mathbf{X_i}}{\tau^2/q_i}\right)^{-1}$.

Step 3: Sample τ^2 from $p(\tau^2|\mathbf{d}, \mu, \boldsymbol{\delta}, \mathbf{q})$.

The conditional distribution of τ^2 is an inverse gamma distribution:

$$\tau^2|\mathbf{d}, \mu, \boldsymbol{\delta}, \mathbf{q} \sim \text{Inv} - \text{Gamma}\left(\frac{k}{2}+1, \frac{\sum q_i(\delta_i-\mu)}{2}\right). \quad (2.19)$$

The mode of the conditoinal posterior of τ^2 is defined as $\tau^2_{\text{mode}} = \frac{\beta}{\alpha+1} = \frac{\sum_{i=1}^{k} q_i(\delta_i - X_i\boldsymbol{\beta})^2}{\text{k}+2}$ under the t-distirbutional assumption. Unlike in the normal model, τ^2_{mode} is now determined by the sum of squared deviations $(\delta_i - \mu)^2$, where each squared deviation is multiplied by its current q_i value.

Step 4: Sample $\mathbf{q}$ from $p(\mathbf{q}|\mathbf{d}, \tau^2, \mu, \boldsymbol{\delta})$.

The full conditional distribution of q_i given all other variables is a gamma distribution as follows:

$$q_i|d_i, \tau^2, \mu, \delta_i \sim \text{Gamma}\left(\alpha = \frac{\nu+1}{2}, \beta = \frac{2}{\nu + \frac{1}{\tau^2}(\delta_i - \mu)^2}\right). \qquad (2.20)$$

Same as before, the grand mean μ in Equations 2.19 and 2.20 will be replaced with a conditional mean $X_i\boldsymbol{\beta}$ in meta-regression models with predictors.

Note that the posterior mode of q_i in Equation 2.20 is $\frac{\nu-1}{\nu+\frac{1}{\tau^2}(\delta_i-\mu)^2}$. In a given iteration of the Gibbs sampler, if the absolute value of the residual, $\delta_i - \mu$ is large, the mode of the conditional posterior for q_i will decrease substantially. In such cases, it is likely that a small value for q_i will be generated by the Gibbs sampler, e.g., a value slightly above 0. This in turn will result in a substantial increase in τ^2/q_i for study i. As can be seen in Equation 2.18, λ_i is now equal to $\frac{\tau^2/q_i}{s_i^2+\tau^2/q_i}$. Thus a substantial increase in τ^2/q_i that we would obtain for an outlier, will result in a value of λ_i that is near a value of 1. This means that a weight in the vicinity of a value of 1 will be placed on the data in forming an estimate of δ_i for the outlier, thus overcoming the severe shrinkage problems connected with outliers.

On the other hand, if the absolute value of the residual $\delta_i - \mu$ is very small, the mode of the conditional posterior for q_i will increase. In this situation, it is likely a large value for q_i is generated that is close to a value of 1 or larger. For example, for a value of q_i of 1, $\tau^2/q_i = \tau^2$, and λ_i would be equal to: $\frac{\tau^2}{s_i^2+\tau^2}$.

In sum, for the outlier, a large τ^2/q_i results in a large λ_i value which places more weight on the data, d_i, rather than on μ**,** when computing δ_i. To put it another way, the shrinkage toward the overall average is less severe for outliers when a heavy-tailed distribution is assumed for random effects. This means that the shrinkage estimate in the Bayes-t models with small df will be less biased toward the grand mean, ensuring some robustness against over-shrinkage of outliers compared to the EB approach or Bayes-normal models. If there is good evidence that a study with an outlying result has been implemented in a careful, rigorous way and thus can provide some sound evidence regarding the magnitude of the true treatment effect for the program being studies, the Bayes-t approach will yield shrinkage estimates of outliers that are more sensible than those based on a Bayes-normal model.

2.4. Software

In this study, all the Bayesian analyses were conducted using R 4.2.3 (R Core Team, 2023). Specifically, we used an R package R2jags (Su & Yajima, 2021) to call JAGS (Just Another Gibbs Sampler; Plummer, 2003) into the R environment. JAGS is a program to analyze Bayesian hierarchical models using MCMC, which is compatible with BUGS (Bayesian inference Using Gibbs Sampling; Lunn et al., 2000). Extending BUGS to more general settings, JAGS has been

widely used across different fields for Bayesian analysis, enabling more flexible modeling (Depaoli et al., 2016).

To implement a Gibbs sampler using JAGS, I set the number of iterations to 11,000 per each chain with the first 1,000 samples discarded to reduce the influence of starting values. I ran five chains in total, which resulted in 50,000 sample draws to approximate the marginal posterior distribution of each variable. The convergence of each parameter was monitored using graphical methods, i.e. trace and autocorrelation plots.

For the EB analysis for empirical and simulated datasets, meta-regression models were fitted using HLM8 (Raudenbush & Congdon, 2021). HLM, as in the case of JAGS, is able to process the effect-size estimates and the corresponding sampling variance from each study, i.e., it does need the full data from each study. That is, it can fit a conventional meta-analysis model using the effect-size estimates and the corresponding error variances (V-known model). The EB estimates of effect-size parameters obtained from HLM are compared with the shrinkage estimates from the FB models based on MCMC in my dissertation.

CHAPTER 3

Empirical Data Analysis Results

This chapter presents the results from analyses of an actual meta-analysis dataset, using the Bayesian approaches described in the previous chapters. The dataset was drawn from a published article by Bangert-Drowns et al. (2004). Specifically, this section focuses on shrinkage estimates for an outlier in the dataset based on four estimation approaches: (1) empirical Bayes (EB), (2) fully Bayes based on normality assumptions in the between-study model (Bayes-normal model), (3) fully Bayes based on t-distributional assumptions of random effects with 3 degrees of freedom (Bayes-t model with 3 df), and (4) fully Bayes based on t-distributional assumptions with 7 degrees of freedom (Bayes-t model with 7 df). The primary interest lies in examining how the amount of shrinkage of the estimate of the true effect size for the outlier changes depending on the heaviness of the tails in the various distributional assumptions being employed.

Analyses of the Bangert-Drowns et al. (2004) were used to compare conventional methods for meta-analysis with the robust modeling approach using a t-distribution, with an emphasis on the results for outliers. The analysis results show that the Bayes-t model with heavier tails (e.g., degrees of freedom of 3) provides robust estimates of the effect sizes of outliers, by protecting against over-shrinkage that often occurs under standard distributional

approaches. This section also demonstrates that this result also holds in the case of meta-regression models with between-study predictor variables.

3.1. Background

The meta-analysis of Bangert-Drowns et al. (2004) summarizes the research results that examine the effects of writing-to-learn programs on learning outcomes. The original sample consists of 48 effect sizes from 46 studies focusing on school-based interventions and comparisons between treatment and control conditions. The effect sizes are reported as Cohen's *d*, the standardized mean differences in academic achievement scores between individuals in treatment and comparison conditions, where a positive effect size favors the treatment group. This study reveals that the overall effect of writing-to-learn programs is positive and significant but small i.e., 0.17 (0.26 when unweighted). Among the moderators of interest that have been examined, grade level (elementary school, middle school, high school and college), the length of the intervention (in minutes) and the presence of prompts that require metacognitive reflection are significantly associated with differences in the magnitudes of the effect sizes. Specifically, in a moderator analysis using grade level, the estimate of the average effect size across the studies with participants in Grades 6-8 are lower than those based on other grades. For the length of intervention materials measured in minutes, longer assignments are associated with smaller effect sizes. Lastly, for the moderator indicating whether the writing prompts have components for metacognitive reflection, the effect sizes are larger for the studies which use prompts requiring in the treatment group in a given study to reflect on the current knowledge they have obtained, what they are confused about, and their learning processes, compared to those with intervention materials that lack such components.

The dataset has a few effect sizes deviating from the rest of the effect sizes, which can potentially influence parameter estimates. For illustrating meta-analysis models that are robust to outliers, this study analyzed a subset of the meta-analysis data obtained from Bangert-Drowns et al. (2004), only including the participants attending high schools and colleges. As the original study does not contain information about the error variances of the effect sizes, they are approximated following Viechtbauer and Cheung (2010), based on the sample sizes for control and treatment groups. The reduced sample consists of 26 effect-size estimates which are heterogeneous across studies (Figure 3-1). For illustration, a single predictor, the length of treatment measured in weeks, is included in the analysis model to explain the variation across effect sizes.

3.2. Dataset

Table 3-1 presents the effect-size estimates from 26 studies, the error variances and the predictor values. As shown in Figure 3-1, the effect-size estimates and the corresponding error variances suggest substantial heterogeneity across studies, ranging from a small-to-medium negative effect of -0.32, to an outlying effect of 1.46 favoring the treatment group. The error variances range from 0.01 to 0.27, which suggest that the precisions of the effect sizes are also variable.

From the plot, it is easily noticeable that the effect size from Study 26 deviates from the rest of the data, and is the largest in this study. Possible reasons for this extreme effect include more intensive treatment with several distinctive components of writing tasks and higher quality of instructors in Study 26 (Willey, 1988), compared to the other studies in the sample that reported relatively large effect sizes and certain study characteristics with the outlier, such as a

focus on high-school participants and the class subject of science and math. Note that the error variance of the outlying study result from Study 26 is moderate as 0.10 compared to those of the rest of the studies in the sample, ranging from 0.01 to 0.27. This suggests that the outlier might not have much influence on the estimates of fixed effects, i.e., the average effect size.

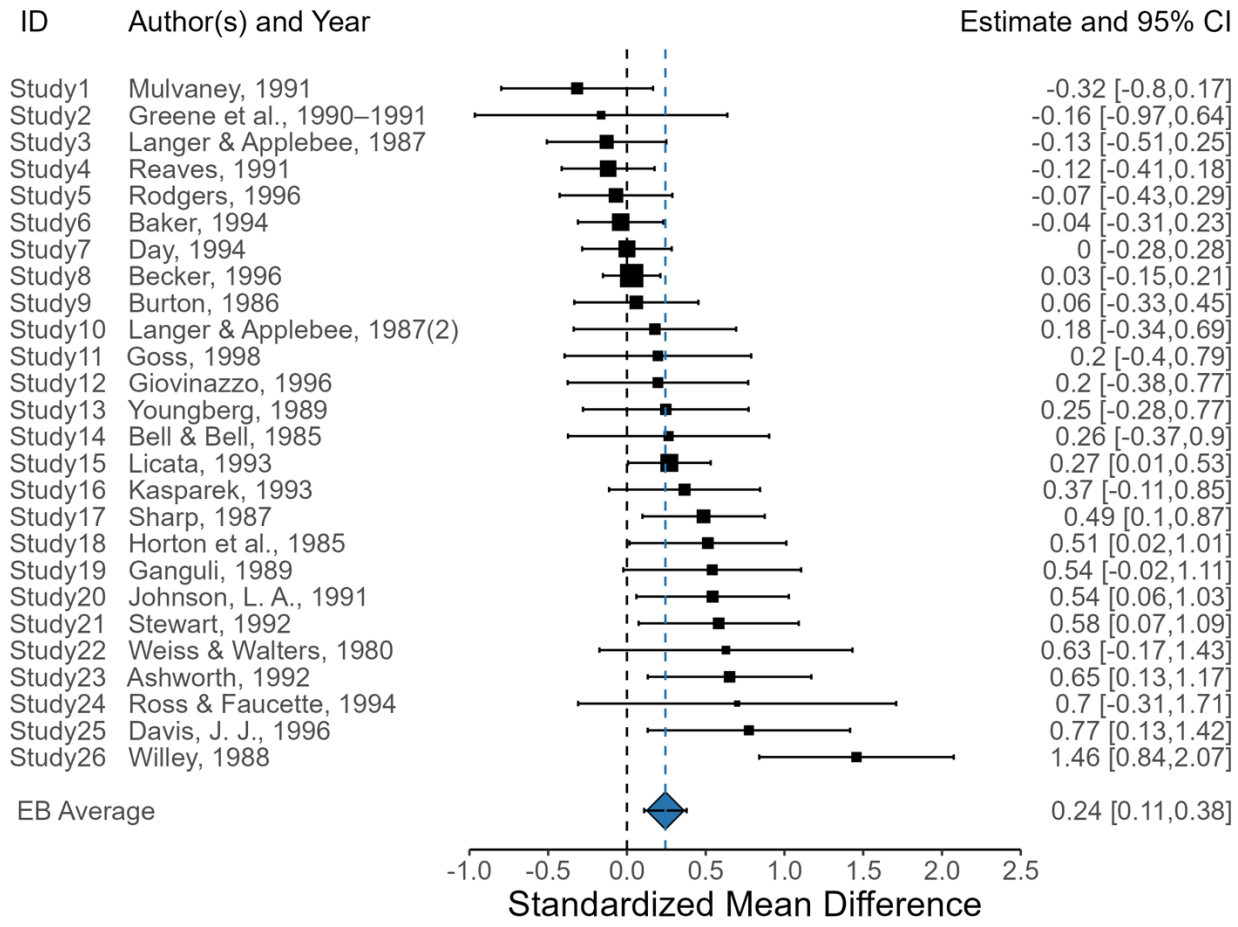

Figure 3-1 Effect-size estimates and their 95% confidence intervals from the meta-analysis conducted by Bangert-Drowns, Hurley and Wilkinson (2004).

The analysis sample consists of 26 effect-size estimates of writing-to-learn interventions implemented in high schools and colleges, which are standardized mean differences in achievement scores between treatment and control conditions. The size of squares indicates how precise an estimate is, for example, a large square means high precision, while a small square represents low precision. The blue polygon is the estimate of average effect size from the empirical Bayes method based on restricted maximum likelihood.

Table 3-1 Meta-analysis dataset from Bangert-Drowns, Hurley and Wilkinson (2004)

Study	Author (s)	Year	Effect Size	Error Variance	Duration of Treatment (In Week)
1	Mulvaney	1991	-0.316	0.060	11
2	Greene et al.	1990-1991	-0.164	0.167	4
3	Langer & Applebee (1)	1987	-0.129	0.037	1
4	Reaves	1991	-0.119	0.023	1
5	Rodgers	1996	-0.070	0.033	15
6	Baker	1994	-0.040	0.019	9
7	Day	1994	0.000	0.021	15
8	Becker	1996	0.030	0.009	1
9	Burton	1986	0.060	0.040	4
10	Langer & Applebee (2)	1987	0.178	0.069	1
11	Goss	1998	0.196	0.091	15
12	Giovinazzo	1996	0.197	0.086	14
13	Youngberg	1989	0.247	0.072	15
14	Bell & Bell	1985	0.264	0.106	4
15	Licata	1993	0.269	0.018	1
16	Kasparek	1993	0.366	0.060	12
17	Sharp	1987	0.486	0.039	2
18	Horton et al.	1985	0.514	0.065	3
19	Ganguli	1989	0.541	0.083	4
20	Johnson, L. A.	1991	0.544	0.061	19
21	Stewart	1992	0.583	0.067	24
22	Weiss & Walters	1980	0.629	0.168	15
23	Ashworth	1992	0.651	0.070	15
24	Ross & Faucette	1994	0.700	0.265	15
25	Davis, J. J.	1996	0.774	0.107	15
26	Willey	1988	1.457	0.099	15

The relationship between effect-size estimates and intervention length in weeks is illustrated in Figure 3-2. While the outlying study has a duration of 15 weeks, which is larger than the average of 9.62, it appears that the outlier is not pulling the slope capturing the relationship between duration and effect size toward it. This is due to the fact that besides the outlier, there is an appreciable number of studies with a duration of 15 weeks that have effect sizes considerably smaller than the outlier that together are exerting a lot of weight on magnitude of the slope. In other words, it is unlikely that the presence of the outlier affects the strength of relationship between the outcome and the predictor, or how steep the regression line is, in this dataset. Regardless of the fixed-effect estimates, the shrinkage estimate of the outlier, which is determined by the amount of heterogeneity across the effect sizes in the sample and the error variance of the outlying effect size, can still suffer from over-shrinkage issues. Using the current dataset, I will demonstrate the problem associated with the conventional approaches such as the EB method and an FB model assuming normally distributed random effects, and how employing a heavy-tailed distributional assumptions for random effects alleviates the excessive amount of shrinkage of the outlier.

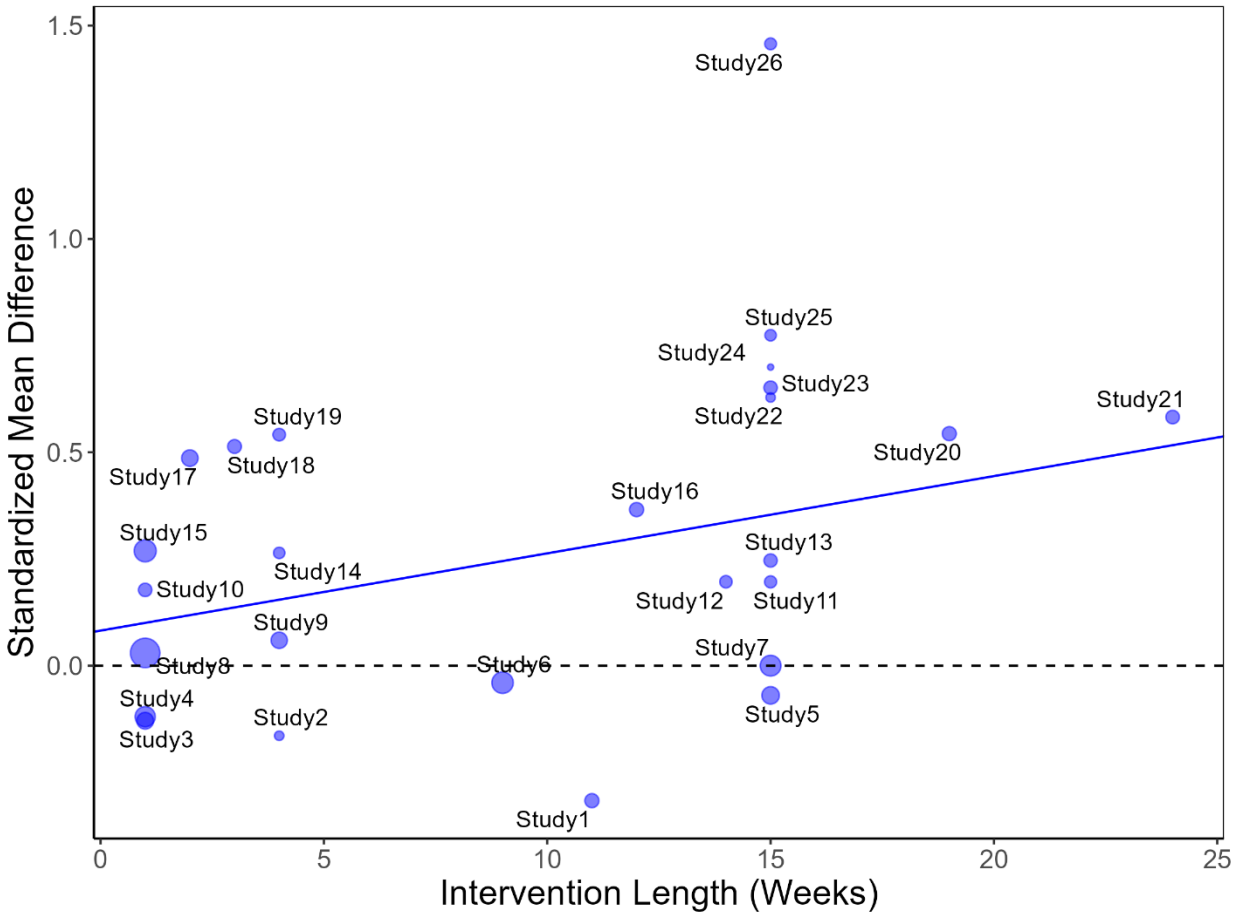

Figure 3-2 Effect-size estimates of studies by intervention length in weeks from the Bangert-Drowns, Hurley and Wilkinson (2004) study.

The average length of treatment of 26 studies in the sample is 9.62, with a minimum length of one week and a maximum duration of 24 weeks. The size of each data point represents the precision of the estimate. The positive slope between effect-size estimates and treatment length suggests that the treatment effect will increase for in the case of a longer intervention.

3.3. Empirical Bayes analysis results

The EB results for the unconditional and conditional models are presented in Table 3-2. As mentioned in the previous chapter, the EB estimates of outliers tend to shrink the effect sizes of outliers toward the grand mean by a substantial amount. The average effect size for high-school and college studies in the sample is 0.244 and the estimate of effect-size variance is 0.064. In this example, the I^2 statistic, a measure of heterogeneity across the studies in a meta-analysis, is 59.707. The I^2 is the ratio of the effect-size variance to the total variance and analogous to the intra-class correlation computed in multilevel models (Higgins & Thompson, 2002; Raudenbush & Bryk, 2002). The I^2 of 59.707 indicates that the effect-size variance is 59.707 % of the total variance. These results suggest that there is a substantial amount of heterogeneity among effect sizes, although the overall effect size is positive and fairly small.

Table 3-2 Empirical Bayes analysis results of Bangert-Drowns et al. (2004) data

	Parameter	Estimate	SE	95% CI	
				Lower	Upper
Unconditional Model	Intercept, γ_{00}	0.244	0.069	0.109	0.379
	Parameter Variance, τ^2	0.064			
	EB Estimate, δ_{26}	0.720	0.202	0.325	1.116
Conditional Model	Intercept, γ_{00}	0.256	0.067	0.125	0.388
	Trt Length, γ_{10}	0.018	0.010	-0.001	0.037
	Parameter Variance, τ^2	0.057			
	EB Estimate, δ_{26}	0.756	0.199	0.367	1.145

Note: The variance ratio, $\frac{\tau^2}{\tau^2+s_i^2}$, λ_{26} for the unconditional model is 0.392, and 0.364 in the conditional model under the EB approach.

In order to explain the varying effect sizes, the treatment length in weeks is included in the model. Table 3-2 shows that the average effect size for studies with an average treatment length (9.62 weeks) is 0.256. The regression coefficient for the treatment length is 0.018, indicating that the standardized differences between treatment and control conditions increase by 0.018 of an effect size when there is a one-week increase in treatment length. Put another way, the treatment effects are larger for the participants who received the treatment for a longer period of time, though this result is border-line significant with a p-value slightly over 0.05. Specifically, the fitted value, or the expected effect size, for a study with a minimum treatment duration of one week is 0.101, whereas the fitted value for a study with a maximum duration of 24 weeks is 0.515. This indicates that the predicted effect size increases from the null effect with a week of treatment to a medium effect when the treatment is continued for the longest period in the sample. The estimate for the remaining effect-size variance is 0.057, which is a decrease from the null model of 11.23%. The I^2 statistic is 56.05 and the *Q* statistic for the residual variance is 50.91, which is still significant after including the predictor. The conditional model results suggest that there might be other moderators related to effect sizes which can explain the residual variance.

In both the unconditional and conditional models, the shrinkage estimate of the outlying study, δ_{26}, is computed and presented in Table 3-2. The estimate of the true effect size is 0.720 with λ_{26} equal to a value of 0.392 in the null model, while it is 0.756 with λ_{26} equal to a value 0.364 in the conditional model. In other words, the weight attached to the data is 0.392, and 0.608 of the weight is attached to the average effect size to compute the shrinkage estimate in the null model. The weight in the conditional model placed on the data is 0.364, and 0.636 is placed on the average, which is similar to the null model. Given that the original effect size of Study 26

is 1.457, it is evident that the estimates of the true effect size are shrunk toward the estimated averages by a large amount under EB.

3.4. Fully Bayesian analysis results

The differences in the shrinkage estimates from EB, Bayes-Normal and Bayes-t models with 3 df and with 7 df are shown in Figure 3-3 for the unconditional model. From the plot, it is evident that employing the t-distributional assumption based on fewer degrees of freedom leads to less shrinkage of the outliers, compared to the more standard approaches.

Table 3-3 presents the summary of the marginal posterior distributions of all parameters in the unconditional model. In terms of the shrinkage estimate of the outlier, the posterior mode of the true effect size for Study 26, δ_{26}, is 0.72 with the posterior mode of λ_{26} equal to 0.47 under the Bayes-normal model without any predictor. This result means that the effect-size estimate from the data, d_{26}, receives a weight of 0.47 when computing the corresponding shrinkage estimate whereas the average effect size, $\hat{\gamma}_{00}$, tends to receive a weight of 1-0.47=0.53. It also suggests that the data are shrunk toward the average effect-size estimate by a large amount. In the Bayes-t models, the mode of the posterior distribution of δ_{26} is 1.14 , and the posterior mode of λ_{26} of 0.81 when 3 degrees of freedom are assumed, whereas the posterior mode for δ_{26} is 0.84 and the posterior mode of λ_{26} is 0.58 when the degrees of freedom are 7. The amount of shrinkage is 4.07 times larger under EB than under the Bayes-t model with 3 df for the unconditional model. Similarly, the amount of shrinkage is 2.79 times larger under the Bayes normal model compared to the Bayes-t model with 3 df without any predictors. Meanwhile, the amounts of shrinkage of outlier under the EB, the Bayes-normal model and the Bayes-t model with lighter tails are very similar to one another. The amount of shrinkage is 1.15

times larger under the EB model when compared to the Bayes-normal model, and it is 1.05 larger under the EB model than under the Bayes-t model with 7 df. These results suggest that employing the assumption of heavier tails for random effects is most effective in decreasing the amount of shrinkage of outliers.

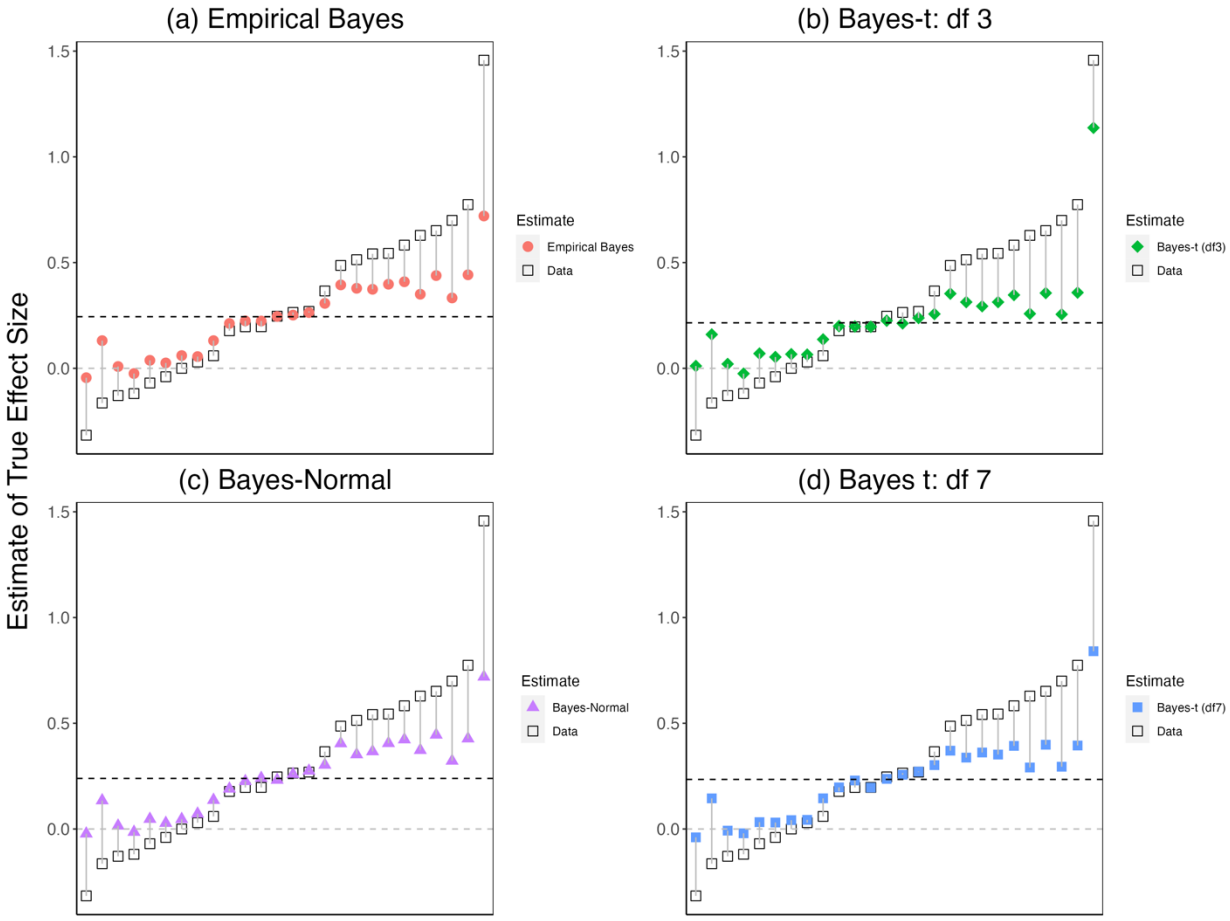

Figure 3-3 Shrinkage estimates of true effect sizes from the unconditional meta-analysis model based on Empirical Bayes, Bayes-normal and Bayes-t (3 df and 7 df) estimation methods.

For the Bayesian models, the posterior modes of shrinkage estimates are used for the comparison with EB estimates. The shrinkage estimate of the outlier with the largest effect-size estimate (the rightmost data point in each plot) shrinks toward the average effect-size estimate by the least amount under the Bayes-t model with 3 df.

Table 3-3 Fully Bayesian analysis results of Bangert-Drowns et al. (2004): Unconditional model

		Mean	Median	Mode	2.5%	97.5%	SD	p > 0
γ_{00}	Normal	0.25	0.25	0.24	0.10	0.41	0.08	1.00
	t: df3	0.22	0.21	0.21	0.08	0.37	0.07	1.00
	t: df7	0.24	0.23	0.23	0.09	0.39	0.08	1.00
τ^2	Normal	0.09	0.08	0.06	0.02	0.22	0.05	1.00
	t: df3	0.05	0.04	0.03	0.01	0.15	0.04	1.00
	t: df7	0.07	0.06	0.05	0.02	0.18	0.04	1.00
$\frac{\nu}{(\nu-2)}\tau^2$	t: df3	0.16	0.13	0.09	0.03	0.44	0.11	1.00
	t: df7	0.10	0.09	0.07	0.02	0.25	0.06	1.00
q_{26}	t: df3	0.32	0.19	0.07	0.02	1.48	0.41	1.00
	t: df7	0.62	0.52	0.37	0.11	1.68	0.41	1.00
τ^2/q_{26}	t: df3	0.50	0.25	0.10	0.02	2.43	1.40	1.00
	t: df7	0.18	0.13	0.07	0.02	0.63	0.18	1.00
δ_{26}	Normal	0.78	0.77	0.72	0.28	1.34	0.27	1.00
	t: df3	1.05	1.06	1.14	0.34	1.77	0.37	1.00
	t: df7	0.91	0.89	0.84	0.31	1.58	0.33	1.00
λ_{26}	Normal	0.44	0.45	0.47	0.17	0.69	0.13	1.00
	t: df3	0.67	0.72	0.81	0.17	0.96	0.21	1.00
	t: df7	0.55	0.56	0.58	0.17	0.86	0.18	1.00

For the conditional model, the posterior distributions of δ_{26} along with λ_{26} show that the Bayes-t models have an advantage of avoiding the over-shrinkage when a predictor is included (Figure 3-4). The conditional model results presented in Table 3-4 show that the mode of the posterior distribution of δ_{26} under the Bayes-normal model is 0.76 with λ_{26} of 0.42. In the Bayes-t models, the posterior mode of δ_{26} is 1.03 with λ_{26} of 0.76 when 3 degrees of freedom are assumed, and δ_{26} is 0.85 with λ_{26} of 0.52 with 7 degrees of freedom.

The λ's decrease in the conditional models, when compared to the unconditional models, occurs because the conditional effect-size variance is usually smaller than the unconditional one due to the variance explained by predictors, resulting in more shrinkage toward the average. Even in such a situation, assuming a heavy-tailed distribution for random effects yields shrinkage estimates which are robust to the over-shrinkage of random-effects models as illustrated in Figure 3-4. Specifically, the amount of shrinkage is 2.65-times larger in the case of the EB estimate when compared to the Bayes-t model with 3 df. The amount of shrinkage is 2.21-times larger under the Bayes-normal model, than under the Bayes-t model with 3 df. The amount of shrinkage is 1.1-times larger under the EB method than under the Bayes-normal model, indicating a negligible difference between these two approaches in terms of shrinkage estimates. The amount of shrinkage from the Baeys-t model with 7 df is also similar to the results based on the Bayes-normal model.

The results from the conditional model demonstrate a similar pattern with the unconditional model results. Specifically, the least amount of shrinkage occurs under the Bayes-t model with 3 df, while the amount of shrinkage is not very different across the EB method, the Bayes-normal model, and the Bayes-t model with 7 df. This is because the estimate of the parameter variance in the unconditional model is similar to the estimate of the conditional

variance from the meta-regression model with a predictor. As the ratio of the parameter variance to the total variance, which is sum of the parameter variance and error variance, determines the amount of shrinkage, the results will be pretty similar across the unconditional and conditional model as long as the variance ratios are close to each other.

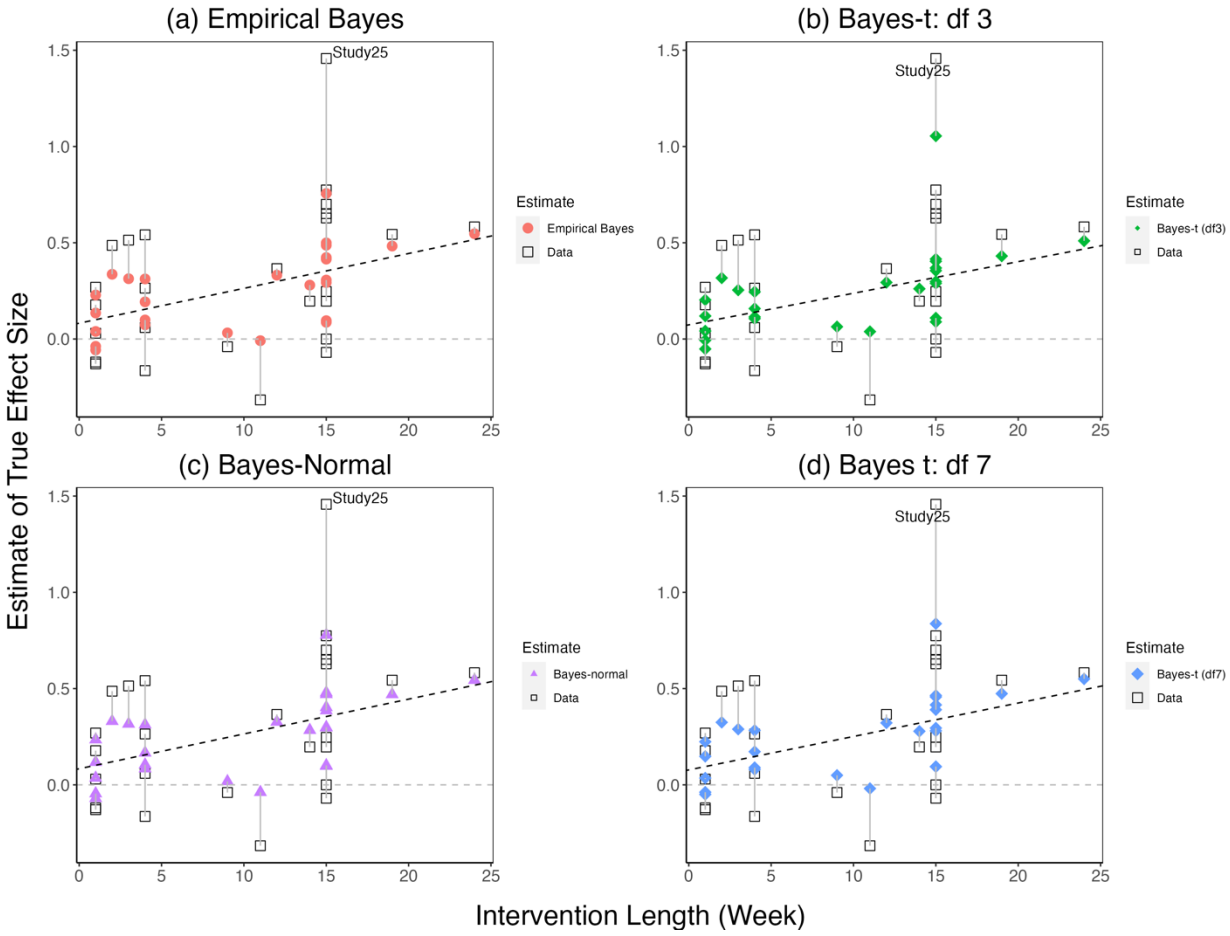

Figure 3-4 Shrinkage estimates of true effect sizes from the conditional meta-analysis model based on Empirical Bayes, Bayes-normal and Bayes-t (3 df and 7 df) estimation methods.

For the Bayesian models, the posterior modes of the shrinkage estimates are plotted to make comparisons with EB estimates. In the meta-regression model, the effect-size estimates shrink toward a conditional mean based on the predictor value, which is the intervention length in weeks of each study. As in the unconditional model, the least amount of shrinkage occurs under the Bayes-t model with 3 df.

Table 3-4 Fully Bayesian analysis results of Bangert-Drowns et al. (2004): Conditional model

		Mean	Median	Mode	2.5%	97.5%	Sd	p > 0
γ_{00}	Normal	0.26	0.26	0.26	0.12	0.42	0.08	1.00
	t: df3	0.24	0.24	0.23	0.10	0.39	0.07	1.00
	t: df7	0.25	0.25	0.24	0.11	0.40	0.08	1.00
γ_{10}	Normal	0.02	0.02	0.02	0.00	0.04	0.01	0.96
	t: df3	0.02	0.02	0.02	0.00	0.04	0.01	0.96
	t: df7	0.02	0.02	0.02	0.00	0.04	0.01	0.95
τ^2	Normal	0.08	0.07	0.05	0.02	0.20	0.05	1.00
	t: df3	0.05	0.04	0.03	0.01	0.14	0.03	1.00
	t: df7	0.07	0.06	0.05	0.02	0.17	0.04	1.00
$\frac{\nu}{(\nu-2)}\tau^2$	t: df3	0.15	0.12	0.09	0.02	0.41	0.10	1.00
	t: df7	0.09	0.08	0.06	0.02	0.24	0.06	1.00
q_{26}	t: df3	0.42	0.24	0.07	0.02	1.83	0.50	1.00
	t: df7	0.69	0.59	0.40	0.13	1.80	0.44	1.00
τ^2/q_{26}	t: df3	0.38	0.18	0.12	0.01	1.91	1.06	1.00
	t: df7	0.15	0.10	0.05	0.02	0.54	0.16	1.00
δ_{26}	Normal	0.82	0.81	0.78	0.35	1.37	0.26	1.00
	t: df3	1.02	1.01	1.05	0.36	1.73	0.36	1.00
	t: df7	0.91	0.89	0.84	0.37	1.56	0.31	1.00
λ_{26}	Normal	0.42	0.43	0.43	0.17	0.67	0.13	1.00
	t: df3	0.61	0.64	0.78	0.11	0.95	0.23	1.00
	t: df7	0.51	0.51	0.56	0.14	0.84	0.19	1.00

As noted in the previous chapter, in the case of FB t analyses with small degrees of freedom, we employ the normal-gamma representation of the t-distribution. Instead of simply assuming a common between-study variance τ^2, a key quantity in the normal-gamma formulation is τ^2/q_i where q_i is Chi-square distributed with ν degrees of freedom, and divided by ν , i.e., $q_i \sim \chi^2_\nu/\nu$.

As mentioned earlier, the posterior mode of q_i is $\frac{\nu-1}{\nu+\frac{1}{\tau^2}(\delta_i-\mu)^2}$. In a given iteration of the Gibbs sampler, if the residual $|\ \delta_i - \mu\ |$ in the denominator is large, the mode of the conditional posterior distribution of q_i will decrease appreciably. As such it is likely that in a given iteration, a small value for q_i will be generated by the Gibbs sampler, for example, a value close to 0. Note that this will result in a substantial increase in τ^2/q_i for study *i*. Note that λ_i is equal to $\frac{\tau^2/q_i}{s_i^2+\tau^2/q_i}$ under the Bayes-t model. Thus a substantial increase in τ^2/q_i for an outlier will result in a value of λ_i that approaches a value of 1. Accordingly, a weight close to 1 will be placed on the data in computing a shrinkage estimate of δ_i for the outlier, thus overcoming the over-shrinkage problems of outliers. Note also that in the case of a study in which the residual $|\ \delta_i - \mu\ |$ is small, a q_i value close to a value of 1 will tend to be generated by the Gibbs sampler, thus resulting in τ^2/q_i values close to a value of τ^2. As such, values of λ_i will tend to be small and a larger weight will be placed on the grand mean (or on a conditional mean).

As mentioned above, the differences we see in shrinkage and weight estimates can be attributed to the differences in the estimates of effect-size variances across the models. In Figure 3-5, the Bayes-t model with 3 df yields the largest posterior mode of parameter variance for δ_{26} because the estimate of the common scale parameter, τ^2, tends to be divided by small weight values of q_{26}, which depend on the residual, $\delta_{26} - \mu$. Specifically, this yields a marginal

posterior mode of $\tau^2/q_{26} = 0.10$. Note that the REML estimate of τ^2 is 0.06. As a large parameter variance places more weight on the data rather than on the overall mean, the mode of the posterior distribution of δ_{26} under the Bayes-t model with 3 df is fairly close to the effect-size estimate based strictly on the data. Figure 3-6 shows how the effect-size variance is connected to the shrinkage estimates, δ_i, using the unconditional model results from the Bayes-t model with 3 df. In the figure, it is clear that the effect-size estimate d_i shrinks toward the overall mean μ by a larger amount when the estimate of parameter variance, τ^2/q_i, is small, while the shrinkage estimate becomes closer to the study result as τ^2/q_i increases. Specifically, the shrinkage estimate of the outlier δ_{26} gets closer to the value of 1.46 from the data, as τ^2/q_{26} increases.

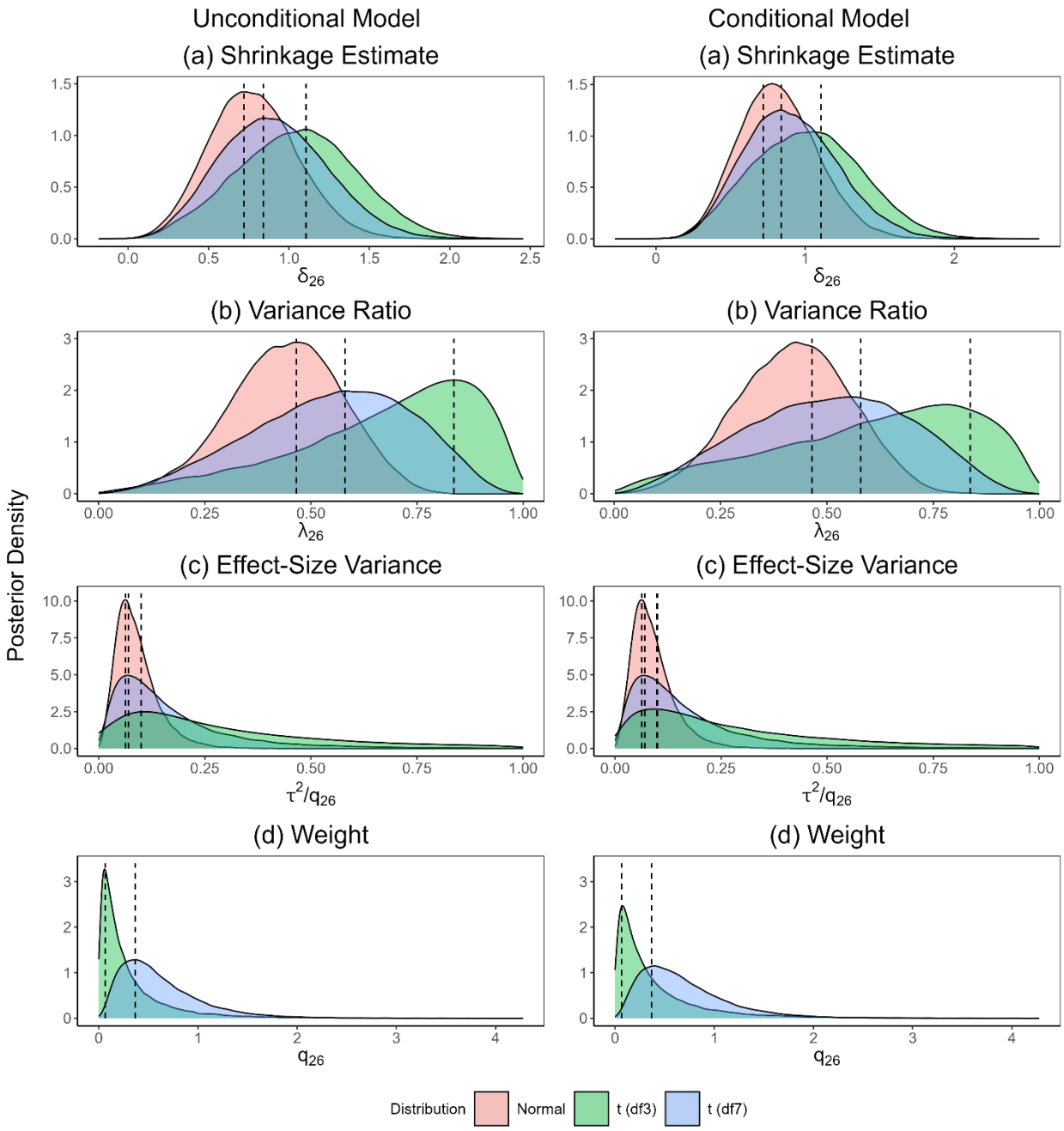

Figure 3-5 Marginal posterior distributions of δ_{26} and λ_{26} of the outlying Study 26 based on the unconditional and the conditional models.

As the tails of random effects become heavier, a small value of q_{26} is assigned for the outlier as shown in (d). With small q_{26}'s, τ^2/q_{26}'s will increase accordingly, resulting large values for λ_{26}'s. A large value of

λ_{26} places more weight on the data, d_{26}, rather than on the estimate of average effect size in the unconditional model or on the estimate of conditional mean in the meta-regression model with a predictor.

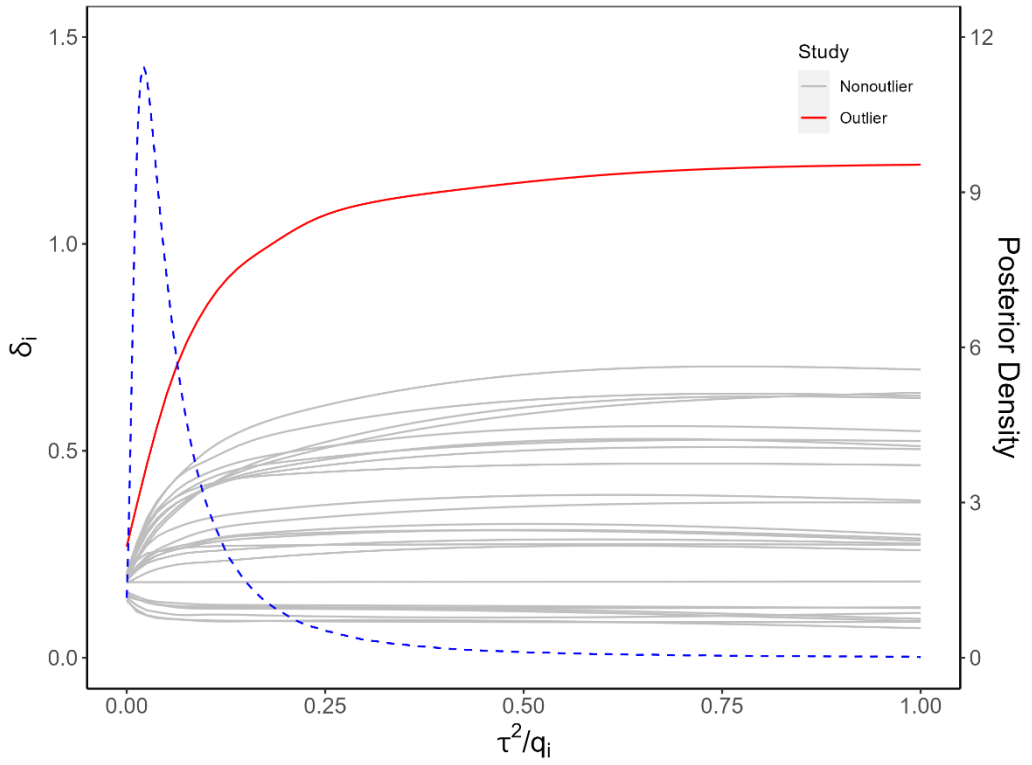

Figure 3-6 The relationship between the effect-size variance of the i[th] study, τ^2/q_i, and the i[th] shrinkage estimate, δ_i, under the Bayes-t model with 3 degrees of freedom.

When τ^2/q_i is very close to 0, all δ_i's shrink toward the grand mean, μ. As τ^2/q_i increases, δ_i is pulled closer to the data, d_i. For the outlier, Study 26 (red line), the amount of shrinkage is substantial compared to other studies (grey lines). A blue dashed line indicates the marginal posterior distribution of τ^2/q_{26}, which shows that about a half of the probability mass lies above the posterior mode. In this case, the EB estimate based on the maximum likelihood estimate of τ^2 doesn't account for the uncertainty regarding the amount of between-study variance.

Using the robust posterior distribution of the true effect size for the outlier obtained by employing a Bayes-t model, one can obtain an estimate of the probability that δ_{26} falls into a particular range of interest from a substantive standpoint. Figure 3-7 demonstrates the 95% Confidence Interval of δ_{26} from the EB method and the posterior distributions of δ_{26} obtained using FB models. In this example, I focus on the probability that δ_{26} is above an effect size value of 1 and above an effect size value of 1.5 under different distributional assumptions for the random effects. Specifically, the probability that δ_{26} exceeds 1 is the largest, i.e., 0.55 under the Bayes-t model with 3 df, while the probability is smallest, i.e., 0.21 under the Bayes-normal model. Further, the probability that δ_{26} lies above a value of 1.5, which is a value similar to the effect-size estimate of the outlier is 10 times larger based on the Bayes-t model with 3 df compared to the probability computed using the results from the Bayes-normal model. This is because the random effects under the Bayes-t model expand to a wider range due to the heavy-tailed assumption, resulting in less of shrinkage toward the average effect size than the other approaches.

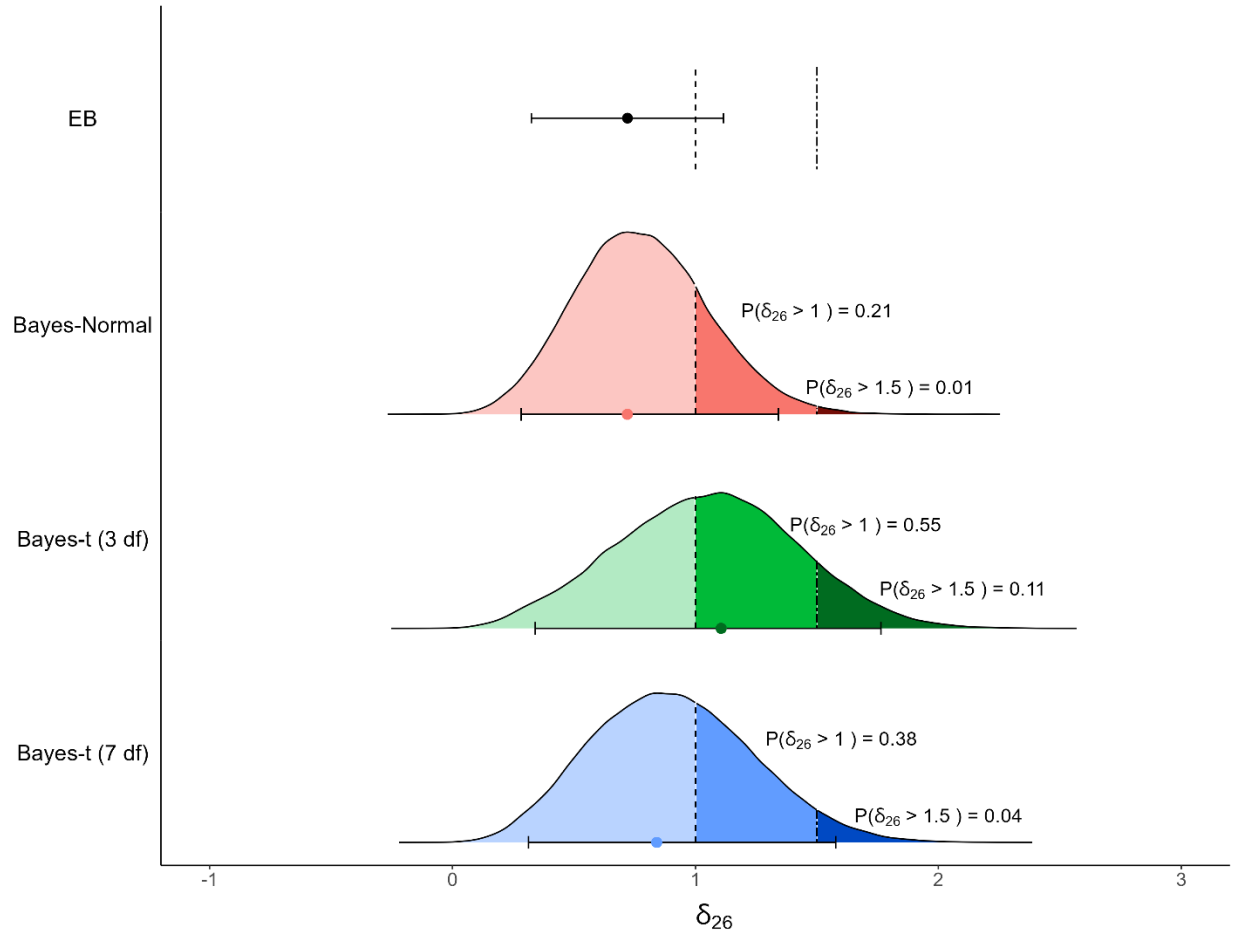

Figure 3-7 The EB 95% Confidence Interval of the shrinkage estimate of the outlier and the posterior distributions of δ_{26} from the FB models.

A darkened data point indicates the point estimate based on the EB method and posterior modes based on the FB methods. A dashed line in each method indicates Standardized Mean Difference (SMD) of 1and a two-dashed line represents an SMD of 1.5. Depending on the distributional assumption, the probability that δ_{26} exceeds 1 or 1.5 differs by a significant amount. For example, more than half of the probability mass lies above 1 under the Bayes-t model with 3 df, while the probability of δ_{26} above 1 is only 0.21 under the Bayes-normal model.

CHAPTER 4

Simulation Study Results

The primary purpose of this chapter is to assess the properties of shrinkage estimates of outliers computed using the four different estimation approaches examined in the data-analysis example. Specifically, this section compares the shrinkage estimates of outliers based on the use of Bayes-t models with 3 df and 7 df, with the estimates from more standard approaches, such as EB and the Bayes-normal model. This simulation study aims to complement the findings from the empirical data analysis in the previous chapter by using hypothetical datasets that will enable examining conditions beyond the Bangert-Drowns et al. (2004) dataset.

In this simulation study, I primarily focus on how the shrinkage estimates of outliers change depending on the ratio of between-study variance (i.e., parameter variance) to total variance (i.e., parameter variance plus error variance), denoted as I^2. The variance ratio is a key factor that determines the amount of shrinkage as discussed in the previous chapters. In particular, an interest of this study lies in figuring out the I^2's in which the Bayes-t model with small degrees of freedom is more advantageous compared to the Bayes-normal model, which is a standard approach widely used in FB meta-analysis.

As will be detailed later, the true parameter values used to generate hypothetical datasets for simulation are drawn from the REML estimates of fixed effects and between-study variance from the previous analyses of the Bangert-Drowns et al. (2004) data presented earlier. In

addition, the outlying effect-size estimate discussed earlier is included as the true effect size in the data sets generated for the simulation study. By virtue of using the parameter estimates based on the actual dataset as data-generating values, the targeted simulation approach enables one to investigate practical, real world data analysis settings, and thus has been widely advocated across the methodology literature (Browne & Draper, 2006; Bryan & Jenkins, 2016; Burton et al., 2006; see also Choi & Seltzer, 2010 for an example in educational research).

The comparisons between EB and FB methods are motivated by the simulation design conducted by Browne and Draper (2006), which uses the summary statistics of posterior distributions of parameter estimates, such as the mean, median and mode to assess the performances of different estimation approaches. In this simulation study, each evaluation criterion, including simple bias, relative bias, Mean Squared Error (MSE) and coverage, is computed for each of 200 generated datasets using each estimation approach (i.e., Bayes-t models with 3 df and 7 df, Bayes-Normal and EB). For example, in a given generated dataset, I obtain a value for simple bias, which is the distance between the posterior mode of the outlier from the true value, using the Bayes-t approaches with 3 df and 7 df, the Bayes-normal approach and the EB approach. And then the simple bias values based on the Bayes-t model with 3 df is averaged across the 200 datasets, the simple bias values based on the Bayes-t with 7 df are averaged, and the like. Note that simulation results based on the modes of posterior distributions are reported for the FB models following Browne and Draper (2006), who found that the evaluation criteria computed using posterior modes yielded the most accurate results.

4.1. Simulation conditions

This simulation study primarily focuses on the performance of each estimation approach considered in this dissertation, in the presence of outliers and depending on the ratio of true effect-size variance to the total variance, which is the sum of true effect-size variance and error variance. In this study, the variance ratio is the form of $I^2 = \frac{\tau^2}{\tau^2 + s_i^2}$. A primary goal is to find the I^2 ratio where the largest differences between the Bayes-normal model and the Bayes-t model occur.

In the simulation study, we set τ^2 equal to a value of 0.06, which is the value of τ^2 obtained in an HLM REML analysis of the dataset from Bangert-Drowns et al. (2004). These REML estimates are based on a between-study model in which the true effect sizes were modeled as a function of a grand mean μ and parameter variance τ^2. Note that the resulting estimate of μ is 0.24.

For simplicity, equal error variances are assumed for each study within a generated dataset, resulting in $I^2 = \lambda_i = \tau^2/(\tau^2 + s^2)$ in the current setting (c,f., Higgins & Thompson, 2002). We consider three I^2 values, i.e., three ratios of parameter variance to total variance: values of 0.25, 0.5, and 0.75 . With τ^2 set to a value of 0.06, which was obtained in the analyses of the Bangert-Drowns et al. (2004) dataset presented earlier, we set s^2 to a value of 0.18 to obtain a variance ratio of 0.25, a value 0.06 to obtain an I^2 value of 0.5, and a value of 0.02 to obtain an $I^2$of 0.75.

Within a given condition of I^2, 200 datasets are generated. In each dataset, there are 26 effect sizes where 25 cases follow a normal distribution, and one effect size which is an outlier. With 200 datasets for each I^2 value, four approaches, EB, Bayes-normal and Bayes-t approaches

based on 3 df and with 7 df, are assessed in terms of key evaluation criteria, for example, bias, MSE, coverage and Bayes risk, which will be elaborated later in more detail later in the next section. This simulation study also examines the properties of shrinkage estimates from these four approaches focusing on an unconditional meta-analysis model. The simulation results using a conditional model with one predictor are presented as supplementary in a later section.

4.2. Data generation process

The datasets for the simulation study are generated based on REML estimates of key parameters in the analysis of the Bangert-Drowns et al. (2004) presented earlier. Recall that the average effect size μ is 0.24 and the effect-size variance τ^2 is 0.06. In this example, the I^2 value is approximately 0.60, which means that the effect-size variance estimate comprises up to 60% of the total variance in effect sizes. In general, there are three components to be sampled to create meta-analytic datasets: 1) true effect sizes, δ_i, 2) error variances attached to each effect-size estimate s_i^2, and 3) effect-size estimates, d_i. For simplicity, this simulation study uses equal error variances for all studies within each I^2. In this section, the specific steps for data generation are presented for an unconditional model.

- Step 1: generate true effect sizes, δ_i

In constructing each dataset for the simulation study, 25 true effect sizes, δ_i, are generated following a normal distribution with a mean, μ and a variance, τ^2, for the unconditional model:

$$\delta_i \sim N(\mu, \tau^2). \quad (4.1)$$

As noted above, the true values for μ and τ^2 are set equal to the corresponding REML estimates from the data-analysis example presented earlier. For the data generation of the unconditional model, μ is set to a value of 0.24 and τ^2 is set to a value of 0.06.

In the meta-analysis dataset of Bangert-Drowns et al. (2004), the estimate of the effect size for the outlying study is 1.46. In this simulation study, we treat this as the true effect size for the outlier, and this constitutes the 26th true effect size in a generated dataset.

- Step 2: generate effect-size estimates, d_i's, of the 26 true effect sizes

In Step 2, for a given dataset that is being constructed, each effect-size estimate d_i is sampled from the corresponding normal distribution based on the true effect sizes, (i.e., the δ_i's) generated in Step 1 and the error variance s_i^2 that corresponds to a given condition I^2:

$$d_i \sim N(\delta_i, s_i^2). \quad (4.2)$$

Specifically, the 25 true effect sizes that were generated from Equation 4.1., plus the true value for the outlier,1.46, are entered into Equation 4.2, with an s_i^2 value of 0.18, 0.06 or 0.02 depending on the $I^2 = \tau^2/(\tau^2 + s^2)$ value we are focusing on (i.e., 0.25, 0.5. 0.75). Note that the outlier for each dataset is generated based on the following distribution: $d_i \sim N(1.46, s_i^2)$ with the mean of this normal distribution set to 1.46. Thus in creating each dataset for the simulation study, the true effect sizes are perturbed with a certain amount of error.

Example datasets for the simulation are presented in Figure 4-1. The plot shows that the patterns of the generated effect sizes change when the amount of error variance differs depending on I^2 ratio we are working with. Specifically, the amount that the study outcome d_i deviates

from true effect size δ_i changes depending on I^2. As the error variance s_i^2 decreases, resulting in a larger I^2, the effect-size estimate d_i will tend to be closer to its true value, δ_i, and vice versa.

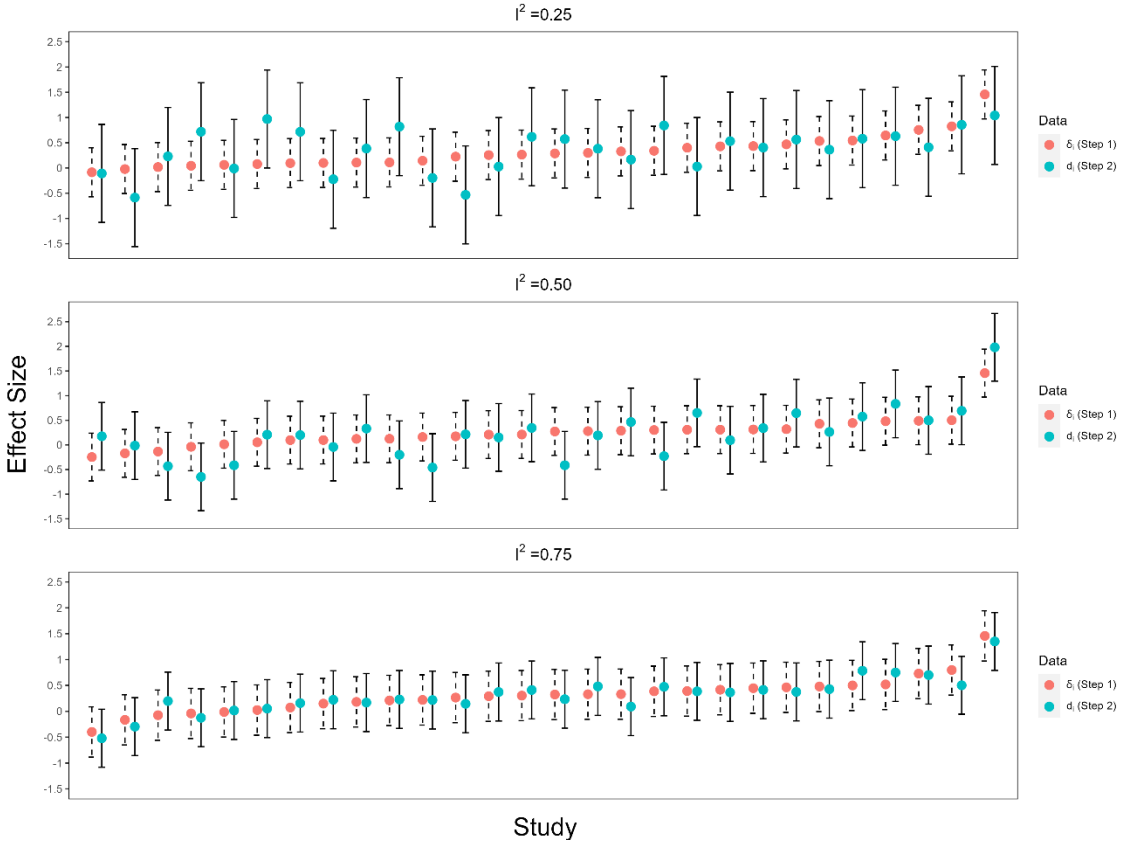

Figure 4-1 Example datasets for the targeted simulation study (Unconditional model)

Red data points represent true effect sizes, δ_i, while blue data points are study outcomes, d_i, generated based on δ_i from the previous step. As the true effect-size variance is assumed to be the same across the conditions, the error variance is the largest when $I^2 = 0.25$ with the widest 95% intervals, while the error variance is the smallest when $I^2 = 0.75$, with the narrowest intervals. Note that the true effect size of the outlier is fixed at 1.46 for all conditions as seen at the rightmost corner of each plot.

4.3. Evaluation criteria

In order to evaluate the strength of each approach in analyzing a small meta-analytic dataset, this simulation study uses the evaluation criteria such as simple and relative bias, Mean Squared Error (MSE), and the actual coverage of 95% FB and EB intervals for the true effect sizes, especially the effect size of the outlier. For each dataset generated for a given I^2 value, the usual evaluation criteria for simulation, such as bias, Mean Squared Error (MSE), and coverage of the shrinkage estimate of outlier are computed and averaged across the 200 datasets. In order to compute these statistics using the FB analysis results, the posterior modes of the marginal posterior distributions of the outlier, for example, as well as the corresponding 95% Credible Intervals are used to make comparisons with EB estimates of the outlier and their 95% Confidence Intervals (Browne & Draper, 2006) For FB models, this study also reports average and outlier Bayes risks, following the evaluation criteria in previous studies (Efron & Morris, 1971, 1972; Raudenbush & Bryk, 2002). In short, Bayes risks are the squared distances between the true and estimated values for a parameter of interest. For the Bayes risks, this study computes the ensemble risk, which is the Bayes risks averaged across all δ_i's, and the component risk only for the outlier. The details of computation of Bayes risks will be presented shortly.

4.3.1. Evaluation criteria for empirical Bayes approach

Suppose that δ_i is true effect size of study *i* for a given dataset, $\hat{\delta}_i$ is the shrinkage estimate of δ_i, and there are 200 datasets in this example. For the simulation results based on using the EB estimation approach, two bias measures, simple and relative bias, are computed for each I^2 condition. The simple bias is the average of the distance between estimates and their true

values, $\hat{\delta}_i - \delta_i$, across all 200 datasets. The relative bias quantifies the amount of bias in the estimates in comparison with their true values. Specifically, the relative bias is the percentage of the differences between estimates and true values, proportional to true effect sizes, $100 * (\hat{\delta}_i - \delta_i)/\delta_i$, averaged over the datasets. In this way, relative bias enables us to compare the results using the same scale of percentage. For both simple and relative biases, positive biases indicate that the analysis models overestimate the parameters of interest, while negative biases indicate that the parameters are being underestimated.

In this section, I also compute MSEs of EB estimates. The MSE measures how much the estimates vary around the corresponding true values, expressed as $\left(\hat{\delta}_i - \delta_i\right)^2$ averaged across the entire set of datasets. A large MSE indicates that there is a substantial amount of uncertainty regarding the parameter estimates. Along with bias and the MSE, the coverage of true parameter values are reported for each condition. The coverage is the percentage of 95% Confidence Intervals of the shrinkage estimate, $\hat{\delta}_i$, that includes the true value, δ_i. An actual coverage value close to the nominal coverage of 95% indicates that a 95% CI is estimated well.

4.3.2. Evaluation criteria for fully Bayesian models

For the marginal posterior distributions from FB models, I modified the usual evaluation criteria for simulation under frequentist approaches detailed earlier. Specifically, a draw from the marginal posterior sample of δ_i is denoted as δ_i^{FB} and the marginal posterior mode is denoted as $\tilde{\delta}_i^{FB}$. For biases and MSEs, the simulation results based on posterior modes are reported since they yield the most accurate results compared to posterior means and medians. While other

statistics remain the same as before, the simple bias is computed as the average of $\tilde{\delta}_i^{FB} - \delta_{i,}$ and the relative bias is obtained as the average of $100 * (\tilde{\delta}_i^{FB} - \delta_i)/\delta_i$. The MSE is calculated as the squared distance between the posterior mode and the true value, $\left(\tilde{\delta}_i^{FB} - \delta_i\right)^2$, averaged over the entire set of datasets. Under an FB approach, the coverage means the proportion of the 95% Credible Intervals of the marginal posterior distributions that include the corresponding true value, δ_i.

For FB models, the average and outlier Bayes risks are also presented for each simulation condition. The Bayes risk is the weighted average of MSE's, i.e., the expected value of $\left(\delta_i^{\mathrm{FB}} - \delta_i\right)^2$where the expectation is taken over the marginal posterior distribution of δ_i. In this example, the risks for each dataset are computed as follows. Within a given dataset, we obtain the marginal posterior distribution of the squared difference for each effect size, $\left(\delta_i^{\mathrm{FB}} - \delta_i\right)^2$, and focus on its posterior mean, $\mathrm{E}\left(\delta_i^{\mathrm{FB}} - \delta_i\right)^2$. The Bayes risk, or the posterior mean of squared differences using the sampled draws, is computed for the average of all 26 effect sizes, as well as for the outlier. Then the 200 posterior-mean values for the average risk and the 200 values for the outlier risk obtained for a given estimation approach are averaged, thus enabling us to compare the amount of risk in effect sizes across the different estimation approaches.

4.4. Simulation results

In this section, the simulation results are presented for the unconditional model as well as the conditional model with one predictor. For each model, the first set of simulation results include bias, MSE and coverage of shrinkage estimates. As mentioned earlier, the posterior

modes of shrinkage estimates are used to compute the evaluation criteria from FB model results. The second set of results include average and outlier Bayes risks. Overall, the simulation results show that the Bayes-t model with 3 df outperforms the other estimation approaches in terms of the shrinkage estimation of outliers, as well as the average of shrinkage estimates in the sample.

4.4.1. Simulation results I: simple bias, relative bias (%), Mean Squared Error (MSE) and coverage

The first set of simulation results include bias, MSE, and coverage of shrinkage estimates from the unconditional model. In Figure 4-1, I show that the Bayes-t model with 3 df yields the shrinkage estimate of the outlier with the least amount of bias across all conditions of I^2, and the smallest MSE for the conditions in which I^2 is 0.5 and 0.75. In addition, the Bayes-t model with 3 df is most likely to yield 95% CIs that capture the true value of the outlier 95% of the time, and this applies across all I^2's . The evaluation criteria averaged over all shrinkage estimates are similar across the estimation approaches.

The details of the results for each evaluation criterion are described as follows.

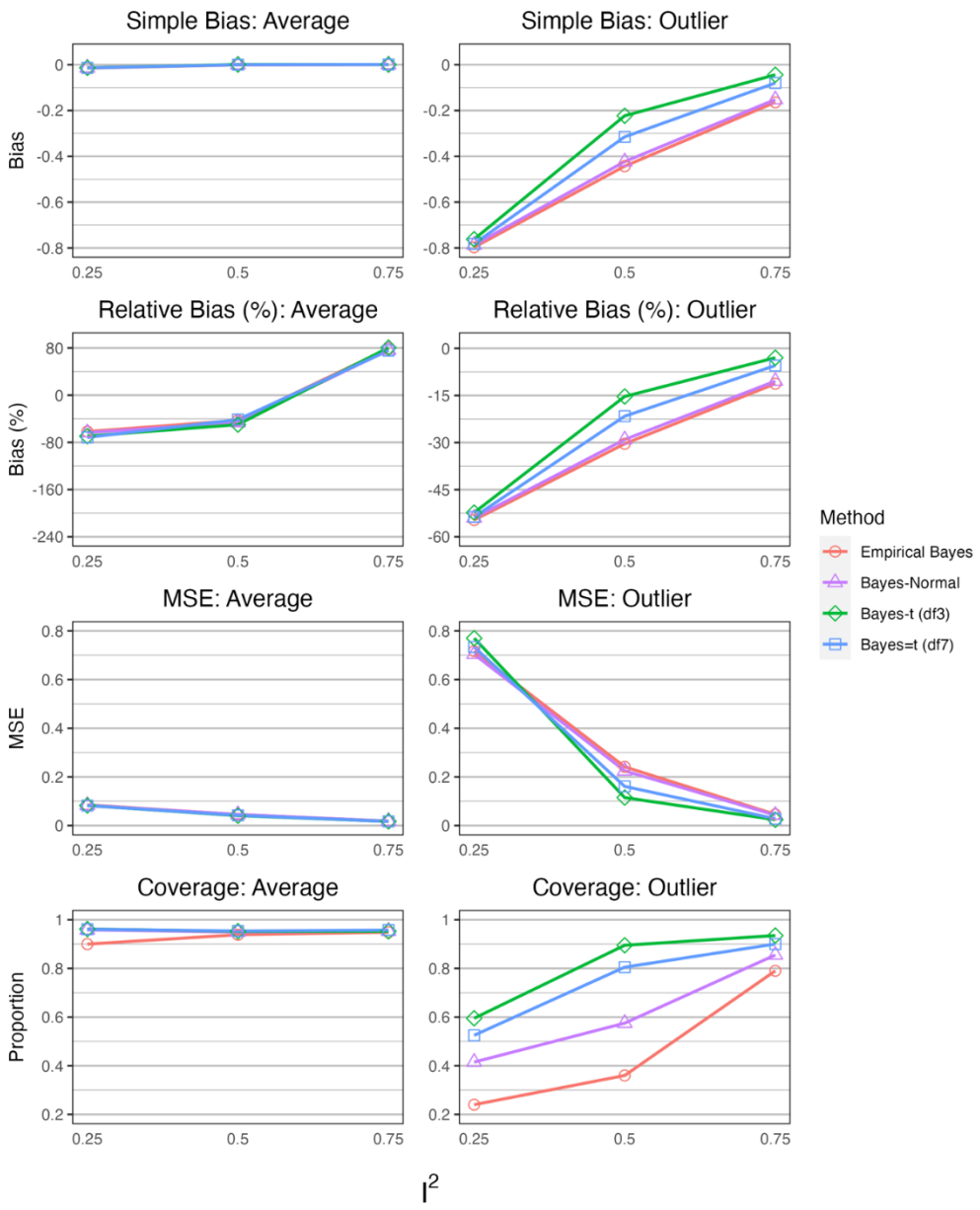

Figure 4-2 Simulation results of unconditional model I: simple bias, relative bias (%), Mean Squared Error (MSE), and coverage of shrinkage estimate of outlying effect size, δ_{26}.

For fully Bayesian models, simple bias, relative bias, and MSE were computed based on the posterior modes of these quantities. The results show that the Bayes-t model with 3 degrees of freedom outperforms the other approaches by the largest amount when $I^2 = 0.5$.

Table 4-1 Simulation results of unconditional model I: Simple bias, relative bias (%), Mean Squared Error (MSE), and coverage

	I^2	EB		FB-Normal		FB-t: df=3		FB-t: df=7	
		Outlier	Average	Outlier	Average	Outlier	Average	Outlier	Average
Simple Bias	0.25	-0.80	-0.01	-0.79	-0.01	-0.76	-0.01	-0.78	-0.01
	0.50	-0.44	0.00	-0.42	0.00	-0.22	0.00	-0.32	0.00
	0.75	-0.16	0.00	-0.15	0.00	-0.04	0.00	-0.08	0.00
Relative Bias	0.25	-54.66	-168.42	-54.00	-174.91	-52.32	-171.06	-53.67	-173.20
	0.50	-30.39	-154.16	-29.05	-122.88	-15.33	-198.00	-21.61	-138.08
	0.75	-11.26	80.23	-10.43	78.18	-2.99	82.24	-5.47	78.19
Mean Squared Error	0.25	0.72	0.08	0.70	0.08	0.77	0.08	0.73	0.08
	0.50	0.24	0.05	0.23	0.05	0.12	0.04	0.16	0.04
	0.75	0.05	0.02	0.04	0.02	0.02	0.02	0.03	0.02
Coverage	0.25	0.24	0.90	0.42	0.96	0.60	0.96	0.53	0.96
	0.50	0.36	0.94	0.58	0.95	0.90	0.95	0.82	0.95
	0.75	0.79	0.95	0.86	0.95	0.94	0.95	0.90	0.96

Simple and Relative Biases. The Bayes-t model with 3 df yields the shrinkage estimate of the outlier with the smallest bias both for simple and relative biases whereas the EB estimate of the outlier contains the largest simple and relative biases for all I^2's. The results show that all models underestimate the true effect size of the outlier, due to the shrinkage toward the grand mean. The amount of bias decreases as I^2 increases for all models. This can be attributed to the fact that the amount of shrinkage decreases as I^2 increases, regardless of the estimation approaches. For the unconditional model, the difference in biases between the Bayes-normal model and the Bayes-t model with 3 df is the largest (i.e., 0.2) when $I^2 = 0.5$.

The shrinkage estimate of the outlier under the Bayes-t model with 3 df contains the least amount of biases because the effect-size variance of the outlier is larger in the Bayes-t model than the estimate of τ^2 from the Bayes-normal model. With a large value of the effect-size

variance, there will be less shrinkage for the outlier. This will yield a shrinkage estimate close to the data (i.e., the observed effect size), resulting in less bias.

On the other hand, the average simple bias of all shrinkage estimates is very similar across the different I^2's and the different approaches. This is because the estimates of the non-outliers under the Bayes-t model will be very similar to those from the Bayes-normal model. For example, the values of q_i that are generated by the Gibbs sampler for an effect size near the grand mean will likely be close to a value of 1, which is larger than the value of q_i for the outlier. As such τ^2/q_i will be approximately equal to τ^2 in such situations, yielding the shrinkage estimates that are close to each other under the Bayes-normal model and the Bayes-t model. Consequently, the average bias of all shrinkage estimates are very similar across the models.

For the relative bias, the average biases are similar across the models for $I^2 = 0.25$, while the Bayes-t model with 3 df has the largest biases for the other conditions. Note that the relative biases are likely to be inflated if some true values for δ_i's are generated to be close to 0. With the small true values in the denominator, relative biases can appear to be extremely large.

Mean Squared Error. The shrinkage estimate of the outlier from the Bayes-t model with 3 df has the smallest MSE except for the condition $I^2 = 0.25$, where MSEs across the models are similar with the smallest MSE in the Bayes-normal model. In particular, the largest difference between the Bayes-normal model and the Bayes-t model with 3 df is observed when $I^2 = 0.5$, where the MSEs of these models are 0.23 and 0.12, respectively. The proportional reduction in MSE is similar when I^2 is 0.5 and 0.75, indicating that employing the Bayes-t model with 3 df, instead of the Bayes-normal model, leads to approximately a 50% reduction in the MSE. The

smallest MSE indicates that the Bayes-t model with 3 df provides the most accurate estimates for the outlier.

For the average MSE, the results are similar across the estimation approaches. As for the bias, the average MSE increases as I^2 decreases for all methods, though the difference across I^2's is small. The results show that the Bayes-t model with 3 df provides shrinkage estimates with MSE similar to the other methods on average, while yielding shrinkage estimates of the outlier that contain the least amount of MSE.

Coverage. The chances that a 95% interval of the shrinkage estimate captures the true effect size of the outlier are the highest in the Bayes-t model with 3 df, whereas the EB estimates suffer from the lowest coverage across all I^2's. For the condition $I^2 = 0.75$, the actual coverage of the Bayes-t model with 3 df is 94%, which is very close to the nominal value of 95%. This is because a larger value of parameter variance under the Bayes-t model with 3df, compared to the estimate of τ^2 under the Bayes-normal model, results in a wider interval for the shrinkage estimate of the outlier.

In terms of the average coverage across all shrinkage estimates, the results are similar across the various conditions, except that the EB approach yields the intervals with the lowest coverage when $I^2 = 0.25$. The actual coverages of the CI's from the FB models are close to one another, reaching a nominal value of 95% for all I^2's.

4.4.2. Simulation results II: average and outlier Bayes risks

As shown in Table 4-2, the outlier Bayes risks are in line with the MSE results in that the risks are similar across the models when $I^2 = 0.25$, whereas in two other conditions the Bayes-t model with 3 df yields shrinkage estimates with the smallest risks. The largest differences in Bayes risks between the Bayes-normal and the Bayes-t model with 3 df appear when I^2 is 0.5 as presented in Figure 4-3. The proportional reduction in the outlier Bayes risk is also the largest for the condition where I^2 is 0.5. Specifically, employing the t-distribution with 3 df for random effects, instead of assuming normally distributed random effects, leads to a 30% reduction in the Bayes risk for the outlier.

The average Bayes risks are similar across all I^2's and different estimation approaches, without further increasing ensemble risks for Bayes-t models at the cost of low outlier risks. It is noteworthy that the Bayes-t model with 3 df also yields the shrinkage estimates with lower risks on average when compared to the Bayes-normal model, though the difference is minimal. Specifically, the average risk decreases under the Bayes-t model with 3 df by the largest amount, i.e., 9% for the condition $I^2 = 0.5$, followed by 6% for $I^2 = 0.75$ and 3% for $I^2 = 0.25$, compared to the Bayes-normal model.

Table 4-2 Simulation results of unconditional model II: Average and Outlier Bayes Risks

	Normal		t: df=3		t: df=7	
I^2	Average	Outlier	Average	Outlier	Average	Outlier
0.25	0.16	0.69	0.16	0.70	0.16	0.69
0.50	0.08	0.27	0.08	0.19	0.08	0.22
0.75	0.04	0.06	0.03	0.05	0.03	0.05

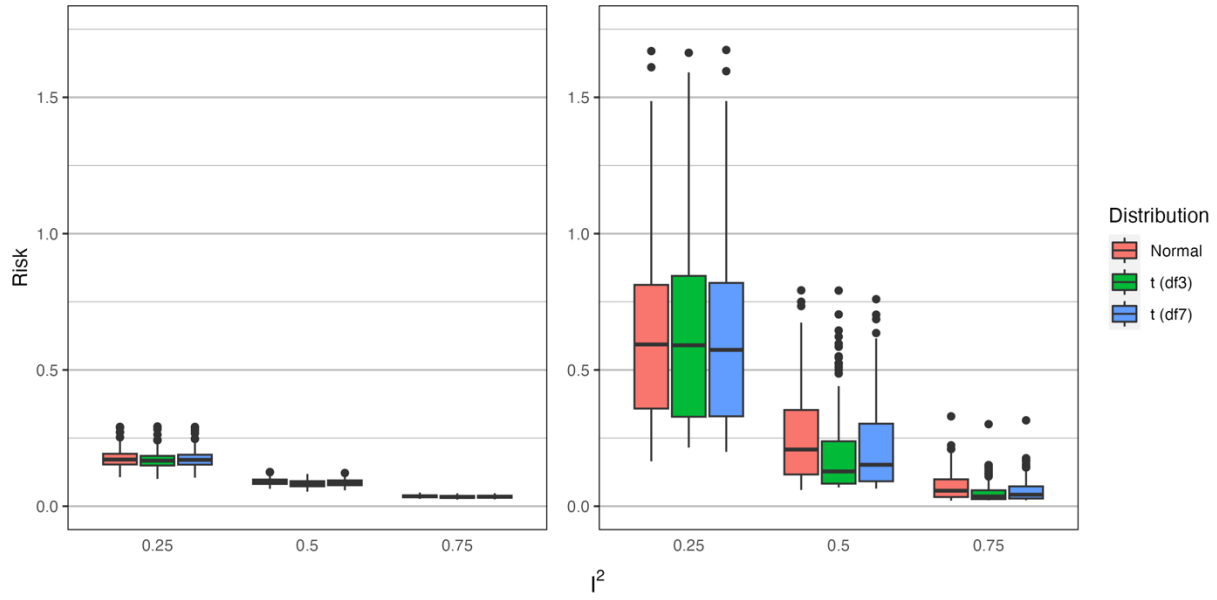

Figure 4-3 Simulation results II: average and outlier Bayes risks of the shrinkage estimate of outlying effect size, δ_{26}, from fully Bayes approaches.

The average Bayes risks are similar across the models, while the outlier Bayes risk is the smallest for the Bayes-t model with 3 degrees of freedom when $I^2 = 0.5$ and $I^2 = 0.75$.

4.4.3. Supplementary results from a conditional model

In order to examine whether the properties of the shrinkage estimates change depending on the inclusion of predictors, a supplementary simulation study is conducted using the hypothetical datasets generated based on the conditional model. Note that the simulation datasets for the conditional model are generated using the results of an empirical data analysis where the REML estimate of the remaining variance τ^2 is 0.06. When compared to the unconditional model, the unexplained variance in the between-study model is reduced by 11% after the predictor (i.e., the duration of treatment in weeks) is added to the model.

The data generation process of a conditional model involves sampling the predictor values, x_i, from a normal distribution. The details of data generation are presented as follows.

- Step 1: generate predictor values, x_i

In the first step of data generation for the conditional model, 25 predictor values are sampled from a normal distribution with a mean of 0 and standard deviation of 6.62. Note that these values are drawn from the simple average and the variance of the centered predictor variables which is the intervention length in weeks in the first set of empirical data-analysis results.

$$x_i \sim N(M_x, SD_x^2).$$

The predictor of the outlier, or the 26th effect size, is fixed at 5.38 which is drawn from the original dataset. Before centering, the intervention length of the outlying study is 15 weeks with predictor values for the other studies ranging from 1 to 24, and the average length of interventions 9.62. The reason for fixing the predictor value of the outlier, rather than sampling it

from a normal distribution as in other studies, is to avoid the situation where the extreme values are assigned to the outlier. If the outlier is located at one of the extremes of the predictor values, the estimated slope will not retain the relationship between the study outcomes and the treatment length in the data-analysis example (e.g., the REML estimate of slope: $\hat{\gamma}_{10} = 0.02$). For this reason, I use the same value of 5.38 for the intervention length of the outlier across all datasets within every condition.

- Step 2: generate true effect sizes, δ_i

For the conditional model, the datasets are generated in a similar way as the unconditional model, however, a normal distribution is specified using the conditional mean and the conditional variance that depends on the predictor value, as follows:

$$\delta_i \sim N(\gamma_{00} + \gamma_{10} * x_i, \tau^2).$$

Using the data-analysis results of the conditional model, γ_{00} is set to 0.26, γ_{10} is set to 0.02, and τ^2 is set to 0.06. For the outlier, the true effect size δ_i is set to 1.46 as in the unconditional model.

- Step 3: generate effect-size estimates, d_i's, of the 26 true effect sizes

In Step 3, study outcomes d_i's are generated based on the normal distribution with mean δ_i, generated from Step 2, and variance of s_i^2 which corresponds to a given I^2:

$$d_i \sim N(\delta_i, s_i^2).$$

Note that the outlier is generated following $d_i \sim N(1.46, s_i^2)$, as in the unconditional model.

Example datasets for the simulation of the conditional model are presented in Figure 4-4.

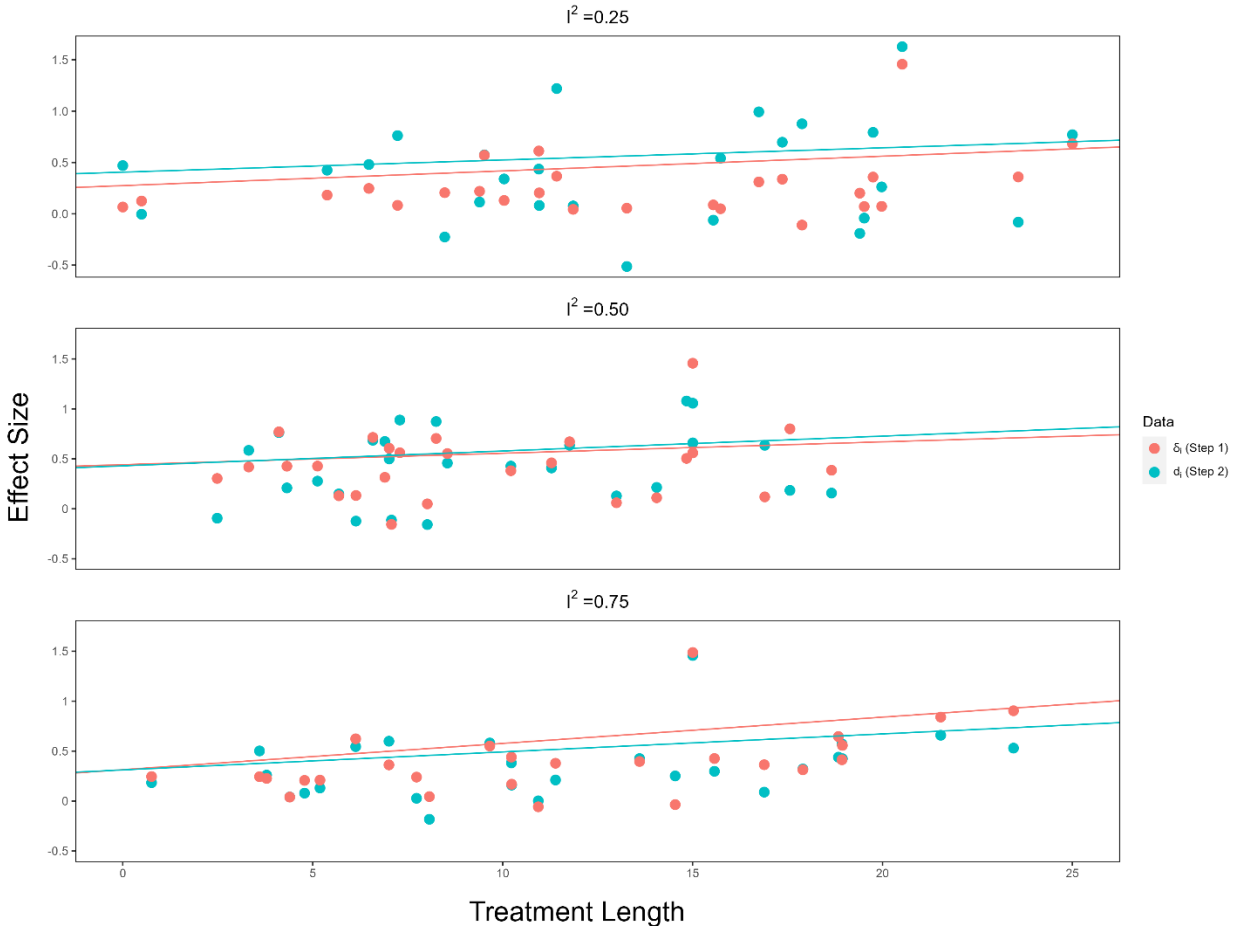

Figure 4-4 Example datasets for the targeted simulation study (Conditional model)

Red data points represent true effect sizes, δ_i, while blue data points are study outcomes, d_i, generated based on δ_i from Step 1. A red line in each plot is the REML estimate of the slope based on true effect sizes, δ_i's, while a blue line indicates the REML estimate of the regression line based on the effect-size estimates, d_i's. As I^2 increases, the degree that the study outcomes deviate from their true effect sizes decreases. For example, the δ_i's and the corresponding d_i's overlap with each other when I^2 is 0.75 and the error variance is the smallest. Note that the true effect size of the outlier is fixed at 1.46 and the predictor value of the outlier is fixed at the treatment length of 15 weeks (5.38 after centering) for all conditions.

As presented in Table 4-3, the simulation results of the conditional model demonstrate a similar pattern to the unconditional model results. In terms of the shrinkage estimate of the outlier, the Bayes-t model with 3 df yields the best results for most of the conditions in terms of the smallest bias and MSE, as well as the coverage closest to the nominal value of 95%. The average bias, MSE and coverage of all shrinkage estimates are also similar across the different estimation approaches, as was the case with the unconditional model results. The large values of average relative bias are due to a few true effect sizes which are generated close to 0, as described in the unconditional model results. The similar results for the unconditional and the conditional models suggests that the Bayes-t approach will yield good shrinkage estimates in meta-regression models given the same I^2.

Table 4-3 Simulation results of conditional model I: Simple Bias, Relative Bias (%), Mean Squared Error, and Coverage

	I^2	EB		FB-Normal		FB-t: df=3		FB-t: df=7	
		Outlier	Average	Outlier	Average	Outlier	Average	Outlier	Average
Simple Bias	0.25	-0.66	0.00	-0.65	0.00	-0.63	0.00	-0.64	0.00
	0.50	-0.40	0.00	-0.38	0.00	-0.24	0.00	-0.30	0.00
	0.75	-0.17	0.00	-0.16	0.00	-0.07	0.00	-0.11	0.00
Relative Bias	0.25	-45.63	101.78	-44.45	110.82	-43.30	75.85	-44.01	95.04
	0.50	-27.11	-91.35	-25.87	-94.68	-16.42	-97.85	-20.78	-94.57
	0.75	-11.90	-36.48	-11.02	-33.07	-4.84	-50.47	-7.18	-45.72
Mean Squared Error	0.25	0.52	0.08	0.50	0.08	0.54	0.08	0.51	0.08
	0.50	0.20	0.04	0.19	0.04	0.14	0.04	0.16	0.04
	0.75	0.05	0.02	0.04	0.02	0.02	0.02	0.03	0.02
Coverage	0.25	0.31	0.89	0.56	0.95	0.71	0.95	0.64	0.96
	0.50	0.48	0.94	0.68	0.95	0.83	0.95	0.76	0.95
	0.75	0.70	0.95	0.80	0.95	0.94	0.95	0.91	0.96

In terms of the Bayes risks presented in Table 4-4, the Bayes-t model with 3 df outperforms the other approaches in terms of the average and outlier risks except the condition $I^2 = 0.5$, where the Bayes-normal model has the smallest outlier risk. When the Bayes-t model with 3 df and the Bayes-normal model are compared, the proportional reduction is 17% when $I^2 = 0.5$ and 33% $I^2 = 0.75$. Overall, the simulation results are very similar across unconditional and conditional models. This is because the variance ratio strongly influences the amount of shrinkage, regardless of the inclusion of the predictor. In other words, as long as the variance ratio remains the same, the properties of the shrinkage estimate of the outlier will be also very similar.

Table 4-4 Simulation results of conditional model II: Average and Outlier Bayes Risks

	Normal		t: df=3		t: df=7	
I^2	Average	Outlier	Average	Outlier	Average	Outlier
0.25	0.16	0.52	0.16	0.53	0.16	0.52
0.50	0.08	0.23	0.07	0.19	0.08	0.20
0.75	0.03	0.06	0.03	0.04	0.03	0.05

CHAPTER 5

Discussion

The previous chapters have illustrated the over-shrinkage of outliers which often occurs under Empirical Bayes (EB) estimation and demonstrated that a fully Bayesian (FB) meta-analysis model assuming t-distributed random effects provide robust shrinkage estimates of outliers. In the last chapter, I will first summarize the primary findings from the empirical data-analysis and simulation study and then discuss the implications of study results. Further, I will also discuss the limitations of this study, which might guide the directions of future research.

5.1. Summary of the findings

This study introduces a fully Bayesian (FB) approach for meta-analysis that provides robust shrinkage estimates of outlying effect sizes in small-sample settings. Shrinkage estimation, in particular Empirical Bayes (EB) estimation, has been widely used in meta-analysis to obtain improved estimates of true effect sizes (Raudenbush, 2009). However, even though shrinkage estimation decreases the overall bias of ensembles of effect-size estimates, there is no guarantee that each individual estimate is close to its true effect size (Efron & Morris, 1971, 1972). In standard meta-analysis models based on shrinkage estimation, the effect-size estimates that are far from the rest of the studies may be shrunk to the overall mean by a substantial amount. The over-shrinkage of outlying study results is particularly problematic when

researchers are attempting to answer substantive questions concerning how large the largest true effect size might be among the studies in one's sample. To alleviate the over-shrinkage problem of outliers, this study employs a heavy-tailed distribution for random effects within FB framework. The primary findings are summarized as follows.

First, this study demonstrates that employing t-distributional assumptions for true effect sizes in FB models, rather than normality assumptions, yields improved shrinkage estimates for outliers. Specifically, the empirical data-analysis results using the meta-analysis dataset from Bangert-Drowns et al. (2004) show that the amount of shrinkage decreases substantially under Bayes-t models, compared to Bayes-normal models, and to the EB approach. The data-analysis results are in line with the early findings discussed in West (1984) where specifying a t-distribution for random effects leads to the least amount of shrinkage of outliers.

Second, the results from a targeted simulation study are also consistent with the data-analysis results, supporting the advantage of Bayes-t models in avoiding over-shrinkage. The simulation study results further reveal that the Bayes-t model with 3 degrees of freedom is more advantageous when the variance ratio I^2 is 0.5 and 0.75. These conditions correspond to the settings where the parameter variance is similar to, or larger than the amount of sampling variance. For the condition $I^2 = 0.25$, the shrinkage estimates of outliers are similar across the estimation methods. Given these results, the Bayes-t model with small degrees of freedom will cover most of the situations where the over-shrinkage of outliers can be problematic.

In addition, the supplementary simulation results show the good properties of shrinkage estimates of outliers under the Bayes-t model with 3 df, specifically the least amount of bias and MSE, also hold for meta-regression models with predictors. This is because the amount of

shrinkage changes depending on the variance ratio, and will not be influenced by the factors that are not relevant to parameter variance or sampling variance. It further suggests that the results under the different conditions will be similar to the unconditional model results as long as the variance ratio is the same. For example, suppose that a meta-analysis model based on another dataset with both a large positive outlier and a large negative outlier yields λ_i's that are around 0.5 for the outliers. Though the number and location of the outliers are different from this study, the difference in the amount of shrinkage between the Bayes-t model with 3 df, and the Bayes-normal model, will be the largest as presented in the analysis results when $I^2 = 0.5$ in Chapter 3.

Lastly, the simulation results also show that the outlier Bayes risks are the smallest under the Bayes-t model with 3 df, while the average Bayes risks are similar across the EB, Bayes-normal model, and Bayes-t models with 3 df and 7 df. This can be attributed to the fact that the estimates of τ^2/q_i under the Bayes-t models will be fairly close to the posterior mode of τ^2 from the Bayes-normal model for the cases which are non-outliers. In the case of non-outliers the relatively similar estimates of τ^2/q_i and τ^2 will yield shrinkage estimates that are close to one another, resulting in the average risks that are similar across the FB models. The simulation results on the Bayes risks suggest that the Bayes-t model with small degrees of freedom is capable of providing robust shrinkage estimates of outliers without increasing the average risk (c.f., Efron & Morris, 1972).

5.2. Implications of the study

This study shows that the FB approach based on t-distributed random effects for meta-analysis yields shrinkage estimates of outliers that are more robust to an excessive amount of shrinkage toward the average effect size, compared to the conventional methods. The over-

shrinkage problem connected to EB estimation and the possibility that employing heavy-tailed distributional assumptions might remedy the issue have been discussed in the literature (e.g., (Dempster, 1983; Seltzer et al., 1996; West, 1984). Motivated by these earlier works, this study demonstrates that the shrinkage estimates from the Bayes-t model have good properties, such as less bias, smaller MSE, and actual coverage closer to the nominal coverage of 95% compared to results for shrinkage estimates based on normality assumptions. Further, this study also reveals situations in which the Bayes-t approach is most needed, i.e., settings in which the parameter variance of the effect sizes is similar to or larger than the error variance, i.e. $I^2 = 0.5$.

The studies that report large positive treatment effects usually attract a lot of attention as they can provide insights regarding how successful a certain intervention might be (Rubin, 1981). While the standard approaches such as EB and FB models based on normality assumptions often fail to provide reliable answers concerning extreme observations, the FB approach employing a heavy-tailed assumption will provide robust estimates for outliers. This will enable a careful estimation and examination of outliers, that will be beneficial with respect to future trials that seek to improve the effects of educational programs.

Over-shrinkage might be acceptable if one has good evidence that a given outlier belongs to the same sub-population of effect sizes with the other cases in the sample, however, the true distribution of effect sizes is unknown in most cases. Further, the information to figure out whether true effect sizes of extreme study results follow the same distribution with other studies in the sample or whether outliers belong to different subpopulations is usually unavailable. When ambiguity around the true effect-size distribution exists, an FB approach assuming heavy-tailed random effects can provide robust shrinkage estimates as well as 95% Credible Intervals, which provide a sensitivity analysis with respect to estimates of true effect sizes from the conventional

models. Given the increasing interests in obtaining improved shrinkage estimates from multilevel models (Aert et al., 2021; Haaf & Rouder, 2023; Jordan, 2022; Longford, 2010; Wang & Lee, 2019) , this study provides a promising approach to obtain reliable estimates for outliers.

Although this study primarily focuses on meta-analysis, the Bayes-t approach expands to a wide range of applications such as multisite trials, replication studies, and cross-national comparisons. When the sample size is small, the Bayes-t model will yield robust fixed-effect estimates by accounting for the uncertainty of the variance estimate (e.g., τ^2) as well as by reducing the influence of outliers on parameter estimates. In this sense, employing heavy-tailed distributional assumptions for random effects would be helpful in drawing sound conclusions from a limited number of cases (e.g., a limited number of countries in a cross-national study, or a small number of sites in a multisite trial).

5.3. Limitations and future research directions

This study has demonstrated the advantage of using FB meta-analysis models based on heavy-tailed distributional assumptions to guard against the over-shrinkage of outliers. However, there are several limitations to the findings, which could be addressed in future research.

First, this study only considered the situation where the precisions of effect-size estimates are fairly similar to each other, both for the empirical data-analysis and the targeted simulation study. However, in meta-analysis or in other applications such as multisite studies, the sampling errors of outliers might vary substantially from the rest of the cases. If the error variance of an outlying effect-size estimate is extremely smaller or larger than the other studies, the I^2 statistic

will not be equal to the λ_i's (i.e., $\lambda_i = \frac{\tau^2}{\tau^2 + s_i^2}$) for outliers. This is because the I^2 statistic is based on the expected average of sampling variances when the precision of each study differs (Higgins & Thompson, 2002). For example, even if the I^2 is around 0.5, the λ_i of an outlier under the Bayes-normal model can be much larger , e.g. $\lambda_i = 0.9$, than I^2, when the error variance is smaller than the parameter variance as well as than those based on other studies. Thus, the shrinkage estimates from the Bayes-normal model or the Bayes-t model with small degrees of freedom will be similar with each other, because the λ_i's will be close to 1 in both of the models. Even in such situations, however, employing t-distributional assumption for random effects reduces the impact of outliers on the estimates of fixed effects and variance components (Seltzer, 1993). Given that the parameter estimates from the EB or the Bayes-normal models are susceptible to the influence of outliers with high precisions, the Bayes-t model with small degrees of freedom will provide more sensible results than the standard approaches.

Consider another extreme case where the λ_i's of outliers are smaller than the I^2 (e.g., $\lambda_i = 0.1$ and $I^2 = 0.5$). This occurs when the error variances of outliers are much larger compared to the parameter variance and those of the rest of the data. As shown in the simulation results for the condition $I^2 = 0.25$, the corresponding shrinkage estimates will be similar across the Bayes-normal model and the Bayes-t model with 3 df when λ_i is small. In such situations, one might want to examine the outlying studies carefully to figure out the possible reasons for their low precisions, e.g., small sample-sizes or imbalance in sample sizes across treatment conditions, the presence of outlying participants within studies, etc. If there is evidence that the low precision of the outlier is connected with problems in study designs or treatment implementations, one might consider excluding the outlier from the analysis.

Second, the primary findings of this study are based on the prior specification of uniform, non-informative distributions on all the unknown parameters in the model, such as the average effect size μ and the parameter variance τ^2. However, in practice, one often encounters datasets smaller than the current example which contains 26 study outcomes. With an even smaller number of cases, e.g., an international assessment dataset with less than 10 countries, the parameter estimates from FB hierarchical models are sensitive to the choice of prior distributions for the variance component τ^2. In such situations, alternative prior distributions, other than uniform distributions, should be examined to see if the conclusions based on the different prior distributions agree well with one another.

One of the promising approaches in small-sample settings is to place a weakly-informative prior on τ^2 (Seltzer et al., 1996; Thompson & Becker, 2020b). When meta-analytic datasets only have a small number of studies, the marginal posterior distribution of τ^2 is likely to be right-skewed, expanding to a very large value. The large values for τ^2's can result in wide intervals of fixed-effect estimates. This can be alleviated by specifying a weakly-informative prior for τ^2, with its mode set to the REML estimate of τ^2. Though not addressed in the current study, it would be worthwhile to examine if an alternative prior distribution for τ^2 like the data-driven prior can offer improved precisions for the shrinkage estimates of outliers.

Lastly, the amount of shrinkage is determined only by the variance ratio under the approaches considered in this study. However, there is another line of alternative methods that are more flexible in terms of the degree of shrinkage. For example, the limited translation rules proposed by Efron and Morris (1971, 1972) enable one to set a limit on how much an effect size, for example, can be shrunk toward a grand mean, depending on the research contexts and the

assumptions related to the data. In addition, one can assume a mixture distribution for the random effects to model the outlying studies when the number of studies is small (Beath, 2014); or replace the assumption of normally distributed random effects with long-tailed distributions by adding extra parameters to accommodate extreme observations (Baker & Jackson, 2008, 2016). On the other hand, semi-parametric and non-parametric methods are also available for meta-analytic datasets, e.g., based on the assumption that the random effects follow multiple underlying distributions, which enables more flexible modeling of extreme study results (Kleinman & Ibrahim, 1998). Future research should address whether these robust approaches for random effects models are capable of providing sensible shrinkage estimates of outliers.

Bibliography

Aert, R. C. M., Schmid, C. H., Svensson, D., & Jackson, D. (2021). Study specific prediction intervals for random-effects meta-analysis: A tutorial: Prediction intervals in meta-analysis. *Research Synthesis Methods*, *12*(4), 429–447. https://doi.org/10.1002/jrsm.1490

Baker, R., & Jackson, D. (2008). A new approach to outliers in meta-analysis. *Health Care Management Science*, *11*(2), 121–131. https://doi.org/10.1007/s10729-007-9041-8

Baker, R., & Jackson, D. (2016). New models for describing outliers in meta-analysis. *Research Synthesis Methods*, *7*(3), 314–328. https://doi.org/10.1002/jrsm.1191

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The Effects of School-Based Writing-to-Learn Interventions on Academic Achievement: A Meta-Analysis. *Review of Educational Research*, *74*(1), 30.

Beath, K. J. (2014). A finite mixture method for outlier detection and robustness in meta-analysis: A finite mixture method for robust meta-analysis. *Research Synthesis Methods*, *5*(4), 285–293. https://doi.org/10.1002/jrsm.1114

Box, G. E. P., & Tiao, G. C. (1992). *Bayesian Inference in Statistical Analysis: Box/Bayesian*. John Wiley & Sons, Inc. https://doi.org/10.1002/9781118033197

Browne, W. J., & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, *1*(3). https://doi.org/10.1214/06-BA117

Bryan, M. L., & Jenkins, S. P. (2016). Multilevel Modelling of Country Effects: A Cautionary Tale. *European Sociological Review*, *32*(1), 3–22. https://doi.org/10.1093/esr/jcv059

Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine, 25*(24), 4279–4292. https://doi.org/10.1002/sim.2673

Choi, K., & Seltzer, M. (2010). Modeling Heterogeneity in Relationships Between Initial Status and Rates of Change: Treating Latent Variable Regression Coefficients as Random Coefficients in a Three-Level Hierarchical Model. *Journal of Educational and Behavioral Statistics*, *35*(1), 54–91. https://doi.org/10.3102/1076998609337138

Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences* (0 ed.). Routledge. https://doi.org/10.4324/9780203771587

Dempster, A. P. (1983). Comment. *Journal of the American Statistical Association*, *78*(381), 57–58. https://doi.org/10.1080/01621459.1983.10477922

Depaoli, S., Clifton, J. P., & Cobb, P. R. (2016). Just Another Gibbs Sampler (JAGS): Flexible Software for MCMC Implementation. *Journal of Educational and Behavioral Statistics*, *41*(6), 628–649. https://doi.org/10.3102/1076998616664876

Efron, B., & Morris, C. (1971). Limiting the Risk of Bayes and Empirical Bayes Estimators—Part I: The Bayes Case. *Journal of the American Statistical Association*, *66*(336), 807–815. https://doi.org/10.1080/01621459.1971.10482348

Efron, B., & Morris, C. (1972). Limiting the Risk of Bayes and Empirical Bayes Estimators—Part II: The Empirical Bayes Case. *Journal of the American Statistical Association*, *67*(337), 130–139. https://doi.org/10.1080/01621459.1972.10481215

Gelfand, A. E., Hills, S. E., Racine-Poon, A., & Smith, A. F. M. (1990). Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling. *Journal of the American Statistical Association*, *85*(412), 972–985. https://doi.org/10.1080/01621459.1990.10474968

Gelman, A., Carlin, J. B., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (3rd ed.). Chapman and Hall/CRC.

Gilks, W. R., Richardson, Sylvia, & Spiegelhalter, D. (1995). *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC.

Haaf, J. M., & Rouder, J. N. (2023). Does every study? Implementing ordinal constraint in meta-analysis. *Psychological Methods*, *28*(2), 472–487. https://doi.org/10.1037/met0000428

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Jordan, P. (2022). On Reverse Shrinkage Effects and Shrinkage Overshoot. *Psychometrika*. https://doi.org/10.1007/s11336-022-09872-8

Kleinman, K. P., & Ibrahim, J. G. (1998). A Semiparametric Bayesian Approach to the Random Effects Model. *Biometrics*, *54*(3), 921. https://doi.org/10.2307/2533846

Lange, K. L., Little, R. J. A., & Taylor, J. M. G. (1989). Robust Statistical Modeling Using the t Distribution. *Journal of the American Statistical Association*, *84*(408), 881–896.

Larose, D. T., & Dey, D. K. (1997). Grouped random effects models for Bayesian meta-analysis. *Statistics in Medicine*, *16*(16), 1817–1829. https://doi.org/10.1002/(SICI)1097-0258(19970830)16:16<1817::AID-SIM621>3.0.CO;2-N

Lindley, D. V., & Smith, A. F. M. (1972). Bayes Estimates for the Linear Model. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(1), 1–18. https://doi.org/10.1111/j.2517-6161.1972.tb00885.x

Longford, N. T. (2010). Estimation of the effect size in meta-analysis with few studies. *Statistics in Medicine*, *29*(4), 421–430. https://doi.org/10.1002/sim.3814

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). [No title found]. *Statistics and Computing*, *10*(4), 325–337. https://doi.org/10.1023/A:1008929526011

Plummer, M. (2003). JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, 1–10.

R Core Team. (2023). *R: A language and environment for statistical computing.* (4.2.3). R Foundation for Statistical Computing. https://www.R-project.org/

Raudenbush, S. W. (2009). Analyzing effect sizes: Random-effects models. In *The Handbook of Research Synthesis and Meta-Analysis* (2nd ed., pp. 295–315). Russell Sage Foundation.

Raudenbush, S. W., & Bryk, A. S. (2001). *Hierarchical Linear Models: Applications and Data Analysis Methods (Advanced Quantitative Techniques in the Social Sciences)* (2nd ed.). SAGE Publications, Inc.

Raudenbush, S. W., & Congdon, R. T. (2021). *HLM 8: Hierarchical linear and nonlinear modeling*. Chapel Hill, NC: Scientific Software International, Inc.

Rubin, D. B. (1981). Estimation in Parallel Randomized Experiments. *Journal of Educational Statistics*, *6*(4), 377–401. https://doi.org/10.3102/10769986006004377

Seltzer, M. H. (1993). Sensitivity Analysis for Fixed Effects in the Hierarchical Model: A Gibbs Sampling Approach. *Journal of Educational Statistics*, *18*(3), 207–235. https://doi.org/10.3102/10769986018003207

Seltzer, M. H., Novak, J., Choi, K., & Lim, N. (2002). Sensitivity Analysis for Hierarchical Models Employing *t* Level-1 Assumptions. *Journal of Educational and Behavioral Statistics*, *27*(2), 181–222. https://doi.org/10.3102/10769986027002181

Seltzer, M. H., Wong, W. H., & Bryk, A. S. (1996). Bayesian Analysis in Applications of Hierarchical Models: Issues and Methods. *Journal of Educational and Behavioral Statistics*, *21*(2), 131–167. https://doi.org/10.3102/10769986021002131

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. https://doi.org/10.1002/sim.4780142408

Spiegelhalter, D. J., Abrams, K. R., & Myles, J. P. (2003). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation* (1st ed.). Wiley. https://doi.org/10.1002/0470092602

Su, Y.-S., & Yajima, M. (2021). *R2jags* (0.7-1) [R package]. https://cran.r-project.org/web/packages/R2jags/

Tanner, M. A., & Wong, W. H. (1987). The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association*, *82*(398), 528–540. https://doi.org/10.1080/01621459.1987.10478458

Thompson, C. G., & Becker, B. J. (2020a). A group-specific prior distribution for effect-size heterogeneity in meta-analysis. *Behavior Research Methods, 52*(5), 2020–2030. https://doi.org/10.3758/s13428-020-01382-8

Thompson, C. G., & Becker, B. J. (2020b). A group-specific prior distribution for effect-size heterogeneity in meta-analysis. *Behavior Research Methods*, *52*(5), 2020–2030. https://doi.org/10.3758/s13428-020-01382-8

Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*(2), 112–125. https://doi.org/10.1002/jrsm.11

Wang, C., & Lee, W. (2019). A simple method to estimate prediction intervals and predictive distributions: Summarizing meta-analyses beyond means and confidence intervals. *Research Synthesis Methods*, jrsm.1345. https://doi.org/10.1002/jrsm.1345

West, M. (1984). Outlier Models and Prior Distributions in Bayesian Linear Regression. *Journal of the Royal Statistical Society: Series B (Methodological)*, *46*(3), 431–439. https://doi.org/10.1111/j.2517-6161.1984.tb01317.x

Willey, L. H. (1988). *The effects of selected writing to learn approaches on high school students' attitudes and achievement.* Mississippi State University.