Bayesian Hierarchical Modeling based on Multi-Source Exchangeability

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Alexander Mark Kaizer

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Advised by Joseph S. Koopmeiners, Ph.D.

July, 2017

Acknowledgements

I have been extremely fortunate to have worked with so many professionals and mentors who have inspired me and helped me during my academic development. I am eternally grateful to my thesis adviser, Joe Koopmeiners, who has served as an excellent role model, provided a seemingly never ending supply of patience as I learned how to be a biostatistician, and was always willing to help me connect the dots when things did not seem to make sense. I am also extremely appreciative of all the support and insights from Brian Hobbs which helped to make this dissertation possible from weekly conference calls in Joe's office to extensive email chains. Additionally, my love for collaborative work with medical professionals would not have been possible without the guidance and encouragement of Kyle Rudser.

I am very grateful to my committee members, Brad Carlin, Haitao Chu, and Julia Drew for taking the time to review my thesis and for asking questions which challenged me, but ultimately led to a stronger dissertation.

Dedication

This thesis is dedicated to my family who have supported me from both near and far, and to my housemates who have become a second family here in Minnesota.

Abstract

Progress in medical practice traditionally takes place over a sequence of clinical studies which are designed to establish clinical efficacy, identify the safety profile, and seek regulatory approval for a novel treatment strategy. These trials in humans can be expensive and present numerous challenges in their implementation. While some challenges may be addressed by the development of innovative trial designs, it may also be advantageous to incorporate supplemental sources of information, which are typically ignored in traditional approaches to analysis. In this dissertation, we introduce Multi-Source Exchangeability Models (MEMs), a general Bayesian hierarchical approach that integrates supplemental data arising from multiple, possibly non-exchangeable, sources into the analysis of a primary source. We first describe the proposed framework and prove some desirable asymptotic properties that show the consistency of posterior estimation. Simulation results illustrate that MEMs incorporate more supplemental information in the presence of homogeneous supplemental sources and exhibit reduced bias in the presence of heterogeneous supplemental sources relative to competing Bayesian hierarchical modeling strategies. Next, we illustrate how MEMs can be used to design a more efficient sequential platform design for Ebola virus disease by sharing information across trial segments. When compared to the standard platform design, we demonstrate that MEMs with adaptive randomization improved power by as much as 51% with limited type-I error inflation. We conclude by extending our work with model averaging to the estimation of multiple mixture distributions in the presence of a hypothesized biological relationship between groups to identify non-compliance in a regulatory tobacco clinical trial. The results of this dissertation illustrate that MEMs yield favorable characteristics across a variety of scenarios and motivates further research to extend the MEM framework to other settings, as well.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Incorporating supplemental sources of data into a primary source

Advances in medical practice arise from evaluating therapeutic interventions over a sequence of clinical studies devised to establish the clinical efficacy and safety profile of a novel treatment strategy. Conduct of clinical trials in humans is expensive and inherently challenging. Furthermore, the current paradigm for evaluating novel interventions, whereby therapies are screened one-at-a-time in phases, remains inefficient with each investigational drug requiring a sequence of disjointed trials designed to acquire and evaluate different types of information. Moreover, design, review, and initiation of a single study, a period often referred to as operational "whitespace," is a gradual process. After initiation, a successful trial usually requires several years to achieve the targeted enrollment. Additionally, a considerable proportion of studies fail due to low recruitment (Williams et al., 2015). This system for clinical testing produces redundancy, whereby similar treatment strategies are replicated, either as experimental or standard-of-care therapy, across multiple studies and development phases.

Some of these limitations to traditional clinical trial progression may be addressed by the development of new designs for clinical trials, such as sequential platform trials, which incorporate and test multiple treatments sequentially within a single master protocol. However, these newer designs present some potential redundancies as well. Within

a sequential platform trial there may be supplemental sources of information from previously completed segments of the trial which are ignored in favor of the traditional approach to analysis which only utilizes contemporaneously collected information. This represents a potential inefficiency, as data from previous segments may provide useful information about the current segment, as well.

Conventional approaches to statistical inference, which assume exchangeable data sampling models, are inappropriate for integrating data from disparate studies because of their failure to account for between-study heterogeneity, which yields statistical estimators that are sensitive to inter-cohort bias. Thus, while ignoring relevant, supplemental information reduces the efficiency of any study, supplemental data acquired from broadly similar therapeutic interventions, patient cohorts, previous investigations, or biological processes is often excluded from statistical analysis, in practice (U.S. Food and Drug Administration, 2001).

1.1.1 Previous approaches to incorporating supplemental sources of information

Statistical methods for integrating information from commensurate trials that relax the assumption of inter-cohort data exchangeability and leverage inter-trial redundancy have been developed. Pocock (1976) was first to propose using Bayesian models to incorporate supplemental information into the analysis of a primary data source through static, data-independent shrinkage estimators that require the extent of between-source variability to be pre-specified. Numerous models have been discussed since, which involve pre-specification of the amount of borrowing under different paradigms related to the power prior (Ibrahim and Chen, 2000; Hobbs et al., 2011; De Santis, 2006; Rietbergen et al., 2011) or inflating the standard error to down-weight supplemental cohorts (Goodman and Sladky, 2005; French et al., 2012; Whitehead et al., 2008).

Hierarchical linear models and models which include adaptive down-weighting of data from supplemental cohorts have been extensively explored, as well. For these models, the extent of shrinkage towards the supplemental sources is not predetermined but is estimated from the data. More strength is borrowed in the absence of evidence for inter-trial effects,

which controls the extent of bias induced from using the supplemental information. One approach is the power prior of Ibrahim and Chen (2000) which can be constructed to discount supplemental sources relative to the primary data. Bayesian (Smith et al., 1995) and frequentist (Doi et al., 2011) methods which utilize hierarchical modeling have been developed to estimate between-source variability with univariate observables or repeated measures. Other authors have considered hyperprior specifications for Bayesian hierarchical models (Daniels, 1999; Natarajan and Kass, 2000; Spiegelhalter, 2001; Gelman, 2006; Browne and Draper, 2006; Kass and Natarajan, 2006). Recently, approaches using Bayesian hierarchical modeling to leverage supplemental controls in data analysis (Neuenschwander et al., 2010; Pennello and Thompson, 2008; Chen et al., 2011; Neelon and O'Malley, 2010) and trial design (Hobbs et al., 2013) have been explored. Furthermore, dynamic approaches to incorporating supplemental information using hierarchical modeling with sparsity inducing spike-and-slab hyperpriors and empirical Bayesian inference have been described (Hobbs et al., 2011, 2012; Murray et al., 2014).

1.2 Motivating examples and plan of dissertation

1.2.1 A regulatory tobacco clinical trial

The Center for the Evaluation of Nicotine in Cigarettes, project 1 (CENIC-p1), was a 6-week randomized multi-center trial designed to evaluate the effect of nicotine reduction on tobacco use and dependence (Donny et al., 2015). 839 current smokers underwent randomization, with 780 completing the 6-week study after being equally randomized to seven treatment groups, including a usual brand control condition or one of six experimental cigarette conditions with nicotine content ranging from 15.8 mg per gram of tobacco (15.8 mg/g group; approximately equivalent to the nicotine content of commercial cigarettes) to 0.4 mg per gram of tobacco (0.4 mg/g group; also referred to as very low nicotine content (VLNC) cigarettes). At the end of six weeks, participants randomly assigned to the lowest nicotine condition had significantly reduced tobacco use, dependence, and nicotine exposure relative to the normal nicotine controls.

In Chapter 2, the data from CENIC-p1 are used to illustrate the benefit of using MEMs to borrow information from a supplemental data source compared to standard analytical

approaches. For purposes of illustration, our analysis focuses on the comparison of the change in cigarettes smoked per day from baseline to week 6 between the 15.8mg/g group and the 0.4 mg/g group. There are multiple supplemental sources of data that could be combined with our primary data source (the 15.8 mg/g and 0.4 mg/g groups from CENIC-p1) to obtain a more precise estimate of the effect of nicotine reduction on cigarette consumption. Specifically, data are available from two previous trials of VLNC cigarettes (Hatsukami et al., 2013, 2010). In addition, CENIC-p1 also included a 0.4 mg/g group with a higher-than-normal tar yield, which could be viewed as supplemental data for the 0.4 mg/g group. Finally, the 15.8 mg/g group could be combined with the usual brand control group from CENIC-p1 to achieve a more precise estimate of cigarette consumption in the control condition. However, assuming that the primary and supplemental sources are exchangeable may be inappropriate for a number of reasons and we desire a statistical method that is flexible enough to down-weight individual supplemental sources that are not consistent with the primary data source, while achieving increased precision when the supplemental sources are consistent with the primary source.

In Chapter 4, model averaging is used to develop a mixture modeling approach for identifying non-compliance in CENIC-p1. Participants self-reported high rates of non-compliance to study product (i.e., smoking cigarettes other than those provided by the study) and, while the primary analysis was completed following the intent-to-treat principle, investigators are interested in estimating the effect of nicotine reduction if all subjects had complied to the treatment to understand nicotine reduction as a potential tobacco regulatory strategy. In addition to high rates of self-reported non-compliance, it has also been noted that self-reported compliance to study product is not a reliable measure of compliance (Nardone et al., 2016) an biomarkers of nicotine exposure, such as cotinine or total nicotine equivalents, have been proposed as an approach for identifying non-compliance in randomized trials of VLNC cigarettes (Benowitz et al., 2015). However, while data from an auxiliary study provide information about the distribution of biomarkers for the VLNC group (Denlinger et al., 2016), no data are available for the intermediate dose levels, which limits the utility of biomarkers for identifying non-compliance in intermediate dose levels. To address this, we develop an approach to identify non-compliance in intermediate dose levels, which utilizes model averaging to potentially share information across nicotine

dose levels based on our biological understanding of the relationship between the nicotine content of the cigarettes and biomarkers of nicotine exposure.

1.2.2 An adaptive platform trial design for emerging infectious disease epidemics

The outbreak of the highly infectious Ebola virus disease (EVD) in West Africa beginning in March 2014 through 2016 resulted in more cases and deaths from EVD than all previous outbreaks combined (WHO, 2016). The outbreak had initial case mortality estimates as high as 74% in some areas and represented a dire situation with no known effective therapeutics (Schieffelin et al., 2014). The West African outbreak called for novel platform-based trial designs which sequentially consider multiple treatments, potentially in combination, within a single trial in order to most effectively identify beneficial therapeutics and quickly incorporate them into the standard of care for EVD to reduce morbidity and mortality. Additional difficulties in designing a trial for the EVD outbreak were the relatively sparse knowledge of this EVD strain, the potential for disease evolution or changes in the course of the trial, the urgent need for identifying any potentially beneficial treatments or combinations of treatments as quickly as possible, and the need to maintain traditional clinical trial benchmarks as much as possible for outcomes such as activity, safety, and efficacy (Dodd et al., 2016).

In response to this need, the National Institutes of Health (NIH) launched the Partnership for Research on Ebola Vaccines in Liberia II (PREVAIL II) trial, a randomized clinical trial to evaluate medical countermeasures against EVD, in March of 2015 (Dodd et al., 2016; PREVAIL II Writing Group, for the Multi-National PREVAIL II Study Team, 2016). PREVAIL II was designed as a modified platform trial, as defined by Renfro and Sargent (2016), in order to improve the efficiency of testing multiple potentially beneficial therapeutics, accelerate clinical development, and maintain flexibility in the context of an emerging infectious disease epidemic. The trial had two unique characteristics driven by the urgent need for new treatments. First, the objective of PREVAIL II was to evaluate multiple treatments within a single, master protocol rather than multiple independent studies. Treatments would be evaluated sequentially against the optimal standard of care (oSOC) which, initially, consisted only of supportive care (intravenous fluids, hemodynamic

monitoring, etc.). As the trial progressed, any treatment which demonstrated a significant improvement over oSOC would be added to the oSOC for future comparisons. Second, PREVAIL II utilized frequent interim monitoring to allow very early termination if a new treatment exhibited sufficient statistical evidence for a decline in the mortality rate with respect to the concurrent oSOC.

1.2.3 Dissertation objectives

In this dissertation, we introduce multi-source exchangeability models (MEM), a general Bayesian approach that integrates supplemental data arising from multiple, possibly non-exchangeable, sources into the analysis of a primary source, while reducing the dimensionality of the prior space by enabling prior specification on supplemental sources and avoiding the limiting assumption of exchangeability among the supplemental sources. Our modeling framework effectuates source-specific smoothing parameters that can be estimated from the data to facilitate dynamic multi-resolution smoothing. We will demonstrate that our proposed approach yields asymptotically consistent posterior estimates, while achieving more desirable small sample properties when compared to competing Bayesian hierarchical modeling strategies.

In Chapter 2 we introduce the general framework for MEMs and apply it to the context of Gaussian-distributed data with an unknown mean and known precision. Asymptotic properties are presented showing the consistency of the posterior MEM estimates with simulation studies demonstrating that desirable small sample properties can be obtained through carefully selected prior distributions. When compared to competing Bayesian hierarchical modeling strategies, the simulation results demonstrate that MEMs achieve up to a 56% reduction in bias when there is heterogeneity among the supplemental sources. When applied to the data from CENIC-p1, MEMs resulted in a 30% improvement in efficiency compared to a standard analysis without borrowing.

In Chapter 3 we propose a multi-source adaptive platform design for EVD using the MEM framework to facilitate the borrowing of information across trial segments. Since the incorporation of supplemental information can induce an imbalance between the arms

within a segment, we also incorporate adaptive randomization to balance overall information within a segment, which improves power relative to a fixed randomization strategy that results in information imbalances across treatment groups. Our proposed design demonstrates improvements of up to 51% in power with limited type-1 error inflation while randomizing more participants to promising treatment arms when compared to the design used in PREVAIL II.

In Chapter 4 we focus on the estimation of mixture distributions to identify non-compliance in CENIC-p1 using a fully Bayesian approach. A model averaging approach, which incorporates our biological understanding regarding the relationship between the nicotine content and biomarkers of nicotine exposure, is proposed to borrow information across dose-levels and achieve more precise estimates of the mixture components. There exist multiple approaches to specifying the MCMC samplers, so model averaging techniques over multiple proposed specifications are explored via reversible-jump MCMC (RJMCMC). Our proposed method results in more precise estimates of the mixture components when the proposed relationship is appropriate, while minimizing bias when the relationship does not hold.

Chapter 2

Bayesian Hierarchical Modeling based on Multi-Source Exchangeability

This chapter introduces the proposed MEM framework, which enables the incorporation of supplemental sources of information into a primary data source. MEMs are developed in the context of Gaussian-distributed outcomes with unknown means and known precision, with an extension to binary outcomes in Chapter 3 and to estimation of mixture distributions in Chapter 4. The remainder of Chapter 2 proceeds as follows. We first introduce MEMs, in generality, in Section 2.1. We then discuss estimation and investigate the asymptotic properties of MEMs in the Gaussian case in Section 2.2. Simulation results comparing the small-sample properties of MEMs to existing methods are presented in Section 2.3 and we apply our proposed method to CENIC-p1 in Section 2.4. Finally, we conclude with a brief discussion in Section 2.5.

2.1 Multi-source Exchangeability Models

Consider the general case where there is a single primary cohort, P, with n observables represented by $\boldsymbol{y}_p$, and a total of H independent supplemental cohorts considered for incorporation into the analysis with n_h observables each, represented by $\boldsymbol{y}_h$, $h = 1, ..., H$.

Together these observables represent the data, D. Our primary goal is to estimate $\boldsymbol{\theta}_p$, which represents the parameters for the primary cohort. $\boldsymbol{\theta}_h$ represents the same parameters from supplemental cohort h. For our framework, exchangeability is defined between the primary cohort and a supplemental cohort h to be where $\boldsymbol{\theta}_p = \boldsymbol{\theta}_h$.

Since it is likely that the supplemental sources may not be exchangeable with the primary cohort or that the supplemental sources may be heterogeneous, themselves, let S_h denote an indicator function of whether or not supplementary source h is assumed exchangeable with the primary cohort. A MEM, noted generally as $\boldsymbol{\Omega}_k$, is defined by considering a set of source-specific indicators, $(S_1 = s_{1,k}, ..., S_H = s_{H,k})$, where $s_{h,k} \in \{0, 1\}$ with indices for source h and model k representing all $K = 2^H$ possible configurations of assumptions regarding exchangeability between the primary and supplemental sources. A conceptual diagram representing these K configurations of exchangeability is depicted by Figure 2.1 with each box representing a potential configuration of exchangeability for a MEM, $\boldsymbol{\Omega}_k$. Solid arrows indicate $s_{h,k} = 1$ and graphically demonstrate what observables are used to estimate $\boldsymbol{\theta}_p$ and therefore indicate which supplemental sources are exchangeable with the primary cohort. If there is no solid arrow connecting the observables, then $s_{h,k} = 0$. The dashed arrows with posterior model weights, ω_k, visualize that posterior inference is completed by a weighted average of each MEM posterior. To quickly identify the sources assumed exchangeable in the K MEMs, the k subscript can be represented as the supplemental sources assumed exchangeable with the primary source such that $\boldsymbol{\Omega}$ represents the MEM which assumes no supplemental information and $\boldsymbol{\Omega}_{1,2,...,\mathrm{H}}$ assumes all supplemental sources are exchangeable.

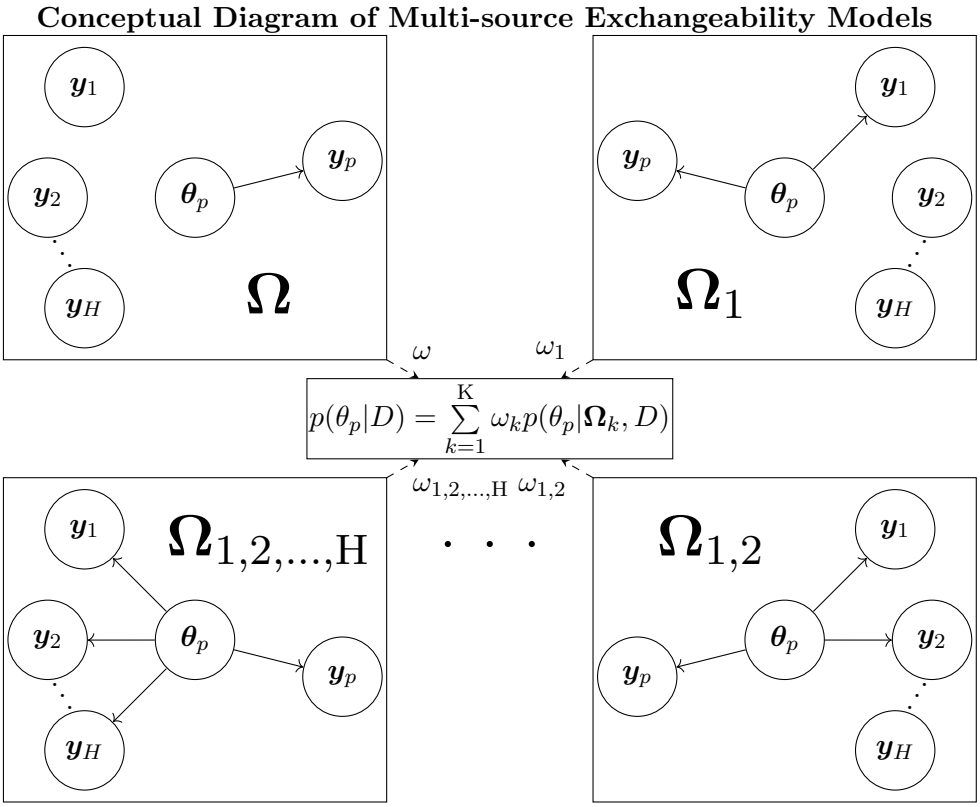

Figure 2.1: Each MEM is a combination of supplemental sources assumed exchangeable with the primary cohort in order to estimate the parameters of interest, $\boldsymbol{\theta}_p$, and is contained within each box for $\boldsymbol{\Omega}_k$. Within a box the solid arrows $\boldsymbol{\theta}_p$ and the observables, $\boldsymbol{y}_h$, represent which supplemental sources are assumed exchangeable with the primary cohort within the given MEM. The dashed arrows represent that the posterior model weights for each MEM, ω_k, are used in calculating the weighted average of each MEM's posterior distribution, $p(\theta_p|\boldsymbol{\Omega}_k, D)$, to be used for posterior inference, $p(\theta_p|D)$.

A natural approach to estimating model-specific weights is through Bayesian model averaging (BMA) (Raftery, 1995; Raftery et al., 1997; Hoeting et al., 1999). BMA accounts for uncertainty in the model specification by facilitating posterior inference that averages over a collection of posterior distributions that are obtained from a set of candidate models. BMA describes the Bayesian framework for using conditional probability to estimate a weight for each candidate model in the presence of the data. In our case, BMA would be used to average over the $K = 2^H$ multi-source exchangeability models representing all possible assumptions regarding the exchangeability of the supplementary data with the

primary data. A model weight given the data, ω_k, which is often random, is determined for each possible model as laid out by Hoeting et al. (1999) and used for posterior inference.

Let L represent the likelihood given the data for $\mathbf{\Omega}_k$, $\mathbf{\Theta} = (\theta_p, \theta_1, ..., \theta_H)$, and $\pi(\mathbf{\Theta}|\mathbf{\Omega}_k)$ denote the prior density of $\mathbf{\Theta}$ under $\mathbf{\Omega}_k$. In the context of MEMs, the integrated marginal likelihood for a particular multi-source exchangeability model given the data is obtained by averaging the likelihood over the posterior distribution for the vector of all model parameters of interest,

$$p(D|\mathbf{\Omega}_k) = \int L(\mathbf{\Theta}|D, \mathbf{\Omega}_k)\pi(\mathbf{\Theta}|\mathbf{\Omega}_k)d\mathbf{\Theta}. \tag{2.1}$$

The posterior model weights for each MEM are given by

$$\omega_k = p(\mathbf{\Omega}_k|D) = \frac{p(D|\mathbf{\Omega}_k)\pi(\mathbf{\Omega}_k)}{\sum_{j=1}^{K} p(D|\mathbf{\Omega}_j)\pi(\mathbf{\Omega}_j)}, \tag{2.2}$$

where $\pi(\mathbf{\Omega}_k)$ is the prior probability that $\mathbf{\Omega}_k$ is the true model. The marginal posterior distribution given the observable data D to be used for inference on θ_p is the weighted average using the posterior model weights of the K multi-source exchangeability model posteriors, $p(\theta_p|\mathbf{\Omega}_k, D)$:

$$p(\theta_p|D) = \sum_{k=1}^{K} \omega_k p(\theta_p|\mathbf{\Omega}_k, D). \tag{2.3}$$

Unfortunately, BMA quickly becomes highly parameterized as the number of models grows exponentially with the number of supplementary sources ($K = 2^H$), and prior specification over a model space of large size is problematic. Fernández et al. (2001) noted that posterior model weights can be very sensitive to the specification of priors in the model, especially in the absence of strong prior knowledge. In the analysis of limited data obtained from a clinical study, these issues with the conventional BMA approach become critical and motivate our proposed approach. With MEMs the supplemental sources are assumed to be distinct and independent, therefore we can specify priors with respect to sources instead of models, $\pi(\mathbf{\Omega}_k) = \pi(S_1 = s_{1,k}, ..., S_H = s_{H,k}) = \pi(S_1 = s_{1,k}) \times \cdots \times \pi(S_H = s_{H,k})$. This results in drastic dimension reduction in that it necessitates the specification of only H source-specific prior inclusion probabilities in place of 2^H prior model probabilities comprising the entire model space. In Section 2.2.2, we propose prior weights for the source-inclusion probabilities that result in consistent posterior model weights and yield desirable

small sample properties, as evaluated by simulation in Section 2.3. In contrast, similarly constructed prior weights on the models did not result in consistent posterior model weights.

2.2 Estimation and theoretical results

We now describe posterior inference using MEMs in the Gaussian case and investigate the asymptotic properties of our proposed approach for two classes of source-specific prior model weights. For our primary cohort (P), let $x_{1,p}, x_{2,p}, ..., x_{n,p}$, denote a sample of i.i.d. and normally distributed observables with mean μ and variance σ^2. Similarly, for supplemental cohorts $h = 1, ..., H$, let $x_{1,h}, ..., x_{n_h,h}$, denote a sample of i.i.d. normally distributed samples with source-specific mean μ_h and variance σ_h^2. Throughout this section, we assume σ^2 and σ_h^2 are known and define $v = \frac{\sigma^2}{n}$ and $v_h = \frac{\sigma_h^2}{n_h}$ for notational simplicity.

Using the MEM approach, the likelihood is a weighted average of multi-source exchangeability models representing all possible assumptions regarding exchangeability:

$$\sum_{k=1}^{K} \omega_k L(\boldsymbol{\mu}, \boldsymbol{\sigma}^2 | (s_{1,k}, ..., s_{H,k})) = \sum_{k=1}^{K} \omega_k \mathcal{N}(\mu, \sigma^2) \prod_{h=1}^{H} \left\{ \mathcal{N}(\mu s_{h,k} + \mu_h (1 - s_{h,k}), \sigma_h^2) \right\}. \quad (2.4)$$

Then, assuming a flat prior on the Gaussian mean, μ_k, as described by (Gelman, 2006), $\pi(\mu_k | \boldsymbol{\Omega}_k) \propto 1$ in (2.1), the Gaussian conditional marginal likelihood can be generally written for any MEM as:

$$\begin{aligned} p(D|\boldsymbol{\Omega}_k) = & \frac{(\sqrt{2\pi})^{(H+1)-\sum_{h=1}^{H}\{s_{h,k}\}}}{\sqrt{\left(\frac{1}{v} + \sum_{i=1}^{H} \left\{ \frac{s_{i,k}}{v_i} \right\}\right) \left(\prod_{j=1}^{H} \left\{ \left[\frac{1}{v_j} \right]^{1-s_{j,k}} \right\} \right)}} \times \\ & \exp\left(-\frac{1}{2} \left[\sum_{l=1}^{H} \left\{ \frac{s_{l,k}(\bar{x} - \bar{x}_l)^2}{v + v_l + v v_l (\sum_{m \neq l} \{ s_{m,k} v_m^{-1} \})} + \right. \right. \right. \\ & \left. \left. \left. \sum_{l<r}^{H} \left\{ \frac{s_{l,k} s_{r,k} (\bar{x}_l - \bar{x}_r)^2}{v_l + v_r + v_l v_r (v^{-1} + \sum_{p \neq l,r} \{ s_{p,k} v_p^{-1} \})} \right\} \right\} \right] \right). \end{aligned} \quad (2.5)$$

The exponential portion of (2.5) is comprised of the squared deviations between the sources included in $\boldsymbol{\Omega}_k$ such that if all included sources are exchangeable, then $\exp(0) = 1$ and the

posterior weights of (3.4) are influenced by the non-exponential terms of (2.5), which do not include sample means, and the priors placed on model weights, $\pi(\mathbf{\Omega}_k)$.

The posterior distribution of μ, derived from (2.5) and used for inference with multi-resolution shrinkage of the supplemental cohorts, is a mixture of normal distributions computed using (3.3) and (3.4):

$$\begin{aligned} p(\mu|D) &= \sum_{k=1}^{K} \omega_k p(\mu|D, \boldsymbol{\sigma}^2, (s_{1,k}, ..., s_{H,k})) \\ &= \sum_{k=1}^{K} \omega_k \mathcal{N}\left(\frac{\bar{x} \prod_{h=1}^{H} \left\{ v_h^{s_{h,k}} \right\} + \sum_{i=1}^{H} \left\{ \frac{v\bar{x}_i}{v_i} \right\} \prod_{j=1}^{H} \left\{ v_j^{s_{j,k}} \right\}}{v \left[\sum_{l=1}^{H} \{ \frac{s_{l,k}}{v_l} \} \prod_{m=1}^{H} \{ v_m^{s_{m,k}} \} \right] + \prod_{r=1}^{H} \left\{ v_r^{s_{r,k}} \right\}}, \left(\frac{1}{v} + \sum_{p=1}^{H} \frac{s_{p,k}}{v_p} \right)^{-1} \right) \end{aligned} \tag{2.6}$$

The posterior mean is obtained as the weighted average of the model-specific posterior means. The posterior variance of a mixture of normal distributions is also available in closed form.

2.2.1 Specification of model-specific prior weights

The properties of any dynamic Bayesian modeling approach are largely determined by its prior specification. More flexible choices attenuate the influence of *a priori* assumptions in the presence of conflicting data, imparting robustness for posterior inference. In the context of MEMs, a flexible prior specification would allocate the posterior weight to models in a manner that reflects the putative evidence for inter-source "exchangeability." As noted earlier, in the Gaussian case, (2.5) demonstrates that the exponential term of the integrated marginal likelihood for each MEM is determined by the difference between the means of the primary and supplemental data sources. If all sources in a MEM are exchangeable, the resulting marginal density is determined by the ratio of the variance components and sample size of the terms not in the exponent. In the absence of exchangeability shrinkage is reduced, as conveyed by a smaller posterior model weight.

Recall that since supplemental sources are independent we can specify the prior model weight as the product of priors on source inclusion/exclusion probabilities, $\pi(\mathbf{\Omega}_k) = \pi(S_1 = s_{1,k}) \times \cdots \times \pi(S_H = s_{H,k})$. While there are numerous strategies for determining the prior

inclusion probability of each source, we present the asymptotic properties and resulting posterior inference of two approaches. Thereafter, we will demonstrate how the prior specification impacts both asymptotic and small sample properties of the resulting posterior estimators. The first prior, denoted by π_e, equally weights all supplementary sources. This is an obvious choice for a prior since it simply provides equal "inclusion" weight to all H sources, e.g., $\pi_e(S_h = 1) = \frac{1}{2}$. This reflects the condition of impartiality as it pertains to which supplemental sources should be considered exchangeable with the primary cohort, and thus on the surface appears advantageous.

The second prior we consider, denoted π_n, is specified in relation to the sampling-level variability which attempts to overcome the likelihood's intrinsic preference for independence found with π_e. Specifying the prior source inclusion probability in relation to the fractional component of the integrated model likelihood results in:

$$\begin{aligned}
\pi_n(S_h = 1) &\propto \sum_{k=1}^{K} s_{h,k} \frac{\sqrt{\left(\frac{1}{\sigma^2} + \sum_{i=1}^{H}\left\{\frac{s_{i,k}}{v_i}\right\}\right)\left(\prod_{j=1}^{H}\left\{\left(\frac{1}{v_j}\right)^{1-s_{j,k}}\right\}\right)}}{(\sqrt{2\pi})^{(H+1)-\sum_{l=1}^{H} s_{l,k}}}, \\
\pi_n(S_h = 0) &\propto \sum_{k=1}^{K} (1 - s_{h,k}) \frac{\sqrt{\left(\frac{1}{\sigma^2} + \sum_{i=1}^{H}\left\{\frac{s_{i,k}}{v_i}\right\}\right)\left(\prod_{j=1}^{H}\left\{\left(\frac{1}{v_j}\right)^{1-s_{j,k}}\right\}\right)}}{(\sqrt{2\pi})^{(H+1)-\sum_{l=1}^{H} s_{l,k}}}.
\end{aligned} \tag{2.7}$$

π_n cancels out most of the fractional component of the integrated model likelihood (2.5). Allowing the extent of uncertainty for a given model, as represented by the variance, to influence the prior source inclusion probability yields a posterior that places greater emphasis on the exponential term of the integrated model likelihood. This enables the extent of exchangeability between the primary and each supplemental cohort to be predominately determined by its standardized mean difference, which is contained in the exponential term of $p(D|\mathbf{\Omega}_k)$. This prior also depends on the sample size of the supplemental data, which allows the source weights to place greater emphasis on the differences produced in the exponential term by reducing the influence of the fractional component. In Section 2.3, we will demonstrate that π_n produces more shrinkage than π_e which is most useful in the presence of small studies often encountered in biomedical applications.

2.2.2 Asymptotic properties

Methods that incorporate supplemental information should endeavor to integrate data from potentially very different sources and arrive at a posterior estimate that minimizes the bias introduced by incorporating the supplemental data. In the case of MEMs, bias arising from using the supplemental data can be minimized if sources which are exchangeable attain a weight of 1 while all other sources attain weight 0.

Using the two prior specifications presented in Section 2.2.1, we demonstrate the frequentist, asymptotic properties of MEMs. Specifically, we describe the conditions whereby the MEM specification yields asymptotically consistent model-specific weights, resulting in consistent estimation of μ by the posterior mean. The asymptotic properties assume a finite mixture of MEMs represented by Gaussian distributions with known variances and posterior model weights calculated in the MEM framework.

Theorem 1 *As* $n, n_1, ..., n_H \to \infty$, $\omega_{k^*} \to 1$ *for model* k^* *defined by* $(S_1 = s_{1,k^*}, \ldots, S_H = s_{H,k^*})$, *where* $s_{h,k^*} = \mathbb{1}_{\{\mu=\mu_h\}}$ *for all* $h = 1, \ldots, H$ *and* $\omega_k \to 0$ *for* $k \neq k^*$ *with priors* π_e *and* π_n.

A proof of Theorem 1 is provided in Appendix A. Theorem 1 establishes the consistency of the model-specific weights.

There are two important points to note regarding Theorem 1. First, consistency is only attained in the presence of large sample sizes for both the primary and the supplemental cohorts. This property is illustrated in Figure 2.2, which presents the model weights as a function of sample size for the case with three supplemental cohorts, where model 2 is the correct model (i.e. $\mu = \mu_1$ and $\mu \neq \mu_h$ for $h = 2, 3$). We see that the model weight converges to 1 when the sample size is large for all cohorts (b) but not when the sample size for the primary cohort is large and the sample size for the supplemental cohorts is constant (a). Second, we illustrate the advantage of specifying the priors on the source-specific inclusion weights, rather than on the model weights, in Supplemental Figures 1c and 1d. These figures present the model weights as a function of sample size for when the prior is specified on the source-specific inclusion weights using π_n and when the prior is specified on the model weights in a standard BMA approach using prior weights analogous to π_n for the case when there are three supplemental cohorts and model 5 is the correct model (i.e.

$\mu = \mu_h$ for $h = 1, 2$ and $\mu \neq \mu_3$). In this case, the model weights are consistent when the prior is specified on the source-inclusion probabilities but not when the prior is specified on the model weights. This is true even when the sample sizes for both the primary and supplemental cohorts are large, suggesting that the prior specification on the model is not adequate for all situations in our context.

Because bias arising from integrating multi-source data is avoided when shrinkage is effectuated using only truly exchangeable sources, consistency is an important property for any multi-source integration approach. The fact that the MEM estimators are consistent for both priors also demonstrates flexibility to accommodate a wide variety of prior beliefs pertaining to source inclusion. As noted in Section 2.2.1, the posterior mean, μ_{MEM}, is a weighted average of the model-specific posterior means, which, in combination with Theorem 1, allows us to conclude that the posterior mean will be consistent for the true mean assuming π_e and π_n.

Corollary 2 *As* $n, n_1, ..., n_H \to \infty$, $E(\mu_{MEM}|D) \xrightarrow{a.s} \mu$ *with priors* π_e *and* π_n.

The proof of Corollary 2 is a result of the consistency of the sample mean by the strong law of large numbers, Theorem 1 and Slutsky's Theorem.

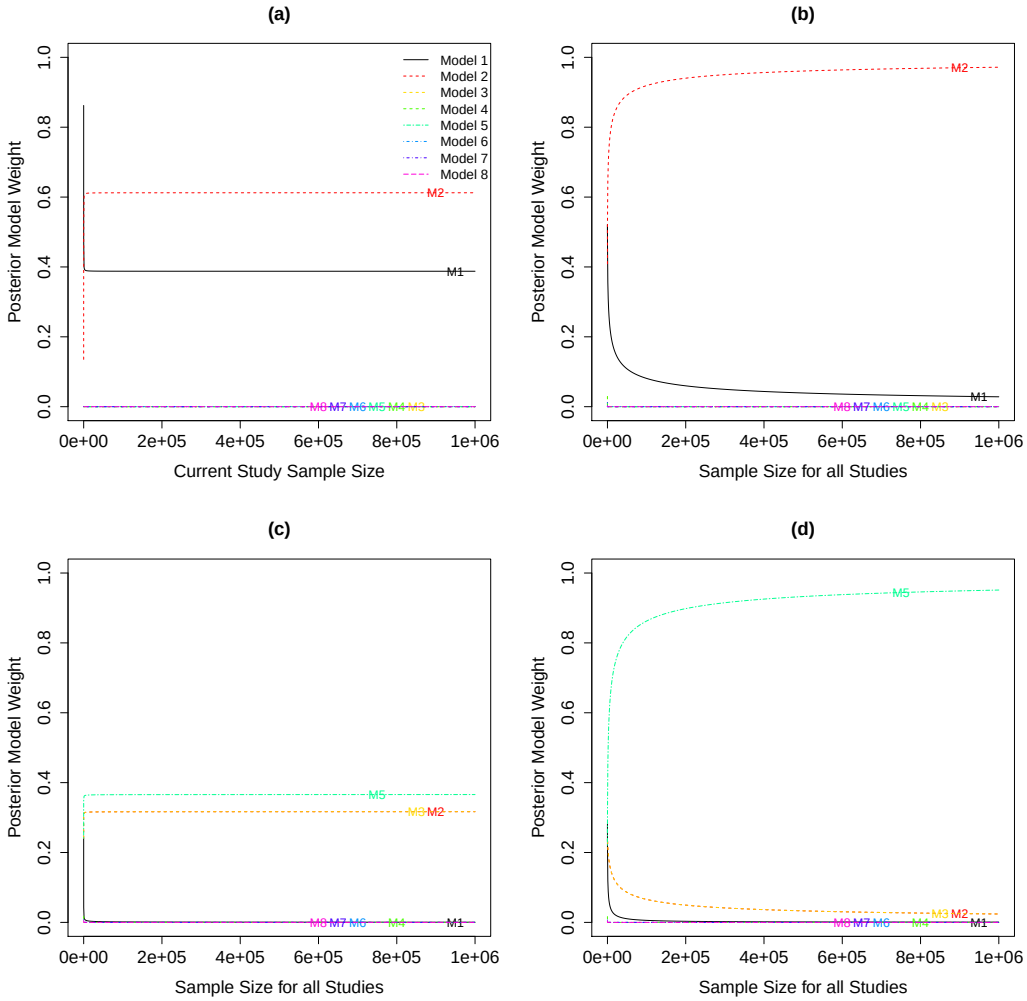

Figure 2.2: Large-sample properties for posterior estimation of model weights for **(a)** MEM with π_n approach as only $n \to \infty$ and **(b)** $n, n_1, n_2, n_3 \to \infty$ assuming S_1 exchangeable (M2). Large-sample properties for posterior estimation of model weights with π_n under **(c)** BMA and **(d)** MEM as $n, n_1, n_2, n_3 \to \infty$ assuming S_1 and S_2 exchangeable (M5). Model labels are imposed along the convergence trajectory to help with identification of overlapping lines.

2.3 Simulation studies to establish small sample properties

In this section we evaluate the small-sample properties of our proposed approach and compare them to two existing methods for integrating supplementary information into the analysis of a primary data set.

2.3.1 Simulation design

Our simulation study considers scenarios that involve a single primary cohort and three supplemental cohorts: S_1, S_2, S_3. We assume that $n = n_h = 100$ and $\sigma = \sigma_h = 4$ for $h = 1, 2, 3$ (i.e., that the sample size and standard deviations are equal for all sources of data). This assumption allows us to evaluate performance without having to discern differences attributable to differing sample sizes or sampling-level variances. Four scenarios are evaluated with different assumptions regarding the sample means. The first scenario considers the case where the sample means were equal for all supplemental sources: $\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = -4$. The second scenario considers the case where the sample means for the first two sources were identical, but the third was very different: $\bar{x}_1 = \bar{x}_2 = -10, \bar{x}_3 = 2$. The third scenario represents the case where the sample means were different for all three supplemental sources: $\bar{x}_1 = -10,\ \bar{x}_2 = -4,\ \bar{x}_3 = 2$. The fourth scenario considers the case where the sample means for the first two supplemental sources were similar, but the third was very different: $\bar{x}_1 = -10,\ \bar{x}_2 = -9.25,\ \bar{x}_3 = 2$.

For each scenario, we simulate data for the primary cohort with a true μ that varied across an equally spaced grid of 420 points from -15 to 6. Data are analyzed with one of four approaches for each simulated trial: (1-2) MEMs with π_e and π_n priors; (3) the empirical Bayes implementation of the commensurate prior approach (CP) (Hobbs et al., 2011); and (4) a standard hierarchical model (SHM) with a uniform hyperprior for the common inter-source standard deviation over the interval (0,50) (Spiegelhalter et al., 2004). Additional values of upper limits for the uniform distribution were explored, but all results performed similarly under the considered scenarios. 10,000 simulated studies are completed for each μ for approaches 1 through 3, while 1000 simulated studies are completed at each μ for approach 4. The performance of each approach is summarized as a function of μ by bias, coverage of the 95% HPD interval, and measures of data integration (as explained in Section

2.3.2). All calculations are completed with `R` (R Core Team, 2013).

2.3.2 Dynamic borrowing

We first consider comparing the extent to which each approach integrated supplemental information using effective supplemental sample size (ESSS). ESSS, which uses the relative gain in posterior precision obtained from a Bayesian model to characterize an additional number of "effective primary" samples effectuated for joint inference, was considered by Hobbs et al. (2013) as an extension of prior effective sample size (Morita et al., 2008). Formally, for any model in which posterior precision is approximately linear in sample size, ESSS is defined as $ESSS = n\left\{\frac{\mathcal{P}(\boldsymbol{x}_p,\boldsymbol{x}_1,...,\boldsymbol{x}_H)}{\mathcal{P}^*(\boldsymbol{x}_p)} - 1\right\}$, where $\mathcal{P}^*(\boldsymbol{x}_p)$ is the posterior precision of the reference model with no borrowing from supplemental sources and $\mathcal{P}(\boldsymbol{x}_p, \boldsymbol{x}_1, ..., \boldsymbol{x}_H)$ is the posterior precision under the joint model which incorporates supplemental information. In the Gaussian case for MEMs, the posterior precision for $\boldsymbol{\Omega}_k$ is $\frac{1}{v}+\sum_{h=1}^{H}\frac{s_{h,k}}{v_h}$ which results in an ESSS which can be calculated exactly as

$$ESSS_{\text{MEM}} = n\sum_{k=1}^{K}\left\{\omega_k\left[\frac{\frac{1}{v}+\sum_{h=1}^{H}\frac{s_{h,k}}{v_h}}{\frac{1}{v}} - 1\right]\right\}. \tag{2.8}$$

Figure 2.3 plots median ESSS curves obtained from our simulation study as a function of μ. In Scenario 1, wherein the primary and supplemental data are truly exchangeable, all methods incorporated a substantial amount of supplementary data when the true mean is equal or close to -4, but the MEM approaches resulted in approximately 1.7 and 2.2 times larger median ESSS (for priors π_e and π_n, respectively) than CP and SHM when $\mu = -4$.

Scenarios 2-4 illustrate that the SHM consistently fails to incorporate supplementary information in the presence of heterogeneity among the supplemental sources. Similar results were observed when using alternative values for the hyperprior upper limit at 5, 10, 20, and 100, as well. The maximum ESSS was larger for π_n than π_e, in general. In addition, π_n resulted in a higher maximum ESSS than CP in 2 of the 3 scenarios, while π_e had a higher maximum sample size than CP in 1 of the 3. Moreover, MEMs facilitate flexible patterns of dynamic borrowing such that ESSS is maximized at the observed supplemental cohort means and minimized in regions devoid of supplemental information. In contrast, CP, which assumes exchangeability among all supplemental sources, yielded high median

ESSS in regions without support from the supplemental information. As a result, the two MEM approaches resulted in larger effective regions of borrowing when compared to CP.

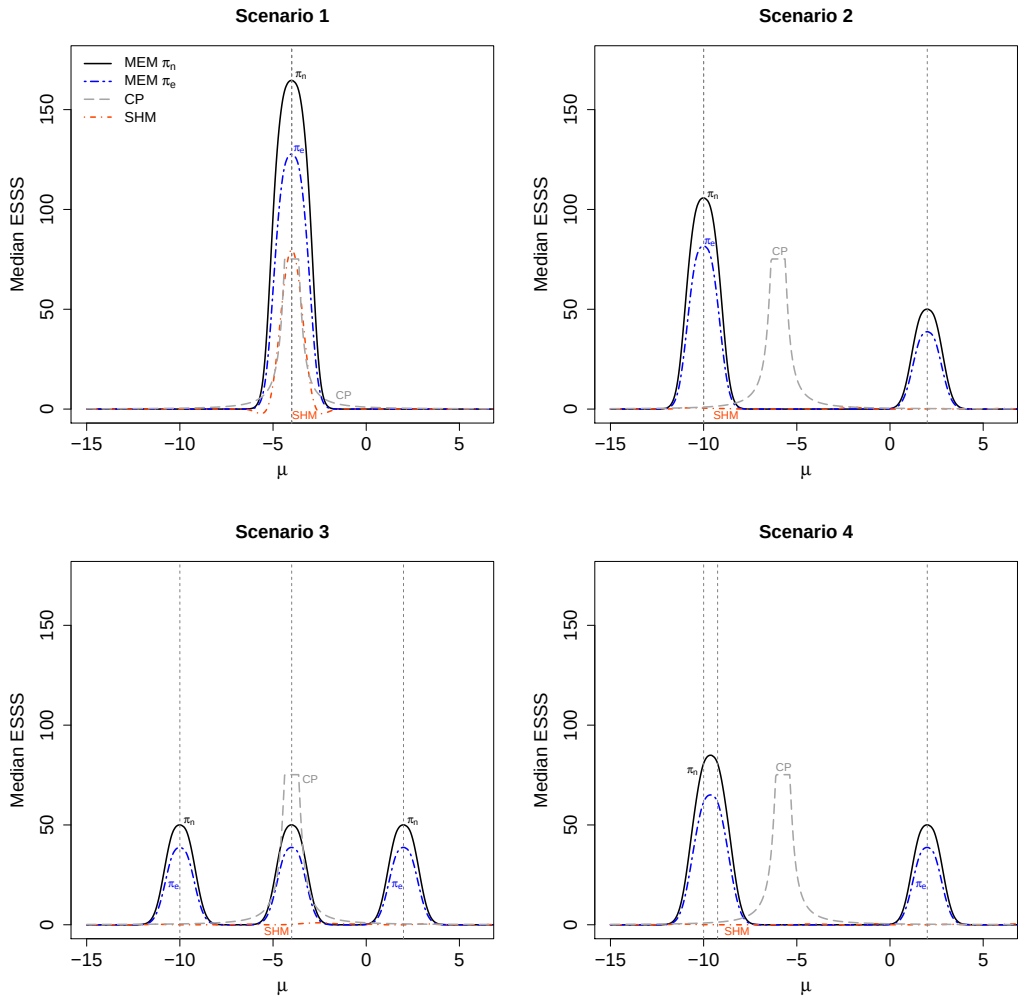

Figure 2.3: Median effective supplemental sample size using MEM with π_e, π_n priors, CP, and SHM under each scenario. Dashed vertical gray lines are used to represent assumed observed values of the supplemental group means for each scenario.

2.3.3 Bias and coverage

This section evaluates trade-offs between the extent to which one can enhance efficiency through integrating supplementary information (as characterized by ESSS) while maintaining other desirable inferential properties. Figures 2.4a and 2.4b present scatter plots that illustrate maximum ESSS as a function of integrated bias per standard deviation (left) and 1-integrated 95% HPD coverage (right). Integrated bias per standard deviation is calculated via Riemann integration of the absolute value of the bias over the simulated points divided by the simulated standard deviation. Integrated coverage is calculated via Riemann integration of the coverage over the simulated points. Additional plots depicting coverage of the 95% HPD interval estimators, bias, and MSE as functions of μ can be found in Appendix B.

Figure 2.4a effectively demonstrates that both CP and SHM intrinsically favor extreme bias versus efficiency trade-offs. This is most evident for the SHM, which either gains efficiency at the expense of a substantial increase in integrated bias (scenario 1) or exhibits no bias but also no increase in efficiency due to its inability to leverage supplemental information in the presence of heterogeneity in the supplemental cohorts (scenarios 2-4). In contrast, CP, represented by diamonds, integrated supplemental information in all scenarios but resulted in an identical level of both maximum ESSS and bias for all four scenarios. As a result, the MEM approach resulted in less integrated bias than CP, in all cases, with π_e illustrating a 47 to 56% decrease in integrated bias and π_n illustrating a 25 to 38% decrease in integrated bias compared to CP. In fact, even in the scenario most favorable to CP and SHM (scenario 1), MEMs effectuated estimators that attained *both* less bias and more efficiency.

Similar trends were observed in Figure 2.4b which describes maximum ESSS as a function of 1-integrated coverage. The MEM approach outperforms CP with π_e yielding greater coverage in 4 of 4 cases and π_n providing greater coverage in 3 of 4 cases while the SHM exhibits either worse integrated coverage than MEMs (scenario 1) or adequate coverage with essentially no borrowing (scenarios 2-4).

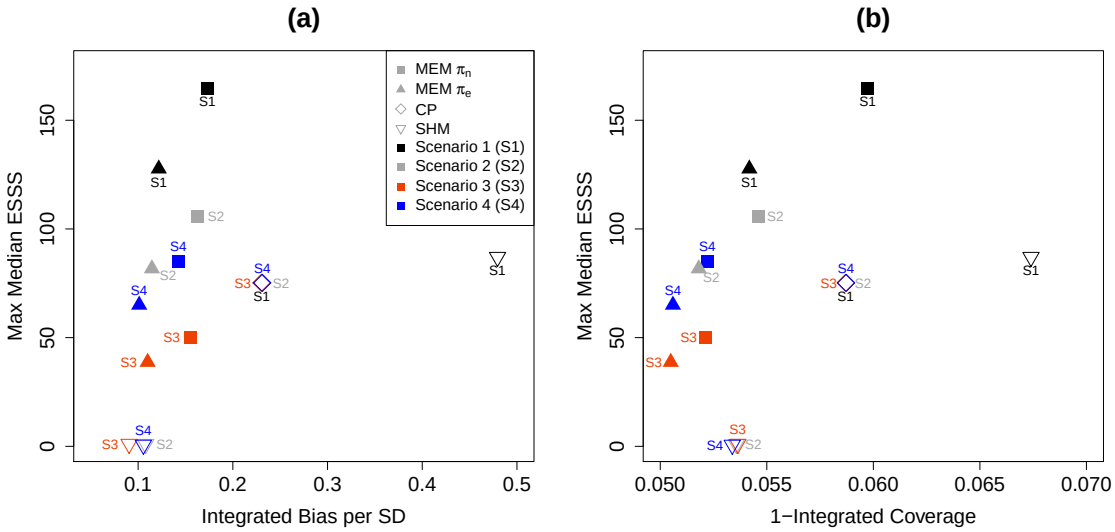

Figure 2.4: Plots demonstrating bias versus shrinkage trade-offs using the methods of CP, SHM, and MEM with π_e, π_n source-inclusion priors. Note that CP overlaps for all four scenarios.

2.4 Application to a regulatory tobacco clinical trial

This section presents a case study illustrating the application of MEMs and competing approaches to data from a recently completed randomized trial of Very Low Nicotine Content (VLNC) cigarettes. CENIC-p1 was a randomized trial devised to evaluate the effect of reducing the nicotine content of cigarettes on tobacco use and dependence (Donny et al., 2015). Subjects were randomized equally to one of seven treatment groups, consisting of a usual brand cigarette condition and six conditions with investigational cigarettes that contained a range of nicotine contents from 15.8 mg of nicotine per gram of tobacco (15.8 mg/g group; approximately the amount of nicotine found in commercially available cigarettes) to VLNC cigarettes with 0.4 mg of nicotine per gram of tobacco (0.4 mg/g group). In addition, one treatment group received cigarettes with 0.4 mg of nicotine per gram of tobacco with high tar to evaluate the impact of tar yield on the effect of VLNC cigarettes (0.4 mg/g, HT group). Our case study compares the change from baseline in cigarettes smoked per day (CPD) between the 0.4 mg/g group and the 15.8 mg/g group.

There are several supplementary sources of data available that could be integrated into our analysis to achieve a more precise estimate of the effect of VLNC cigarettes. First, the usual brand and 0.4 mg/g, HT groups from CENIC would be expected to have similar outcomes to the 15.8 mg/g and 0.4 mg/g groups, respectively, if nicotine is the primary driver of smoking behavior. These data should not be assumed exchangeable, however, due to the impact of "product switching" (i.e. assigning subjects to cigarettes different than their usual brand is likely to impact smoking behavior regardless of the nicotine content of the new cigarette) and tar yield. In addition, data are also available from two historical trials of VLNC cigarettes (Hatsukami et al., 2010, 2013). We will consider the VLNC cigarette group from Hatsukami et al. (2013) and the 0.05 mg nicotine cigarettes group from Hatsukami et al. (2010). Both trials used VLNC cigarettes with similar nicotine content to the 0.4 mg/g group from CENIC-p1, but neither trial used an equivalent to the 15.8 mg/g control condition and, therefore, these historical sources will only provide supplemental data for the VLNC condition. We note that the amount of nicotine in a cigarette can be quantified by either the nicotine yield or the nicotine content. CENIC-p1 used the nicotine content, whereas the two historical studies used the nicotine yield, but the nicotine content of the cigarettes was similar. In summary, we have three potential supplemental sources for the 0.4 mg/g group (0.4 mg/g, HT group from CENIC-p1, Hatsukami et al. (2013), Hatsukami et al. (2010)) and one potential supplemental source for the 15.8 mg/g group (usual brand group from CENIC-p1).

A summary of the observed data describing the mean and standard deviation for the change in CPD for the primary and supplementary data sources can be found in the upper panel of Table 2.1. Results for the CENIC, 0.4 mg/g group and the CENIC, 0.4 mg/g, HT group are nearly identical, suggesting that exchangeability might be a reasonable assumption for these two cohorts, whereas the data from Hatsukami et al. (2013) and Hatsukami et al. (2010) were not consistent with the CENIC 0.4 mg/g group. An ideal method would exhibit the flexibility to integrate the data from the 0.4 mg/g, HT group into the primary analysis, while giving little weight to the data from Hatsukami et al. (2013) and Hatsukami et al. (2010). The two control populations (CENIC 15.8 mg/g and CENIC usual brand) exhibited similar but not entirely consistent results, indicating that some amount of smoothing is appropriate but not to the same extent as for the two 0.4

mg/g conditions.

The lower panel of Table 2.1 provides results obtained from five competing approaches: a standard model that does not allow borrowing, MEMs with the π_e and π_n priors, the empirical Bayesian commensurate prior approach, and the SHM. Figure 2.5 represents the "bias-variance" trade-off from Table 2.1 using the percent change in mean estimation and the percent reduction in the posterior standard deviation relative to the standard model that does not allow borrowing. Compared to the standard approach, the MEM approach with the π_e prior and the SHM show minimal change from the mean estimate at the expense of minor decreases of 4% and 2%, respectively, in the posterior standard deviation. The MEM approach with π_n results in only a 10% change in mean estimation as compared to the standard approach while considerably reducing the posterior standard deviation (29%). This is in contrast to the commensurate prior approach which has a larger percent change in mean estimation than MEMs with π_n, but borrowed half as much as the MEM with π_n prior with a 14% decrease in the posterior standard deviation relative to the standard approach.

		Study		**Source**	**Mean Change**	**SD**
Control Groups		CENIC, 15.8 mg/g group (n=110)		P	5.90	9.15
		CENIC, usual brand group (n=112)		1	7.33	8.38
Treatment Groups		CENIC, 0.4 mg/g group (n=109)		P	-0.23	6.79
		CENIC, 0.4 mg/g, HT group (n=116)		1	-0.15	6.71
		Hatsukami et al. (2013) (n=55)		2	-4.24	9.02
		Hatsukami et al. (2010) (n=32)		3	-7.08	7.02
Group	**Weight**	**No Borrowing**	**MEM π_e**	**MEM π_n**	**CP**	**SHM**
Control	ω	1.000	0.861	0.212	-	-
	ω_1	0.000	0.139	0.788	-	-
Treatment	ω	1.000	0.691	0.423	-	-
	ω_1	0.000	0.305	0.566	-	-
	ω_2	0.000	0.003	0.007	-	-
	ω_3	0.000	0.000	0.000	-	-
	$\omega_{1,2}$	0.000	0.001	0.005	-	-
	$\omega_{1,3}$	0.000	0.000	0.000	-	-
	$\omega_{2,3}$	0.000	0.000	0.000	-	-
	$\omega_{1,2,3}$	0.000	0.000	0.000	-	-
Control		5.90 (0.87)	6.01 (0.89)	6.52 (0.61)	6.21 (0.77)	5.87 (0.85)
Control ESSS		0.0	18.5	105.0	31.3	5.2
Treatment		-0.23 (0.65)	-0.22 (0.55)	-0.21 (0.47)	-0.84 (0.53)	-0.29 (0.65)
Treatment ESSS		0.0	36.5	68.2	56.8	0.7
Δ(Trt-Con)		-6.12 (1.09)	-6.22 (1.05)	-6.73 (0.77)	-7.05 (0.93)	-6.17 (1.07)

Table 2.1: Summary statistics for change in cigarettes smoked daily since baseline by group and posterior model estimates and ESSS for no borrowing, MEM π_e, MEM π_n, CP, and SHM.

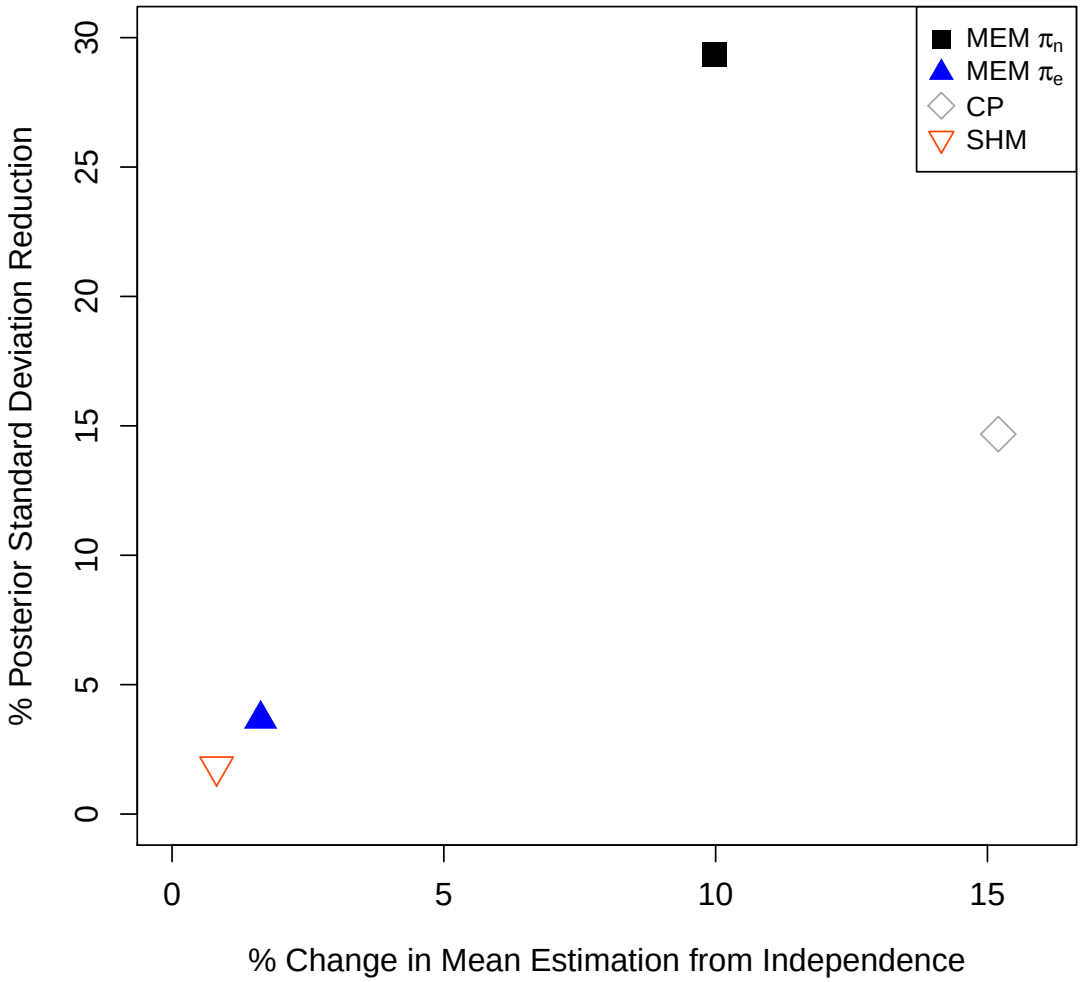

Figure 2.5: Plot comparing percent change in mean estimation from the standard model with no borrowing and the percent reduction in the posterior standard deviation for the difference between the treatment and control groups in Table 2.1 for the standard approach with no borrowing to the CP, SHM, and MEM with π_e, π_n source-inclusion priors approaches.

2.5 Discussion

We proposed multi-source exchangeability models, a Bayesian approach for integrating multiple, potentially non-exchangeable, supplemental data sources into the analysis of a primary data source. The modeling strategy was devised to overcome both limitations arising from "single-source" exchangeability models which produce estimators that tend to ignore supplemental data in the presence of heterogeneity as well as the challenges with implementation of BMA, which is limited by high dimensionality of model prior specification. By way of contrast, the MEM strategy characterizes source-specific shrinkage parameters that synthesize all possible exchangeability relationships between primary and supplemental cohorts, thereby inducing robustness to heterogeneity. Moreover, the method relies on source-specific prior inclusion probabilities for model specification which, unlike BMA, yield consistent shrinkage estimators. The general approach presented in this chapter can be adopted in conjunction with any statistically valid likelihood specification and any set of appropriate supplementary cohorts to be considered for potential integration into the primary cohort. The Gaussian case demonstrates how one can influence the shrinkage of supplemental sources through specification of the prior probability of source inclusion, while preserving consistency.

The MEM model formulation reduces the prior space by placing prior weights on the source-inclusion probabilities but other solutions have also been proposed within the BMA framework. Fernández et al. (2001) propose ten 'benchmark' priors which use little or no subjective prior information for specification of the priors on parameters by choice of a single scalar hyperparameter with a corresponding uniform prior for model weights. Eicher et al. (2011) explore these benchmark priors in addition to a prior specification corresponding to BIC and identify better performance utilizing the BIC-related prior in situations where there are an extremely large number of models to consider (e.g, 2^{40}) while also noting that the different priors can result in very different posterior results for optimal models. These approaches to prior specification demonstrate that there is a strong relationship between sample size and the number of models that must be averaged over when identifying prior specification that result in optimal performance in finite sample sizes. Eicher et al. (2011) also note that different prior specifications on the model weights did not have much impact, but our results suggest that different priors can greatly affect

posterior estimation. BMA may also be approximated using the MC^3 algorithm for model averaging as proposed by Madigan et al. (1995), but this may be infeasible for large K due to computation considerations as the chain may fail to consider all models. By way of contrast, relying on source-inclusion prior probabilities, the MEM model formulation addresses these limitations.

A limitation of any method which attempts to utilize supplemental information is that data integration naturally induces bias. We attempt to control bias through a model specification that facilitates source-specific shrinkage parameters, thereby inducing flexibility in the presence of non- or partially exchangeable supplemental cohorts. Our simulation studies and analytical findings demonstrate the extent to which the proposed MEM approach effectively integrates supplemental information while minimizing bias. Averaging over our four scenarios and two priors, MEMs achieved 41% less integrated bias while effectuating a 93% larger maximum effective supplemental sample size when compared to the commensurate prior approach.

The definition of exchangeability used, where $\mu = \mu_h$, may be seen as unnecessarily stringent. The proposed MEM framework enables source-specific shrinkage to an extent defined by the empirical evidence for exchangeability such that sampling observables arise from identical distributional forms. As an alternative model, the strict assumption of equality could be relaxed by allowing some small term, $\epsilon_h > 0$, potentially specified for each source, such that the extent of shrinkage is determined by $\mu = \mu_h + \epsilon_h$. This model, however, would not be identifiable without inducing sparsity on the domain of ϵ_h, perhaps with a spike-and-slab prior. This would require additional hyperparameter specification beyond that of the MEM framework, which would complicate the calibration to control frequentist error in practical application.

Another limitation is that the type of study, retrospective/observational versus prospective/randomized, affects the quality and reliability of the resulting estimates, and therefore consideration of design type should be taken into account. For example, as conveyed in Pocock's "Acceptability Criteria," (1976) decisions pertaining to which supplemental sources should be considered for integrative analysis should be based on the study objectives, as well as controlling sources of potential bias arising from inconsistent eligibility criteria, application of the interventions, or outcome ascertainment. Similarly, when using

data arising from non-randomized designs, confounding due to selection bias needs to be accounted for in some manner, such as through existing methods for causal inference using propensity scoring, inverse-probability weighting, or matching. We endeavor to extend the methodology to facilitate integration of multiple trials while accounting for confounding from non-randomized designs as future work.

Chapter 3

A Multi-source Adaptive Platform Design for Emerging Infectious Diseases

3.1 Introduction

In the context of the dire, rapidly developing Ebola outbreak introduced in Chapter 1, the NIH launched the Partnership for Research on Ebola Vaccines in Liberia II (PREVAIL II) trial, a randomized clinical trial to evaluate medical countermeasures against Ebola virus disease (EVD). PREVAIL II was designed as a sequential platform design, where multiple experimental therapeutics were evaluated sequentially compared to the optimal standard of care (oSOC). Treatments that showed a significant improvement relative to the oSOC were added to the oSOC for all future comparisons. This allowed PREVAIL II the potential to rapidly evaluate multiple treatments, accelerate clinical development, and maintain flexibility in the context of an emerging infectious disease epidemic within one master protocol. However, particular aspects of the platform design methodology could be enhanced for future outbreaks. One major limitation is that the proposed design only allowed the use of contemporaneous controls. For instance, if the initial drug (drug A) represented a significant improvement over the oSOC in the initial segment of the trial, the second segment would evaluate drug B in addition to drug A in a randomized design,

i.e. oSOC + drug A vs. oSOC + drug A + drug B, but only subjects from segment two would be used to evaluate drug B. This ignores subjects randomized to receive oSOC + drug A in the initial segment of the trial.

In this chapter, we illustrate how MEMs and adaptive randomization (AR) can be incorporated into the PREVAIL II master protocol to achieve improved operating characteristics relative to the original PREVAIL II design. In Chapter 2, we illustrated that MEMs can be used to integrate information arising from potentially non-exchangeable normal populations while minimizing bias. Integrating supplemental information using MEMs dynamically determines if the non-contemporaneous segments in PREVAIL II are exchangeable (e.g., if the segments have equivalent mortality rates) and thereby shares information across segments, if appropriate, to achieve more precise estimates of the disease-response rate and increase power.

In addition, we will incorporate AR through an extension of the dynamic allocation procedure proposed by Hobbs et al. (2013), which targets information balance across treatment groups. Borrowing information from previous segments using MEMs may lead to imbalances in statistical information across treatment groups within a segment because supplemental information are only available for the control group, but, by adapting the randomization ratio to allocate more subjects to the treatment group, we are able to balance statistical information within a segment. Balancing information by boosting allocation to the novel treatment arm improves statistical power to detect effective treatments. As effective novel therapies emerge in the platform, boosting allocation to the treatment arm in the presence of evidence for inter-segment exchangeability among controls also has the potential to improve outcomes for trial participants.

It is important to note that the proposed multi-source AR differs fundamentally from conventional response- or outcome-AR methods. A recent article by Thall et al. (2015) evaluated the impact of outcome-AR, concluding that designs that use outcome-AR attain diminished power, often to a considerable extent, necessitating a larger overall sample size. Moreover, outcome-AR risks large imbalances in sample size in the wrong direction, assigning more patients to inferior treatments in the presence of the small to moderate effect sizes observed in practice. In contrast, our proposed AR scheme endeavors to balance total effective information in the presence of potentially non-exchangeable supplemental cohorts,

which maximizes power for comparing treatment groups.

The remainder of the chapter proceeds as follows. First, the standard design of the PREVAIL II master protocol is introduced in Section 3.2, followed by our proposed design which incorporates MEMs and AR in Section 3.3. The scenarios considered for the simulation studies, the process for design calibration, and results for the simulation studies are presented in Section 3.4. We conclude with a brief discussion in Section 3.5.

3.2 Standard design of PREVAIL II master protocol

The initial objective of PREVAIL II was to sequentially evaluate multiple candidate therapies for the treatment of EVD. Each treatment was to be evaluated in a separate trial segment and each segment to consist of a separate randomized trial to compare the new treatment versus the current oSOC. Treatments found to offer a significant survival benefit compared to the standard of care would be added to the standard of care for all future segments. Figure 3.1(I) graphically depicts an example with 3 segments and four different treatment combinations represented by color. In segment 1, the oSOC arm (blue) is compared to an experimental arm of drug A + oSOC (yellow) with the proportion randomized to the experimental arm fixed at $\tau = 0.5$ (also represented by the equally sized triangles signifying equal enrollment throughout the segment). If the experimental arm was determined to provide significant improvement over the standard of care, then the next trial segment would consist of a comparison between the updated oSOC, drug A + oSOC (yellow), and a new experimental arm, drug B + drug A + oSOC (green). However, if the drug B + drug A + oSOC does not demonstrate a significant improvement, drug A + oSOC is carried forward to the next segment where drug A + oSOC (yellow) is compared to drug C + drug A + oSOC (maroon).

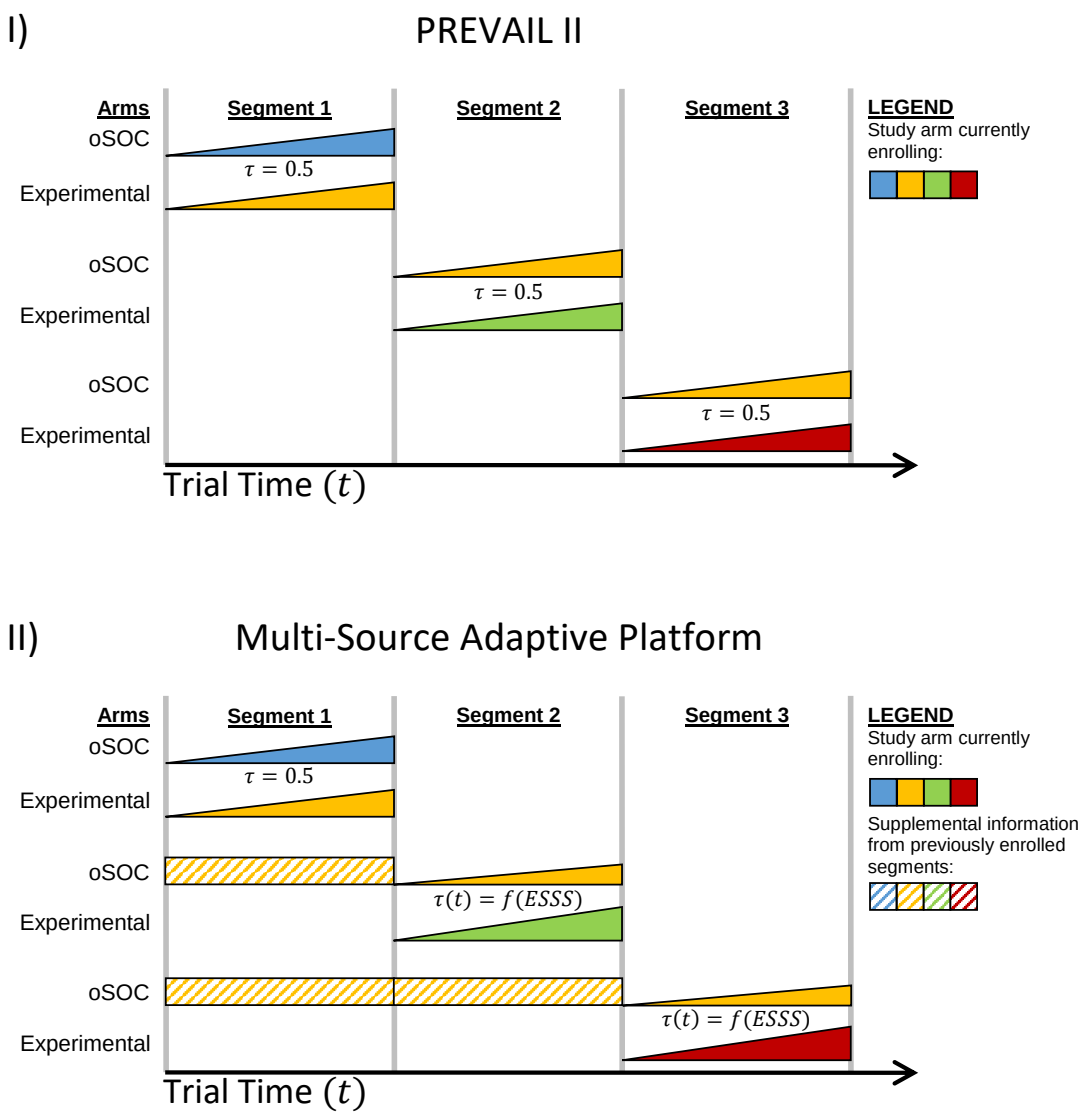

Figure 3.1: Example comparing three segments of a trial with (I) PREVAIL II master protocol which only compares contemporaneously enrolled subjects with equal allocation to the study arms, $\tau = 0.5$, versus (II) framework with methods to potentially incorporate non-contemporaneous data and ability to adaptively alter the randomization ratio as a function of the effective supplemental sample size, $\tau(t) = f(ESSS)$. Equally sized triangles further indicate segments with equal allocation versus smaller oSOC triangles which indicate the potential for greater allocation to the experimental arm in the presence of supplemental information.

The PREVAIL II master protocol allowed for the rapid evaluation of multiple candidate treatments. Within a segment, PREVAIL II used frequent sequential monitoring in the Bayesian paradigm to allow early termination in the event that a treatment provided a substantial survival benefit over the standard of care (Dodd et al., 2016; Proschan et al., 2016). The primary outcome for PREVAIL II was a binary indicator of 28-day mortality. Let x_A be the number of deaths, n_A be the total number of subjects randomized, and p_A be the 28-day mortality rate for hypothetical treatment arm A. Define x_B, n_B and p_B analogously for hypothetical control arm B. Assuming independent beta($\alpha = 1, \beta = 1$) priors for p_A and p_B results in independent beta posteriors with $\alpha = 1+x_A$, $\beta = 1+n_A-x_A$ for p_A and $\alpha = 1 + x_B$, $\beta = 1 + n_B - x_B$ for p_B.

Formal inference on the 28-day mortality rate was based on the posterior distribution. The posterior probability that the 28-day mortality rate in arm A is less than arm B is:

$$P(p_A < p_B | x_A, x_B) = \sum_{k=x_A+1}^{n_A+1} \frac{\binom{n_A+1}{k}\binom{n_B+1}{x_B}(n_B - x_B + 1)}{\binom{n_A+n_B+2}{k+x_B}(n_A + n_B - k - x_B + 2)}. \tag{3.1}$$

PREVAIL II was planned with a within-segment maximum sample size of 100 subjects per arm. If the maximum sample size were reached, treatment A would be declared a significant improvement over the control if $P(p_A < p_B | x_A, x_B) \geq 0.975$.

PREVAIL II enabled aggressive interim monitoring to stop the trial early if the new treatment represented a substantial improvement over the standard of care. Interim analyses were first completed after six subjects were randomized to each arm and were completed after every 2 subjects until data were available for 40 subjects. After the first 40 subjects, interim monitoring was carried out after every 40 subjects until a maximum of 200 subjects were enrolled (100 per arm). The trial would stop and declare treatment A a significant improvement over the control if $P(p_A < p_B | x_A, x_B) \geq 0.999$. Simulation results demonstrated that this design had an overall within-segment type-I error rate near 0.03 and 86% power to detect significant difference assuming a relative risk of 0.5 for the new treatment (Dodd et al., 2016).

3.3 Methods

A limitation of the PREVAIL II design is that only contemporaneous information from the current segment is included in the analysis. Thus, relevant data from prior segments was ignored. In fact, after the initial segment, non-contemporaneous, supplemental data is always available for the control arm from one or more previous segments. While avoiding the introduction of inter-cohort bias, the PREVAIL II design makes inefficient use of the data. Utilizing Bayesian methods that estimate partial exchangeability across segments, however, overcomes this inefficiency, resulting in increased power (or decreased total sample size), while protecting against bias in the presence of systematic differences between supplemental and contemporaneous controls.

3.3.1 General framework of multi-source adaptive designs

To address this limitation, we propose the general conceptual design graphically represented in Figure 3.1(II). Segment 1 is identical to the original design proposed for PREVAIL II; patients are randomized with a fixed allocation ratio of $\tau = 0.5$ to the oSOC (blue) or experimental drug A + oSOC (yellow). However, after the first segment, there will always be supplemental, non-contemporaneous control arm information acquired from past segments which can be integrated into future comparisons using a dynamic Bayesian model. The figure depicts data acquired from prior segments by rectangles with diagonal lines placed in the segment of observation. For example, in Segment 2, supplemental data for the controls are available from Segment 1, which in our figure comes from the yellow study arm. In Segment 3, supplemental data for the controls are available from the first two segments.

Incorporating supplemental information from previous segments can potentially result in imbalances in the total effective information between treatment groups within a segment if a fixed allocation to the experimental arm of $\tau = 0.5$ is maintained. Extending the AR method proposed by Hobbs et al. (2013) to the setting of a sequential platform design attenuates this imbalance and maximizes power. This is achieved by allowing the allocation ratio to vary as a function of the effective supplemental sample size (ESSS). Recall, ESSS is a measure reflecting the extent of relative gain in the posterior precision

obtained from a Bayesian model when compared to a model that neglects the supplemental sources. The measure is intended to characterize the effective number of samples incorporated from supplemental sources. By defining the allocation ratio as a function of ESSS ($\tau(t) = f(ESSS)$), the proposed AR methodology aims to balance total information across treatment groups within a segment. Within Figure 3.1(II) this potentially unequal allocation to the treatment arms is represented by the differing slopes, which are adjusted in relation to the extent of estimated exchangeable data contributed by concordant treatment regimes during segments 2 and 3.

The remainder of this subsection presents notation to explain the general framework for multi-source AR. During the initial period of a segment, equal allocation between arms is used until sufficient information is acquired to facilitate estimation of inter-segment exchangeability, after which block-randomization is used to update the allocation ratios according to estimates of ESSS. Let n_{burn} represent the number of patients which must be observed for the "burn-in" period at the start of any segment where supplemental information is available, $\mathcal{B}$ represent the total number of blocks to adaptively randomize patients after the burn-in period, and t_b be the "time" of the b^{th} interim analysis at the start of a block. Additionally, define $n_A(t_b)$ and $n_B(t_b)$ as the total sample size assigned in the current segment to the treatment and control arms at t_b, respectively, $ESSS(t_b)$ as the estimated effective supplemental sample size from the data at t_b, and $R(t_b)$ as the number of subjects left to be randomized at t_b assuming the maximum sample size of the segment is achieved. Recall, the objective is to balance total effective information at trial completion such that, at t_b, $n_A(t_b) = ESSS(t_b) + n_B(t_b)$. Thus, allocation is needed in relation to $n_A(t_b) + \tau R = ESSS(t_b) + n_B(t_b) + (1 - \tau)R$. Therefore, under the the aim of balanced allocation, assignments to treatment for the next block of patients attains the following formulation

$$\tau(t_b) = \frac{1}{2}\left(\frac{ESSS(t_b) + n_B(t_b) - n_A(t_b)}{R(t_b)} + 1\right). \tag{3.2}$$

$\tau(t_b)$ can range between 0 and 1 depending on the extent of shrinkage to supplemental information with a value of 0 implying all patients are randomized to the control arm, a value of 0.5 implying a 1:1 allocation ratio, and a value of 1 implying all patients are randomized to the treatment arm.

3.3.2 Incorporating supplemental information with Bayesian modeling using MEMs

Our proposed multi-source adaptive design as described thus far is general and can be implemented using any method for incorporating supplemental information. While there are many potential methods available for incorporating supplemental information, the MEM framework is specifically considered herein on the basis of recent efforts demonstrating its desirable properties for yielding shrinkage estimators in the presence of non- or partially exchangeable cohorts while avoiding highly parameterized models (which are computationally infeasible for implementation and calibration of sequential design).

Recall from Chapter 2, the MEM framework takes the H supplemental segments available for incorporation and maps them to $2^H = K$ multi-source exchangeability models, denoted $\mathbf{\Omega}_k$, which represent all possible combinations of assumptions of exchangeability between the current segment and the H supplemental segments. For an example in the context of our proposed design, refer back to Figure 3.1(II) where analyses at segment 3 would have four possible MEMs: no supplemental segments assumed exchangeable with segment 3 ($\mathbf{\Omega}$), only segment 1 assumed exchangeable with segment 3 ($\mathbf{\Omega}_1$), only segment 2 assumed exchangeable with segment 3 ($\mathbf{\Omega}_2$), and both segments 1 and 2 assumed exchangeable with segment 3 ($\mathbf{\Omega}_{1,2}$). The MEM framework produces a posterior estimate over these K models using posterior model weights, ω_k, such that $\sum_{k=1}^{K} \omega_k = 1$. MEMs comprising non-exchangeable supplemental data receive smaller posterior weights whereas models contributing only exchangeable sources carry more influence in the posterior distribution. A resultant smoothed posterior estimator synthesizing all possible exchangeability relationships is used for inference.

If the standard beta-binomial model is updated to accommodate MEMs in the control arm, a similar structure to the current PREVAIL II master protocol can be utilized which is able to incorporate supplemental data from previous segments, making more efficient use of available evidence and potentially improving the power of the trial. In the setting of PREVAIL II, we have supplementary data for the control arm, arm B, but not for the treatment arm, arm A. Therefore, we will model arm A using the beta-binomial model and model arm B using MEMs. Introducing formal notation, the marginal posterior distribution of p_B given the observable data, D, from the current segment's controls and the observable

data from H supplemental segments is derived as the weighted average of the posterior distributions for the K multi-source exchangeability models, $q(p_B|\mathbf{\Omega}_k, D)$:

$$q(p_B|D) = \sum_{k=I}^{K} \omega_k q(p_B|\mathbf{\Omega}_k, D). \tag{3.3}$$

The posterior model weight, ω_k, for each MEM is given by

$$\omega_k = pr(\mathbf{\Omega}_k|D) = \frac{p(D|\mathbf{\Omega}_k)\pi(\mathbf{\Omega}_k)}{\sum_{j=I}^{K} p(D|\mathbf{\Omega}_j)\pi(\mathbf{\Omega}_j)}, \tag{3.4}$$

where $p(D|\mathbf{\Omega}_k)$ is the integrated marginal likelihood for $\mathbf{\Omega}_k$ and $\pi(\mathbf{\Omega}_k)$ is the prior probability that $\mathbf{\Omega}_k$ is the true model. The formulation of posterior model weights in (3.4) utilizes a framework similar to Bayesian model averaging (BMA), however the MEM framework reduces the dimension of the prior weight space by enabling specification on the supplemental sources rather than models as described in Chapter 2.

Using the notation from Section 3.2 for arm A and the current segment for arm B, let $x_{B,h}$ be the number of deaths observed in arm B for non-contemporaneous supplemental segment h ($h = 1, ..., H$), $n_{B,h}$ be the number of subjects randomized to arm B in segment h, and $p_{B,h}$ be the 28-day mortality rate for arm B in segment h. Let S_h denote an indicator function of whether or not the non-contemporaneous supplementary source h is assumed exchangeable (i.e., if $S_h = 1$, $p_{B,h} = p_B$). A model, $\mathbf{\Omega}_k$, is then defined by considering a set of source-specific indicators, $(S_1 = s_{1,k}, ..., S_H = s_{H,k})$, where $s_{h,k}$ is a binary indicator of whether or not source h is assumed exchangeable with the primary data in $\mathbf{\Omega}_k$. Assuming independent beta(α, β) priors on p_B and $p_{B,1}$,...$p_{B,H}$, the integrated marginal likelihood for each MEM can be written as follows:

$$\begin{aligned} p(D|\mathbf{\Omega}_k) = \frac{B\left(x + \alpha + \sum_{h=1}^{H} s_{h,k}x_h, n + \beta - x + \sum_{j=1}^{H} s_{j,k}(n_j - x_j)\right)}{B(\alpha, \beta)} \times \\ \prod_{i=1}^{H} \left(\frac{B(x_i + \alpha_i, n_i + \beta_i - x_i)}{B(\alpha_i, \beta_i)}\right)^{1-s_{i,k}}, \end{aligned} \tag{3.5}$$

where $B()$ represents the beta function. The marginal likelihood results in the following

MEM-specific posterior distribution used to calculate the marginal posterior distribution in (3.3):

$$q(p_B|D,\mathbf{\Omega}_k) = \text{Beta}\left(x+\alpha+\sum_{h=1}^{H} s_{h,k}x_h, n+\beta-x+\sum_{j=1}^{H} s_{j,k}(n_j-x_j)\right). \tag{3.6}$$

Therefore the posterior distribution of (3.3) for the 28-day mortality rate for the MEM estimator is a mixture of beta distributions encompassing all possible exchangeability relationships.

Since supplementary data are only available for the control arm (arm B), the marginal posterior probability that $p_A < p_B$ is a weighted average of the conditional posterior probability that $p_A < p_B$ for all possible assumptions about exchangeability:

$$P_{MEM}(p_A < p_B|x_A, x_B) = \sum_{i=1}^{K} \omega_i P(p_A < p_{B,\mathbf{\Omega}_i}|x_A, x_{B,\mathbf{\Omega}_i}). \tag{3.7}$$

In the context of AR with MEMs, a 1:1 allocation ratio is assumed during the burn-in period. The specific calculations used for ESSS in (3.2) are defined as follows. For each individual MEM, in the context of the beta-binomial model, the posterior effective sample size (ESS) can be generally derived as

$$ESS(\mathbf{\Omega}_k) = \alpha+\beta+n_B+\sum_{h=1}^{H} s_{h,k}n_{B,h}. \tag{3.8}$$

The posterior ESSS for the overall MEM estimate is then calculated as the weighted average of the difference from each individual MEM's ESS and the current control arm's sample size: $ESSS = \sum_{k=1}^{K} \omega_k[ESS(\mathbf{\Omega}_k) - n_B]$. Further, it should be noted that the beta(α,β) prior in the beta-binomial model confers the effective information of $\alpha+\beta$ subjects in (3.8). Therefore, the MEM facilitates a non-zero ESSS of $\alpha+\beta$ when assuming the prior probability of 1 on the independence model (e.g., the model which does not borrow strength across segments).

3.3.3 MEM prior probability specification

As with any Bayesian model, the properties of MEMs depend on the prior specification assumed for the model weights, with more flexible choices imparting robustness for posterior inference. Since supplemental sources are assumed independent in the MEM framework, the prior model weight formulation can be specified as the product of the source-specific prior inclusion probabilities: $\pi(\mathbf{\Omega}_k) = \pi(S_1 = s_{1,k}, ..., S_H = s_{H,k}) = \pi(S_1 = s_{1,k}) \times \cdots \times \pi(S_H = s_{H,k})$ (Kaizer et al., 2017). While there are numerous strategies that could be used to identify potential priors for each source, this section considers specific fully Bayesian and empirical Bayesian approaches which were found to achieve desirable operating characteristics in our simulation study.

Our proposed fully Bayesian prior, denoted by π_e, assumes equal prior weight for inclusion and exclusion for all supplementary sources: $\pi_e(S_h = 1) = \frac{1}{2}$. This prior provides impartiality to which supplemental segments should be considered exchangeable with the primary segment.

In contrast to the fully Bayesian approach, an empirical Bayesian (EB) approach utilizes the data collected to inform the prior distribution by maximizing the marginal likelihood with respect to the prior weights. For the proposed MEM model for binary data discussed above, the marginal likelihood is maximized by placing a prior inclusion weight of 1 on sources assumed exchangeable, while all other supplemental sources receive a prior inclusion weight of 0. This induces posterior weights of 1 for the model which maximizes the marginal density and 0 for all other models. The proposed EB prior is denoted by π_{EB}.

However, placing all of the weight on a single MEM may induce less than ideal operating characteristics under circumstances where the marginal density of multiple MEMs may be close to the maximum marginal density. Therefore, a constrained EB prior is proposed, denoted π_{EB_c}, where $0 \leq c \leq 1$, such that the marginal density is maximized under the constraint that the prior source inclusion probabilities must be less than c. This results in a prior inclusion probability of c for segments assumed exchangeable in the MEM that maximizes the marginal density, with all other segments receiving a prior inclusion probability of 0. When $c = 1$, the standard EB formulation is achieved, and when $c = 0$, there is no borrowing of supplemental information. Constraining the optimization over $\mathbf{\Omega}_k$ with π_{EB_c} attenuates bias and avoids over smoothing in the presence of limited evidence

for exchangeability.

As an illustrative example of the behavior of the posterior weights for these priors, consider the case where there are three supplemental segments. Let $n_B, n_{B,1}, n_{B,2}, n_{B,3} = 100$, $x_B = 50$, $x_{B,1} = 52$, $x_{B,2} = 45$, and $x_{B,3} = 65$. The calculated posterior weights for each MEM are provided in Table 3.1. We see that besides π_e, no other prior gives any weight to S_3 which has an estimated mortality of 0.65 and is the most different from our primary segment with its 0.50 estimated mortality. Further, as c decreases for the π_{EB_c} priors we see that $\mathbf{\Omega}_5$, which assumes both S_1 and S_2 are exchangeable, receives less weight while $\mathbf{\Omega}_1$, which assumes no supplemental sources are exchangeable, receives increasing amounts of the posterior weight.

	Sources in $\mathbf{\Omega}_k$								
MEM	P	S_1	S_2	S_3	π_e	π_{EB_1}	$\pi_{\text{EB}_{.9}}$	$\pi_{\text{EB}_{.5}}$	$\pi_{\text{EB}_{.1}}$	π_{EB_0}
$\mathbf{\Omega}_1$	1	0	0	0	0.025	0.000	0.000	0.030	0.420	1.000
$\mathbf{\Omega}_2$	1	1	0	0	0.136	0.000	0.026	0.165	0.256	0.000
$\mathbf{\Omega}_3$	1	0	1	0	0.111	0.000	0.021	0.134	0.208	0.000
$\mathbf{\Omega}_4$	1	0	0	1	0.015	0.000	0.000	0.000	0.000	0.000
$\mathbf{\Omega}_5$	1	1	1	0	0.556	1.000	0.953	0.672	0.116	0.000
$\mathbf{\Omega}_6$	1	1	0	1	0.064	0.000	0.000	0.000	0.000	0.000
$\mathbf{\Omega}_7$	1	0	1	1	0.012	0.000	0.000	0.000	0.000	0.000
$\mathbf{\Omega}_8$	1	1	1	1	0.081	0.000	0.000	0.000	0.000	0.000

Table 3.1: Posterior model weights for each multi-source exchangeability under various priors to demonstrate the impact different priors can have on the resulting weights. P represents the current segments, S_h represents the supplemental segments potentially available for incorporation ($h = 1, 2, 3$).

3.4 Simulation Study

Simulation was used to evaluate and compare the operating characteristics of the PREVAIL II master protocol and the proposed multi-source AR approach. Data were generated assuming an underlying mortality rate, p_{oSOC}, for oSOC alone, and the mortality rate for each potential drug combination was defined through a multiplicative model utilizing the relative risk (RR) of each drug and assuming no interactions. For example, the mortality

rates for the various combinations of oSOC and two potential treatments are:

$$\begin{aligned}
\text{oSOC} &= p_{oSOC}, \\
\text{oSOC+Drug A} &= p_{oSOC} \times \text{RR}_A, \\
\text{oSOC+Drug B} &= p_{oSOC} \times \text{RR}_B, \\
\text{oSOC+Drug A+Drug B} &= p_{oSOC} \times \text{RR}_A \times \text{RR}_B.
\end{aligned}$$

The multi-source AR approach assumes independent beta($\alpha = 1, \beta = 1$) priors for p_A and p_B and hypothesis testing at the interim and final analyses is based on the marginal posterior probability that p_A is less than p_B. As in PREVAIL II, our proposed design will stop and declare the experimental arm to be a significant improvement over the control if $P(p_A < p_B | x_A, x_B) > 0.999$ for all interim analyses. The posterior probability thresholds used at the final analysis will be calibrated for each prior to achieve the desired operating characteristics, as described in Section 3.4.1.

Operating characteristics for the multi-source AR approach using MEMs with π_e and π_{EB_c} were compared to the naive approach of pooling all available supplemental information regardless of exchangeability, which will maximize the amount of supplemental information available, but introduces a prohibitive extent of bias in the presence of non-exchangeable supplementary segments. Rather than adopting the aggressive interim monitoring of PREVAIL II, we propose interim monitoring after the enrollment of every 40 subjects until the end of the burn-in period, where it will then follow the practical schedule of interim analyses at the start of each block, at t_b, with $n_{burn} = 60$ and $\mathcal{B} = 5$. This implies interim analyses after 40, 60, 95, 130, and 165 patients were observed, but updating the allocation ratio using (3.2) only occurs after 60, 95, 130, and 165 patients are observed. If the trial does not terminate early for superiority it will proceed to enroll a total of 200 patients in the current segment and conduct the final analysis after all information is collected. Additionally, bounds are placed on (3.2) to ensure $\tau(t_b) \in [0, 1]$.

We considered the sequential testing of 5 potential therapeutics in the context of our platform design with two scenarios for the underlying oSOC mortality rate: (1) a constant mortality rate for all segments: $p = 0.40$ and (2) a decreasing mortality rate by segment, $p = (0.74, 0.61, 0.48, 0.36, 0.23)$. These decreasing values reflect observed mortality rates as

the Ebola epidemic progressed from May 2014 to December 2014 in Sierra Leone (Dodd et al., 2016). The varying mortality scenario is more challenging for MEMs than the constant mortality case since the supplemental controls are not exchangeable, in which case minimal borrowing is preferred.

Further, five different therapeutic RR profiles are examined: (1) all drugs have a null effect (RR=1) and (2-5) one drug in the treatment pipeline has a moderate effect in segment 2, 3, 4, or 5 with RR=0.7. In a rapidly evolving epidemic it may very well be that no or very few of the included therapeutics demonstrate an improvement. Moreover, location in the pipeline may impact the platform's operating characteristics.

3.4.1 Parameter calibration

To provide context, the original PREVAIL II design had a type-I error rate of around 0.03 with 86% power to reject the null hypothesis assuming a RR of 0.5, a baseline 28-day mortality rate of 0.4, and a posterior probability threshold of 0.975 at the final analysis within a segment if the trial did not terminate early. While the same posterior probability threshold could be used for MEMs, performance will be optimized if the posterior probability threshold is optimized to achieve the desired operating characteristics. Furthermore, optimal characteristics may be achieved with a threshold that varies by segment because more supplemental information will be available at later segments as compared to earlier segments.

Given the two scenarios for the underlying mortality rate, two potential processes to calibrate the posterior probability thresholds are considered. First, thresholds can be calibrated to achieve a type-I error rate of approximately 0.025 within a segment for the constant mortality scenario and the operating characteristics under the varying mortality scenario can be evaluated to determine if the inflation in the type-I error rate is within acceptable levels. Alternatively, thresholds can be calibrated to limit the inflation of the type-I error rates in the varying mortality scenario while maintaining similar power in the constant mortality scenario to that observed for the PREVAIL II design. To address the first case, potential thresholds are identified via simulation without interim monitoring using a gradient descent algorithm with $n_{burn} = 60$ and $\mathcal{B} = 5$ until the average segment-wise type-I error rate is between 0.024 and 0.026. These estimates are used as the initial values

Scenario	Approach	Segment 1	2	3	4	5
Constant	$\pi_{\text{EB}_{10}}$	0.975	0.97125	0.96625	0.95875	0.95750
	π_e	0.975	0.96375	0.95875	0.94375	0.93250
	POOL	0.975	0.97375	0.96375	0.95375	0.94000
Both	$\pi_{\text{EB}_{10}}$	0.975	0.97500	0.97750	0.97500	0.97500
	π_e	0.975	0.98150	0.98500	0.98750	0.98750
	POOL	0.975	0.99500	0.99900	0.99750	0.99900

Table 3.2: Posterior probability thresholds for MEMs with empirical Bayesian prior ($\pi_{\text{EB}_{10}}$), MEMs with fully Bayesian uniform prior (π_e), and the naive pooling (POOL) approach presented to optimize to maintain average type-1 error rate across scenarios for constant underlying mortality scenario only ("Constant") or to minimize trade-off in inflation of type-1 error in constant underlying mortality scenario while maintaining power equal to, or greater, than the PREVAIL II scenario under the varying underlying mortality scenario ("Both").

to further refine thresholds to achieve desired performance. In the second case, thresholds are identified by increasing the posterior threshold segment-by-segment to achieve equal or greater power compared to the PREVAIL II design in the constant mortality scenario while attempting to minimize the inflation of the type-I error rate in the varying mortality scenario as compared to the PREVAIL II design. The latter approach may result in a trade-off between power and type-1 error but, in the context of an emerging disease outbreak, this is beneficial due to the importance of maintaining power to detect effective treatments.

3.4.2 Results

25,000 simulated trials were completed for each scenario. Operating characteristics are presented for the PREVAIL II master protocol, MEMs with the fully Bayesian uniform prior (π_e), MEMs with the constrained EB prior (π_{EB_c}), and naive pooling. Results are presented for thresholds calibrated to achieve the desired type-I error rate in the constant mortality scenario as described in Section 3.4.1, with the thresholds identified under each approach to calibration presented in Table 3.2. The value of $c = 0.10$ was selected for π_{EB_c} based on extensive sensitivity analyses (not presented) that considered values of c from 0.05 to 0.50.

The operating characteristics presented for each scenario include the probability of attaining a positive test within a segment based on the Bayesian posterior probability thresholds, the mean (sd) total number of subjects (N) treated throughout all segments as a measure of early termination, the mean (sd) proportion randomized to the treatment arm in segments 2-5 as a measure of adaptive randomization performance, and the mean (sd) proportion who survived either across segments 2-5 in the null case or in the specific non-null segment in scenarios with an efficacious therapy. When considering a drug under the null case it is ideal to rarely attain a positive test within a segment (i.e., have a probability of rejecting near 0), whereas it is desirable to attain a positive test within segments that include efficacious treatment (i.e., have a probability of rejecting near 1). Further, the proportion surviving in the null scenario is expected to be identical in PREVAIL II and the proposed AR design, but an improvement in survivorship is expected in non-null segments when the AR design effectuates more allocation to the treatment arm.

Table 3.3 presents simulation results for the constant mortality scenario. The average type-I error rate across segments for MEMs is similar to or less than the average type-I error for the PREVAIL II master protocol. However, the power to detect an effective drug is higher in every non-null segment for $\pi_{\text{EB}_{10}}$ and π_e compared to PREVAIL II, with increases in power ranging from 9% to 27% and 29% to 69% for $\pi_{\text{EB}_{10}}$ and π_e, respectively. Naive pooling, which represents the best-case upper bound on performance in the presence of a constant mortality rate, results in similar or reduced type-I error rates compared to those observed in PREVAIL II with increases in power of 34% to 76% across all segments.

Operating characteristics are summarized visually in the left panel of Figure 3.2, where open triangles represent the results for the PREVAIL II design, closed shapes represent approaches incorporating supplemental information, and the different colors identify the segment. In the presence of a constant mortality rate, all 8 MEM-based designs have increased power compared to PREVAIL II while 6 also have lower average type-1 error rates.

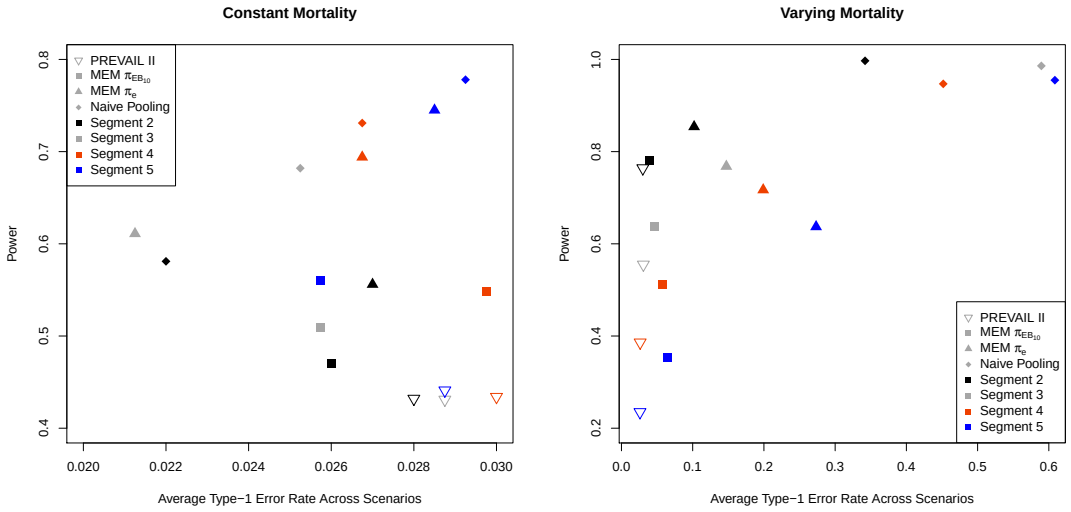

Figure 3.2: Plots demonstrating power versus average type-1 error rate across scenarios for segments 2-5 for PREVAIL II, MEMs with $\pi_{\text{EB}_{10}}$ and π_e priors, and the naive pooling case.

While a moderate RR=0.7 results in minimal early termination, as observed by the average sample size estimates near 1000 for each overall trial, AR balances the information available for evaluating the control and experimental arms, with the positive byproduct of more patients receiving a potentially beneficial treatment in the MEM-based designs than the PREVAIL design. Across segments, we observed an absolute maximum increase of 15.5% and 29.7% in the proportions assigned to the treatment arm for $\pi_{\text{EB}_{10}}$ and π_e, respectively. This also corresponds to increases in the proportion surviving within non-null segments for the MEM-based designs compared to the PREVAIL II design with improvements observed for both $\pi_{\text{EB}_{10}}$ and π_e. Figure 3.3(a) presents the proportion randomized to the treatment group and Figure 3.3(b) presents the proportion surviving by segment for all designs in the constant mortality scenario. π_e and naive pooling randomize a similar number to the control group, while $\pi_{\text{EB}_{10}}$ is more conservative. The MEM-based designs and naive pooling show clear increases in the median proportion surviving in each non-null scenario compared to the PREVAIL II design.

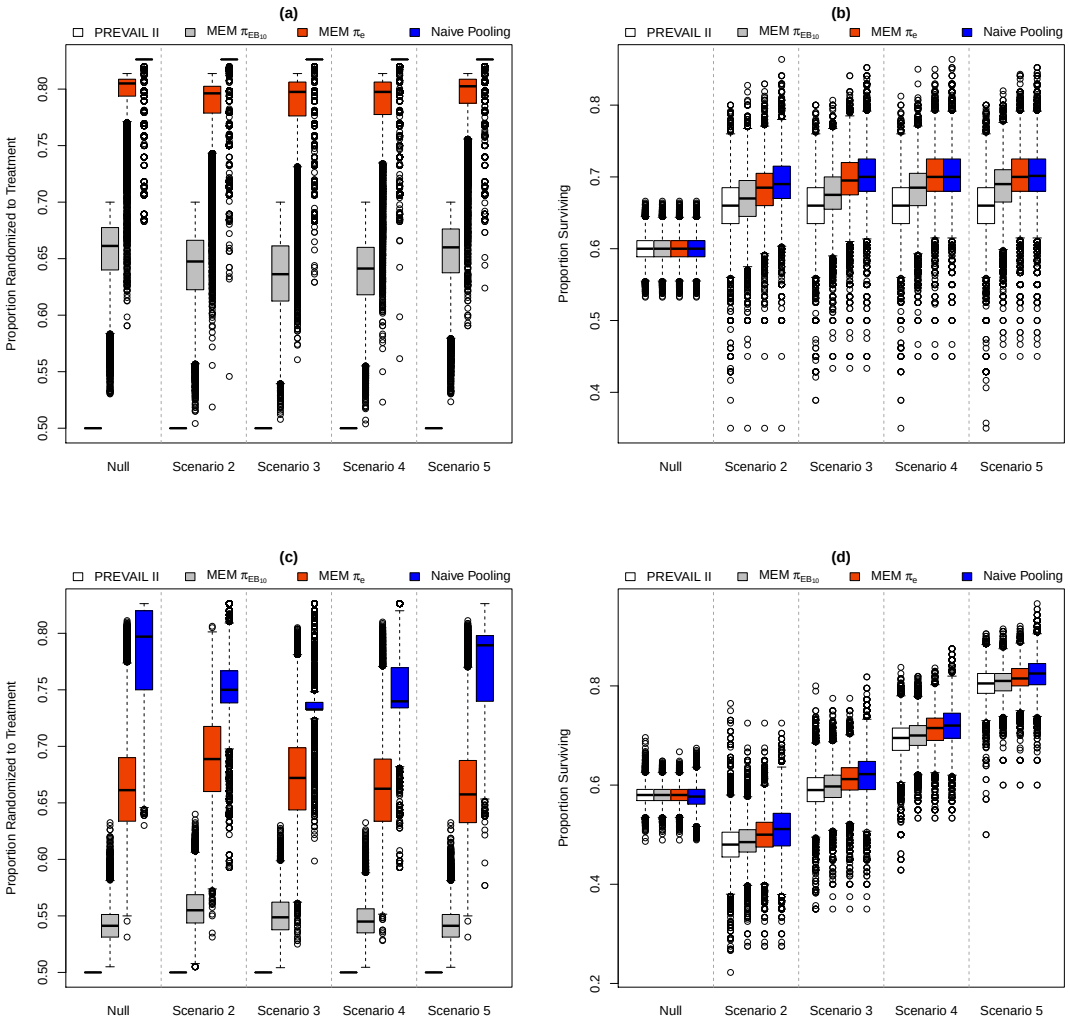

Figure 3.3: Proportion assigned to treatment arm across segments 2-5 (left) and proportion surviving across segments 2-5 (Null scenario) or within the non-null segment (Scenarios 2-5) (right) under the constant mortality scenario (top) and the varying mortality scenario (bottom) for the PREVAIL II design, MEM-based designs, and naive pooling.

Table 3.4 presents simulation results for the varying mortality scenario using the thresholds calibrated for the constant mortality scenario. Recall that this scenario is motivated by the mortality rate observed in Sierra Leone during the West Africa Ebola outbreak and is more challenging than the constant mortality scenario. The PREVAIL II design

maintains a type-I error rate between approximately 0.025 to 0.03 since no supplemental information is incorporated across segments, but the power steadily decreases with each segment as the underlying mortality rate continues to drop and the absolute difference in the mortality rate due to an effective treatment shrinks. Figure 3.2 clearly identifies that the MEM-based design with π_e (filled-in triangles) and naive pooling (filled-in diamonds) result in drastic inflation to the average type-I error rates, which negate any benefit due to increased power compared to the PREVAIL II design. However, the more conservative $\pi_{\text{EB}_{10}}$ prior (filled-in squares) demonstrates a more acceptable trade-off, where the largest inflation of the type-I error rate occurs in segment 5, increasing from 0.026 for the PREVAIL II design (open triangles) to 0.064 for $\pi_{\text{EB}_{10}}$, while power increases from 2% to 51% across non-null segments. Figures 3.3 (c) and (d) demonstrate similar operating characteristics to the constant mortality scenario, where the MEM-based designs and naive pooling assign a higher proportion of subjects to the treatment arm with increases in the proportion surviving the non-null segments as compared to the PREVAIL II design.

When considering Tables 3.3 and 3.4 together, the MEM-based design with $\pi_{\text{EB}_{10}}$ offers perhaps the best performance when taking into account all operating characteristics and trial properties. While there is an inflation of the type-I error rate in the varying mortality case, we believe it is within acceptable levels when you consider the context of a rapidly developing disease outbreak and that there are clear improvements to power, the proportion randomized to the treatment arms, and the proportion surviving within non-null segments compared to the PREVAIL II design.

Section B of the Supplementary Materials provides additional simulation results. Section B.1 provides results when thresholds are calibrated to control the type-I error rates in the varying mortality scenario while maintaining the power for the PREVAIL II design in the constant mortality scenario. Section B.2 provides results for scenarios with RR=0.5. Section B.3 provides results with different parameter values for n_{burn}, $\mathcal{B}$, and c. Section B.4 provides results with two effective therapeutics.

		Probability Reject in Segment					Mean (sd)	Mean (sd)	Mean (sd)
	Trt	1	2	3	4	5	N	Prop Trt	Prop Surv
P-II	S0	0.032	0.028	0.029	0.031	0.029	996 (25.58)	0.5 (0)	0.600 (0.017)
	S2	0.032	0.432	0.028	0.029	0.028	988 (38.41)	0.5 (0)	0.659 (0.036)
	S3	0.032	0.028	0.431	0.029	0.030	988 (38.89)	0.5 (0)	0.659 (0.036)
	S4	0.032	0.028	0.029	0.434	0.028	988 (38.79)	0.5 (0)	0.659 (0.036)
	S5	0.032	0.028	0.029	0.031	0.441	988 (38.78)	0.5 (0)	0.659 (0.036)
$\pi_{\text{EB}_{10}}$	S0	0.027	0.026	0.026	0.030	0.026	998 (14.97)	0.655 (0.029)	0.600 (0.017)
	S2	0.027	0.470	0.025	0.028	0.024	990 (32.03)	0.642 (0.032)	0.669 (0.035)
	S3	0.027	0.026	0.509	0.031	0.025	990 (31.49)	0.634 (0.035)	0.676 (0.035)
	S4	0.027	0.026	0.026	0.548	0.028	990 (31.09)	0.638 (0.032)	0.682 (0.035)
	S5	0.027	0.026	0.026	0.030	0.560	991 (30.96)	0.654 (0.029)	0.686 (0.036)
π_e	S0	0.027	0.027	0.026	0.033	0.037	998 (14.63)	0.797 (0.021)	0.600 (0.017)
	S2	0.027	0.556	0.017	0.020	0.026	990 (32.08)	0.785 (0.032)	0.683 (0.035)
	S3	0.027	0.027	0.611	0.021	0.024	989 (32.85)	0.782 (0.038)	0.696 (0.035)
	S4	0.027	0.027	0.026	0.694	0.027	988 (33.67)	0.784 (0.034)	0.701 (0.035)
	S5	0.027	0.027	0.026	0.033	0.745	987 (34.80)	0.794 (0.023)	0.701 (0.035)
POOL	S0	0.027	0.022	0.030	0.036	0.040	998 (15.38)	0.825 (0.009)	0.600 (0.017)
	S2	0.027	0.581	0.011	0.017	0.025	984 (37.24)	0.817 (0.024)	0.691 (0.036)
	S3	0.027	0.022	0.682	0.018	0.024	982 (38.71)	0.814 (0.030)	0.701 (0.036)
	S4	0.027	0.022	0.030	0.731	0.028	981 (38.51)	0.814 (0.028)	0.702 (0.036)
	S5	0.027	0.022	0.030	0.036	0.778	981 (38.75)	0.821 (0.013)	0.702 (0.036)

Table 3.3: Operating characteristics and trial properties for the utilized platform design as well as alternative adaptive platform designs. 25,000 simulations for the constant underlying mortality case ($p = 0.4$ for all segments) with RR=0.7 for non-null segments for the PREVAIL II (P-II) master protocol; MEMs incorporating adaptive randomization with the constrained empirical Bayes, $c = .10$ prior ($\pi_{\text{EB}_{10}}$) and the fully Bayesian uniform prior (π_e); and the naive pooling (POOL) of all supplemental information incorporating adaptive randomization using posterior probability thresholds optimized for the constant mortality case. Results provided for power/type-I error for each segment, average (sd) total sample size (N) across entire trial, average (sd) proportion allocated to treatment arm in segments 2-5, and average (sd) proportion surviving in the non-null segments (for Trt=S2-S5) or across segments 2-5 (for Trt=S0).

		Probability Reject in Segment					Mean (sd)	Mean (sd)	Mean (sd)
Prior	Trt	1	2	3	4	5	N	Prop Trt	Prop Surv
	S0	0.028	0.030	0.032	0.027	0.025	997 (22.73)	0.5 (0)	0.580 (0.017)
	S2	0.028	0.764	0.027	0.025	0.027	972 (51.38)	0.5 (0)	0.482 (0.041)
P-II	S3	0.028	0.030	0.555	0.026	0.026	984 (42.33)	0.5 (0)	0.591 (0.038)
	S4	0.028	0.030	0.032	0.386	0.026	990 (35.35)	0.5 (0)	0.693 (0.035)
	S5	0.028	0.030	0.032	0.027	0.235	994 (27.80)	0.5 (0)	0.804 (0.029)
	S0	0.027	0.040	0.048	0.058	0.061	998 (16.20)	0.543 (0.016)	0.580 (0.017)
	S2	0.027	0.781	0.042	0.063	0.070	971 (47.98)	0.557 (0.019)	0.487 (0.038)
$\pi_{\text{EB}_{10}}$	S3	0.027	0.040	0.637	0.052	0.069	983 (39.13)	0.551 (0.018)	0.597 (0.036)
	S4	0.027	0.040	0.048	0.511	0.057	990 (31.19)	0.546 (0.016)	0.699 (0.032)
	S5	0.027	0.040	0.048	0.058	0.354	994 (24.99)	0.542 (0.016)	0.807 (0.027)
	S0	0.027	0.102	0.165	0.213	0.262	995 (23.33)	0.665 (0.044)	0.580 (0.017)
	S2	0.027	0.854	0.094	0.236	0.326	961 (52.95)	0.689 (0.042)	0.500 (0.038)
π_e	S3	0.027	0.102	0.768	0.134	0.307	972 (46.86)	0.673 (0.042)	0.611 (0.036)
	S4	0.027	0.102	0.165	0.717	0.198	980 (41.23)	0.665 (0.043)	0.712 (0.032)
	S5	0.027	0.102	0.165	0.213	0.637	986 (36.15)	0.663 (0.043)	0.816 (0.027)
	S0	0.027	0.342	0.729	0.561	0.736	926 (54.93)	0.785 (0.038)	0.576 (0.023)
	S2	0.027	0.997	0.171	0.486	0.674	878 (48.89)	0.749 (0.034)	0.510 (0.047)
POOL	S3	0.027	0.342	0.986	0.199	0.648	884 (39.12)	0.740 (0.029)	0.621 (0.047)
	S4	0.027	0.342	0.729	0.947	0.376	893 (44.84)	0.752 (0.036)	0.720 (0.039)
	S5	0.027	0.342	0.729	0.561	0.955	884 (56.21)	0.773 (0.039)	0.824 (0.033)

Table 3.4: Operating characteristics and trial properties for the utilized platform design as well as alternative adaptive platform designs. 25,000 simulations for the varying underlying mortality case ($p = (0.74, 0.61, 0.48, 0.36, 0.23)$ for segments 1-5, respectively) with RR=0.7 for non-null segments for the PREVAIL II (P-II) master protocol; MEMs incorporating adaptive randomization with the constrained empirical Bayes, $c = .10$ prior ($\pi_{\text{EB}_{10}}$) and the fully Bayesian uniform prior (π_e); and the naive pooling (POOL) of all supplemental information incorporating adaptive randomization using posterior probability thresholds optimized for the constant mortality case. Results provided for power/type-I error for each segment, average (sd) total sample size (N) across entire trial, average (sd) proportion allocated to treatment arm in segments 2-5, and average (sd) proportion surviving in the non-null segments (for Trt=S2-S5) or across segments 2-5 (for Trt=S0).

3.5 Discussion

In the context of rapidly developing disease outbreaks, there is a need for flexible, dynamic methods to identify effective therapeutics as quickly as possible. In the EVD outbreak from 2014–16, the PREVAIL II study was designed to test multiple therapies within one overarching trial while incorporating aggressive interim monitoring to identify effective treatments as early as possible. However, the PREVAIL II trial design did not utilize previous segments of the control arm to improve the efficiency of the analyses and is perhaps suboptimal given the availability of adaptive design features and Bayesian hierarchical modeling techniques. To address this shortcoming we proposed incorporating MEMs with AR to estimate the extent of exchangeability across segments administering identical treatment regimes and balance allocation in relation to ESSS. Note that the modification to the design is general and can be applied with other methods for incorporating supplemental information.

There are many approaches to AR which could have been incorporated to our proposed design. For example, Berry and Eick (1995) compare four different response adaptive approaches to a standard equal randomization scheme and identify improvements using AR in many scenarios. More recently, Thall and Wathen (2007) explored the use of a positive constraint on Bayesian AR in order to limit the extent of adjusting the allocation ratio. Methods also exist which use biomarker information to adaptively randomize individuals during a study to more advantageous trial arms based on an individual's biomarker profile (Zhou et al., 2008) or to adjust patient allocation in trials with binary outcomes to address covariate imbalances such that more patients can access the superior treatments identified in the study (Ning and Huang, 2010). Other methods utilize predictive probability and Bayesian AR to treat more patients with the more efficacious treatment while enabling early termination if superiority or equivalence can be demonstrated before trial completion (Yin et al., 2012).

However, outcome-AR techniques, which can lead to an imbalance in the sample size or poor operating characteristics, are controversial (Thall et al., 2015). Our proposed design utilizes AR to target information balance, which results in additional allocation to experimental therapies in the presence of exchangeable information contributed by supplemental controls. In the context of the EVD outbreak, where concerns were raised about

the appropriateness of randomized trials due to ethical or practical concerns (Adebamowo et al., 2014; Ippolito et al., 2016), the modified multi-source adaptive platform design offers perhaps an ideal trade-off: controlling for cohort bias with the potential to randomize more study participants to emerging therapies. The methods presented herein represent a useful tool for designing platform trials to address future infectious disease outbreaks, such as the Zika virus or future Ebola outbreaks.

It can be noted that 1:1 allocation is maintained in the absence of evidence supporting exchangeable controls. Further, while AR has historically been shown to result in potentially undesirable operating characteristics, MEMs with AR under $\pi_{\text{EB}_{10}}$ maintained type-I error in the constant mortality scenario with reasonable inflation in the varying underlying mortality case, while increasing the power to detect an effective drug in both scenarios. This approach to AR addresses the concerns of sample size imbalance and potentially undesirable operating characteristics raised by Thall et al. (2015).

Even though the simulations and designs presented in this chapter are unique to the EVD outbreak, they can be generalized to other settings of clinical study, such as screening platform for intermediate-phased drug trials as well as sequential experimental designs of biomarker assays. In addition, a number of other parameters can be adjusted (e.g., priors on the MEM weights, the components of the adaptive randomization, and the frequency of interim monitoring) to achieve the desired operating characteristics in other settings where incorporating supplemental control information is desired.

The number of different parameters, while demonstrating the flexibility of our design, also represents a potential limitation. It may be challenging to identify the most appropriate parameters to use given the many unknowns of a rapidly developing disease outbreak. However, this can be moderated by assuming conservative priors on the MEM weights, such as $c = 0.1$ for π_{EB_c}, which ensures that posterior model weight is given to the MEM which assumes no exchangeable segments while incorporating information in a manner that results in an improvement to the power, survivorship, and proportion randomized to the treatment arm. The value chosen for c reflects a boundary imposed on the probability of exchangeability and should be evaluated in the context of each trial. One should explore values between 0 and 1 to identify values of c with better performance, but in our experience it should be set at a low value in order to moderate the influence of π_{EB} which assigns

all prior weight to a single source, and as a result, posterior, weight to a single MEM. Further, while calibration may be challenging, the proposed π_{EB_c} prior is based on a single hyperparameter, a benefit compared to Bayesian non-parametric density estimation with finite- or infinite-mixtures or the traditional BMA framework which requires sets of priors for all 2^K models.

Another potential limitation is that simulation results presented in this chapter only consider the scenario with one effective treatment across all segments. However, in the context of a rapidly emerging infectious disease, it is unlikely that an effective treatment will be present in all segments, with many therapies potentially nominated for inclusion based on pharmacokinetic studies, laboratory research, or hypothetical causal drug effects by simulation rather than more rigorous clinical trials. Results for cases with two out of five effective treatments are presented in Section B.4 of the Supplementary Materials and have similarly encouraging operating characteristics and trial properties as presented in Section 3.4.2.

Chapter 4

A Fully Bayesian Mixture Model Approach for Identifying Non-Compliance in a Regulatory Tobacco Clinical Trial

4.1 Introduction

The Center for the Evaluation of Nicotine in Cigarettes, project 1 (CENIC-p1), was a 6-week randomized multi-center trial designed to evaluate the effect of nicotine reduction on tobacco use behavior (Donny et al., 2015). 839 current smokers underwent randomization, with 780 completing the 6-week study after being equally randomized to one of seven groups, including a usual brand control condition and one of six experimental cigarettes with nicotine content ranging from 15.8 mg per gram of tobacco (normal nicotine controls) to 0.4 mg per gram of tobacco (very low nicotine content (VLNC) cigarettes). At the end of six weeks, participants randomly assigned to the lowest nicotine condition had significantly reduced tobacco use, dependence, and nicotine exposure relative to the normal nicotine controls.

In 2009, Congress passed the Family Smoking Prevention and Tobacco Control Act (FSPTCA), which gave the Food and Drug Administration (FDA) the authority to regulate

the content, marketing, and sale of tobacco products. In particular, the FSPTCA gives the FDA the authority to reduce the nicotine content of cigarettes to non-addictive levels (but not zero) if it can be shown that this will improve public health. CENIC-p1 provides key scientific evidence in support of a new nicotine standard for cigarettes, but understanding the impact of mandated nicotine reduction on tobacco use behavior is difficult due to the presence of substantial non-compliance to randomized treatment assignment.

In CENIC-p1, non-compliance is defined as smoking any commercially available, non-study cigarettes in place of, or in addition to, the study cigarettes provided by the trial. Identifying non-compliance in CENIC-p1 is difficult because the most direct measure of non-compliance, self-reported smoking of non-study cigarettes, is known to be unreliable (Nardone et al., 2016). An alternative strategy is to use biomarkers of nicotine exposure (i.e., cotinine or total nicotine equivalents (TNEs)) to identify non-compliance among self-reported compliers (Benowitz et al., 2015). For example, a recent study characterized the distribution of biomarkers of nicotine exposure in participants who were sequestered in a hotel for five days to ensure that they only smoked cigarettes with 0.4 mg of nicotine per gram of tobacco (Denlinger et al., 2016). In this study, the 95th percentile for TNEs in fully compliant subjects was estimated at 6.41 nmol/ml, which has been used as a threshold for identifying non-compliance in secondary analyses of CENIC-p1 (Tidey et al., 2017; Rupprecht et al., 2017). Alternatively, Boatman et al. (2017) used a mixture model approach, which incorporated the data from Denlinger et al. (2016), to estimate the probability that a subject randomized to the VLNC groups was compliant conditional on their biomarker values, which facilitated estimation of the causal effect of nicotine reduction on cigarettes smoked per day.

While the results of Denlinger et al. (2016) provide thresholds for identifying non-compliance in the VLNC group, they can not be used to identify non-compliance at the intermediate dose-levels (i.e., nicotine dose-levels between 0.4 mg/g and 15.8 mg/g). The mixture model approach proposed by Boatman et al. (2017) could be used to identify non-compliance at the intermediate dose-levels but fitting mixture models is known to be challenging due to label switching or overlap between the mixture components (Marin et al., 2005; Jasra et al., 2005; Stephens, 2000). Alternatively, we could leverage our understanding of the biological relationship between the nicotine content of cigarettes and

biomarkers of nicotine exposure to better estimate the mixture components (Benowitz et al., 2015). We expect the nicotine content of the cigarettes to have an additive effect on biomarkers of nicotine exposure and we could specify a model that incorporates this relationship, in which case the mean of the compliant mixture component is estimated relative to the control or usual brand groups. This could improve estimation of the mixture components, but is also susceptible to model misspecification.

Having two different modeling frameworks represents a challenge given the uncertainty as to which model is most appropriate for the data. The Bayesian framework is naturally able to account for the uncertainty around selecting the "best" approach by considering multiple models simultaneously and averaging the results, thus avoiding the need to select the "best" model. For example, the posterior estimates for each model can be estimated using Markov chain Monte Carlo (MCMC) techniques which account for the different assumptions of each approach, such as the proposed biological relationship or various parameter constraints, and model averaging can smooth over the posteriors from each model. However, model averaging in our context is made more difficult because the dimensionality of the parameter space differs between the two proposed approaches. The reversible-jump MCMC (RJMCMC) algorithm addresses this difficulty by providing the flexibility to account for the different dimensionalities while implementing model averaging within each dose-level (Green, 1995). This allows us to "mix and match" which approach is most appropriate for each dose-level so that any gains in efficiency from the biological relationship can be used within the applicable dose-levels while avoiding the introduction of bias when it is inappropriate.

The remainder of this chapter proceeds as follows. In Section 4.2 we define finite Gaussian mixture models with two components (compliant and non-compliant) and discuss how our model specification can be adapted to account for the hypothesized relationship between the nicotine content of cigarettes and biomarkers of nicotine exposure. A RJMCMC algorithm for averaging over various models is described in Section 4.2.2. Simulation results evaluating our modeling approach when the hypothesized relationship does and does not hold are presented in Section 4.3 and we apply our proposed method to CENIC-p1 in Section 4.4. Finally, we conclude with a brief discussion in Section 4.5.

4.2 Methods for estimating compliance within randomized groups

We first introduce the notation that will be used throughout the remainder of this chapter. Let $i = 1, ..., n$ be an index for each observation, $j = 1, ..., J$ index the different randomized groups, and $k = 1, 2$ represent the two components of the mixture distribution within a level j, where $k = 1$ implies compliance and $k = 2$ implies non-compliance. Further, let $\boldsymbol{y}$ represent the observed outcome, such that y_{ij} represents participant i who belongs to randomization group j.

The goal of this Chapter is to identify non-compliance to randomized treatment assignment. As discussed previously, two approaches to this problem utilizing biomarkers of nicotine exposure have been proposed in the literature. Denlinger et al. (2016) identified quantiles of the biomarker distribution (e.g., 90th, 95th, etc.) that can be used as thresholds for identifying non-compliant individuals, while Boatman et al. (2017) estimated the probability of individual compliance conditional on the observed biomarkers through an application of Bayes theorem. Regardless of which approach is used, both require at least estimating the distribution of the biomarkers in fully compliant subjects. This requires estimation of a mixture distribution because we do not know which subjects were compliant, which is challenging in the absence of auxiliary data from known compliers.

We first specify a two component mixture distribution for an arbitrary treatment group j, assuming no dose-response relationship across dose-levels, which we will label the "IND model". For group j, let μ_{j1} be the mean of the compliant component ($k = 1$), μ_{j2} be the mean of the non-compliant component ($k = 2$), τ_j be the common precision of the Gaussian distributions for compliant and non-compliant subjects, and p_j be the mixture weight, which represents the proportion of non-compliant subjects in group j. This results in the following likelihood for group j:

$$p(\boldsymbol{y}|\mu_{j1}, \mu_{j2}, \tau_j, p_j) = \prod_{i=1}^{n} \left[(1 - p_j)\mathcal{N}(y_i|\mu_{j1}, \tau_j) + p_j\mathcal{N}(y_i|\mu_{j2}, \tau_j)\right], \tag{4.1}$$

where $\mathcal{N}$ is a Gaussian density with the given mean and precision. While intuitive, this form of the likelihood is susceptible to the issue of label-switching (i.e., the peaks of the posterior

distribution for the two components switch such that μ_{j1} is actually estimating the mean for the second component and μ_{j2} is estimating the first component mean) and makes deriving the conditional posterior distributions under conjugate priors very challenging.

To address these issues, the likelihood can be reparameterized as follows. First, to address the potential for label-switching, redefine μ_{j1} as μ_j and μ_{j2} as $\mu_j + \theta_j$, where $\theta_j > 0$. This implies that $\mu_j < \mu_j + \theta_j$ and ensures that the non-compliant component mean will be larger than the compliant component mean. Second, to address the challenges in deriving the conditional posteriors, we introduce a latent indicator variable $\boldsymbol{z}$ through data augmentation (Chib, 1996). Define $z_{ijk} = 1$ when y_i belongs to both group j and component k, and 0 otherwise. These changes induce a complete data likelihood considering both our known $\boldsymbol{y}$ and unknown $\boldsymbol{z}$ of

$$p(\boldsymbol{y}, \boldsymbol{z}|\mu_j, \theta_j, \tau_j, p_j) = \prod_{i=1}^{n} [(1 - p_j) \times \mathcal{N}(y_i|\mu_j, \tau_j)]^{z_{ij1}} [p_j \times \mathcal{N}(y_i|\mu_j + \theta_j, \tau_j)]^{z_{ij2}}. \quad (4.2)$$

In the setting where $J = 1$, the notation can be simplified by replacing $z_{ij2} = 1 - z_{ij1}$, however with multiple j this substitution implies that $1 - z_{ij1} = 1$ not only when participant i in group j belongs to $k = 2$, as desired, but also when a participant does not belong to group j. Extending (4.2) to multiple j, let $\Theta = (\boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\tau})$, where the bold notation represents vectors including the corresponding parameter for each of the j groups. The complete data likelihood for all subjects can be written as:

$$p(\boldsymbol{y}, \boldsymbol{z}|\Theta) = \prod_{i=1}^{n} \left\{ \prod_{j=1}^{J} [(1 - p_j)\mathcal{N}(y_i|\mu_j, \tau_j)]^{z_{ij1}} [p_j \mathcal{N}(y_i|\mu_j + \theta_j, \tau_j)]^{z_{ij2}} \right\}. \quad (4.3)$$

The complete data likelihood can be modified to incorporate the hypothesized biological relationship discussed earlier, which we refer to as the "REL model" for the remainder of the Chapter. First, within CENIC-p1, assume that group J represents the combined usual brand control and 15.8 mg/g groups, where all participants are assumed to be compliant. Then assume that some relationship exists between μ_j and μ_J that can be defined by an arbitrary function h: $\mu_j^* = h_j(\mu_J)$. In this case we no longer estimate each μ_j, $j \neq J$, but only μ_J. Estimates for each μ_j are based on the proposed relationship, $h_j(\mu_J)$. Note, data are now utilized across all levels to estimate μ_J, but estimates of θ_j are derived within each

level to account for differences in the mean for the non-compliers across dose-levels due to differences in the prevalence of partial compliance or smoking behavior across dose-levels.

Specifically in the context of CENIC-p1 for the log-normally distributed TNE biomarker, we will consider the following proposed relationship. Letting w_j represent the nicotine content for dose level j, we propose the following function, h_j, that maps the ratio of the nicotine contents to the mean of the biomarker distribution:

$$\mu_j^* = h_j(\mu_J) = \log\left(\frac{w_j}{w_J}\right) + \mu_J. \tag{4.4}$$

Biomarkers of nicotine exposure are thought to be linearly related to the nicotine content of the cigarettes and the proposed relationship imposes a linear relationship between the nicotine content of the cigarettes and the median on the original scale. Given this relationship, the complete data likelihood is very similar, but with μ_j^* in place of μ_j:

$$p(\boldsymbol{y}, \boldsymbol{z}|\Theta) = \prod_{i=1}^{n}\left\{\prod_{j=1}^{J}[(1-p_j)\mathcal{N}(y_i|\mu_j^*, \tau_j)]^{z_{ij1}}[p_j\mathcal{N}(y_i|\mu_j^* + \theta_j, \tau_j)^{z_{ij2}}]\right\}. \tag{4.5}$$

When considering all j levels, it may be the case that some levels are better estimated with the IND approach and others the REL approach. In this case, we can also average over intermediate models where some dose-levels are estimated assuming the IND model, while others follow the REL model. For example, within CENIC-p1 there are a total of 16 potential models, $m = 1, ..., 16$, considering all the possible combinations of modeling frameworks assumed for the four non-control groups where the state of the model can be denoted by a vector, $\boldsymbol{\xi}_m$, where 0 indicates the IND model and 1 the REL model. For example, $\boldsymbol{\xi}_m = (1, 0, 0, 0)$ indicates that group 1 is estimated from the REL model, while groups 2-4 are estimated from the IND model. The RJMCMC algorithm can traverse the model space of 16 models with different parameter dimensionality to average over all possible combinations.

4.2.1 Prior distributions and posterior inference

In this section, we discuss prior distributions and posterior inference for the IND and REL approaches. In general, the posterior will not be available in closed form and must be

approximated using MCMC.

For treatment group j, we assume the following prior distributions for the various model parameters:

$$\begin{aligned}\mu_j \sim& \mathcal{N}(0, s_{\mu j}),\\ \theta_j \sim& \mathcal{N}(0, s_{\theta j}),\\ \tau_j \sim& \Gamma(a_j, b_j),\\ p_j \sim& \text{Beta}(\alpha_j, \beta_j).\end{aligned} \tag{4.6}$$

Letting n_{jk} represent the number of observations in component k of group j and $n_j = \sum_{k=1}^{K} n_{jk}$, the joint posterior for the IND model is proportional to the following:

$$\begin{aligned}p(\Theta|\boldsymbol{y}, \boldsymbol{z}) \propto \prod_{j=1}^{J} \Bigg\{ &(1-p_j)^{n_{j1}+\beta_j-1} p_j^{n_{j2}+\alpha_j-1} \times \tau_j^{\frac{n_j}{2}+a_j-1} \exp\big(-b_j\tau_j\big) \times \exp\Big(-\frac{s_{\mu j}}{2}\mu_j^2\Big)\\ &\times \exp\Big(-\frac{s_{\theta j}}{2}\theta_j^2\Big) \times \exp\left(-\frac{\tau_j}{2}\left[\sum_{i=1}^{n} z_{ij1}(y_i-\mu_j)^2 + z_{ij2}(y_i-[\mu_j+\theta_j])^2\right]\right)\Bigg\}.\end{aligned} \tag{4.7}$$

From (4.7), the conditional posterior for each parameter in the IND model within group j is:

$$p(\mu_j|\theta_j, \tau_j, \boldsymbol{y}, \boldsymbol{z}) \propto \mathcal{N}\left(\frac{\tau_j(n_{j1}\bar{y}_{j1} + n_{j2}(\bar{y}_{j2}-\theta_j))}{n_j\tau_j + s_{\mu j}}, n_j\tau_j + s_{\mu j}\right), \tag{4.8}$$

$$p(\theta_j|\mu_j, \tau_j, \boldsymbol{y}, \boldsymbol{z}) \propto \mathcal{N}\left(\frac{n_{j2}\tau_j(\bar{y}_{j2}-\mu_j)}{n_{j2}\tau_j + s_{\theta j}}, n_{j2}\tau_j + s_{\theta j}\right), \tag{4.9}$$

$$\begin{aligned}p(\tau_j|\mu_j, \theta_j, \boldsymbol{y}, \boldsymbol{z}) \propto \Gamma\Big(&\frac{n_j}{2} + a_j, \frac{1}{2}\big[(n_{j1}-1)s_{j1}^2 + (n_{j2}-1)s_{j2}^2 +\\ &n_{j1}(\bar{y}_{j1}-\mu_j)^2 + n_{j2}(\bar{y}_{j2}-(\mu_j+\theta_j))^2 + 2b_j\big]\Big),\end{aligned} \tag{4.10}$$

$$p(p_j|\mathbf{y}, \mathbf{z}) \propto \text{Beta}(n_{j2}+\alpha_j, n_{j1}+\beta_j), \tag{4.11}$$

$$p(z_{ij2}=1|y_{ij}, \Theta) \propto \text{Ber}\left(\frac{p_j\mathcal{N}(y_{ij}|\mu_j+\theta_j, \tau_j)}{(1-p_j)\mathcal{N}(y_{ij}|\mu_j, \tau_j) + p_j\mathcal{N}(y_{ij}|\mu_j+\theta_j, \tau_j)}\right), \tag{4.12}$$

where $\bar{y}_{jk}$ is the mean of y_{ijk} for group j and component k.

The previous derivation can be easily extended to the REL model where $h_j(\mu_J)$ is

specified as in (4.4). The conditional posteriors for the REL model, with some additional notation, where $\mu_j^* = h_j(\mu_J)$ is

$$d_j = \log\left(\frac{w_j}{w_J}\right), \tag{4.13}$$

$$g_j = \tau_j(n_j d_j + n_{j2}\theta_j - n_{j1}\bar{y}_{j1} - n_{j2}\bar{y}_{j2}), \tag{4.14}$$

$$p(\mu_J|\Theta, \boldsymbol{y}, \boldsymbol{z}) \propto \mathcal{N}\left(\frac{n_J\tau_J\bar{y}_J - \sum_{j=1}^{J-1} g_j}{\sum_{j=1}^{J}(n_j\tau_j) + s_{\mu J}}, \sum_{j=1}^{J}(n_j\tau_j) + s_{\mu J}\right), \tag{4.15}$$

$$p(\theta_j|\mu_j, \tau_j, \boldsymbol{y}, \boldsymbol{z}) \propto \mathcal{N}\left(\frac{n_{j2}\tau_j(\bar{y}_{j2} - \mu_j^*)}{n_{j2}\tau_j + s_{\theta j}}, n_{j2}\tau_j + s_{\theta j}\right), \tag{4.16}$$

$$\begin{aligned} p(\tau_j|\mu_j, \theta_j, \boldsymbol{y}, \boldsymbol{z}) \propto \Gamma\Big(&\frac{n_j}{2} + a_j, \frac{1}{2}\big[(n_{j1}-1)s_{j1}^2 + (n_{j2}-1)s_{j2}^2 + \\ &n_{j1}(\bar{y}_{j1} - \mu_j^*)^2 + n_{j2}(\bar{y}_{j2} - (\mu_j^* + \theta_j))^2 + 2b_j\big]\Big), \end{aligned} \tag{4.17}$$

$$p(p_j|\mathbf{y}, \mathbf{z}) \propto \text{Beta}(n_{j2} + \alpha_j, n_{j1} + \beta_j), \tag{4.18}$$

$$p(z_{ij2} = 1|y_{ij}, \Theta) \propto \text{Ber}\left(\frac{p_j\mathcal{N}(y_{ij}|\mu_j^* + \theta_j, \tau_j)}{(1-p_j)\mathcal{N}(y_{ij}|\mu_j^*, \tau_j) + p_j\mathcal{N}(y_{ij}|\mu_j^* + \theta_j, \tau_j)}\right). \tag{4.19}$$

Note that the previously derived conditional posteriors for $\boldsymbol{\theta}$, $\boldsymbol{p}$, $\boldsymbol{\tau}$, and $\boldsymbol{z}$ are unchanged from the IND model, except that estimates for μ_j are derived from the proposed relationship.

4.2.2 Model averaging with RJMCMC

We now discuss how we will average over the IND, REL, and intermediate models using RJMCMC. RJMCMC accounts for the varying dimensionality of the parameter space during model averaging. For example, the IND model estimates μ_j from the group j observed data only, but in the REL model μ_j is derived from the proposed relationship with μ_J, representing a change in the dimensionality of the parameter space due to the assumed relationship between μ_j and μ_J.

As noted above, we will consider intermediate models, in addition to the IND and REL models, to account for variability in the appropriateness of the hypothesized relationship between nicotine content and biomarkers of nicotine exposure across dose-levels. That is, the hypothesized relationship may be appropriate for some dose levels but not others. For

example, if $\boldsymbol{\xi}_m = (1, 0, 0, 1)$, μ_J is updated using *only* groups $j = 1, 4$ and the control group $j = 5$, while groups $j = 2$ and 3, are not included because only data from their own group are used to derive parameter estimates.

While described in greater detail below, the general steps of the RJMCMC algorithm are:

1. Update the parameters conditional on the current model using the Gibbs sampler discussed in Section 4.2.1.

2. Update the model, m, conditional on the current parameter values using the RJMCMC algorithm which includes the following steps:

   2.a Randomly select a group j' from $j = 1, ..., J-1$ with equal probability as a candidate to change states from REL to IND or vice versa.

   2.b Accept this proposed move with some probability.

We introduce additional notation for RJMCMC. Let $q(\cdot)$ be an (arbitrary) proposal density for parameters that must be generated as the result of proposed moves that increase the dimensionality of the parameter space, $f(\boldsymbol{y}|\Theta, m)$ represent the likelihood of our data given Θ in model m, and $p(m)$ represent the prior probability of model m. Let $I_j = 1$ be an indicator that group j is estimated using the IND model, with $I_j = 0$ indicating that group j is estimated using the REL model and define $p(I_j = 1)$ to be the prior probability that $I_j = 1$ with $p(I_j = 0) = 1 - p(I_j = 1)$, the prior probability that $I_j = 0$. We define $p(m)$ as the product of these priors: $\prod_{j=1}^{J-1}[p(I_j = 1)]^{I_j}[1 - p(I_j = 1)]^{1-I_j}$. If $p(I_j = 1) = 0.5$, all models receive equal prior weight, whereas $p(I_j = 1) > 0.5$ favors IND for group j and $p(I_j = 1) < 0.5$ favors REL for group j.

In Step 2a, the intermediate approaches can be conceptualized as a set of nested models of the REL approach where all groups are estimated assuming the relationship, so the proposal scheme suggested by Green and Hastie (2009) for scenarios with nested models is implemented. In the scheme for nested models, at each iteration we propose a move from (Θ, m) to (Θ', m') by randomly selecting $j' \in 1, ..., J-1$ with equal probability as the candidate dose-level to flip from IND to REL or vice versa. The change in the state of j' will increase or decrease the dimension of the parameter space by 1. For example,

if $\boldsymbol{\xi}_m = (1,0,0,0)$ and $j' = 3$, then $\boldsymbol{\xi}_{m'} = (1,0,1,0)$ with group 3 switching from IND to REL and the parameter dimensionality decreases by 1. We note that, while we are considering a RJMCMC algorithm that only considers state-changes one-dose-level-at-a-time, an alternative proposal framework would be to consider jumping to any of the 15 other permutations of the model states for m' rather than flipping one group at each iteration.

In greater detail, if the proposed move changes group j' from the REL model to the IND model (increasing dimensionality), we define the bijective functions (i.e., one-to-one and onto functions that map to only a single value in the range of possible values) for our proposed values to be:

$$\begin{aligned} \boldsymbol{p}' &= \boldsymbol{p}, \\ \boldsymbol{\tau}' &= \boldsymbol{\tau}, \\ \boldsymbol{\theta}' &= \boldsymbol{\theta}, \\ \mu'_{j\neq j'} &= \mu_{j\neq j'}, \\ \mu'_{j'} &= u, \end{aligned} \tag{4.20}$$

where $u \sim q(\cdot)$, such that $q(\cdot)$ is some (arbitrary) proposal distribution for $\mu'_{j'}$. We define $q(\cdot)$ as the corresponding conditional posterior distribution from (4.8).

Conversely, the reverse move for group j' from IND to REL (decreasing dimensionality) is determined completely by the move from REL to IND described above:

$$\begin{aligned} \boldsymbol{p} &= \boldsymbol{p}', \\ \boldsymbol{\tau} &= \boldsymbol{\tau}', \\ \boldsymbol{\theta} &= \boldsymbol{\theta}', \\ \mu_{j\neq j'} &= \mu'_{j\neq j'}, \\ u &= \mu'_{j'}. \end{aligned} \tag{4.21}$$

We note that when decreasing the dimensionality we do not have to simulate from a proposal distribution, but set $\mu_{j'}$ at the value determined by the proposed relationship to μ_J.

In step 2b, the probability of accepting a proposed move is equal to $\min(1, A)$, where:

$$A = \frac{p(\Theta', m'|\boldsymbol{y})P(m|m')q'(\boldsymbol{u}')}{p(\Theta, m|\boldsymbol{y})P(m'|m)q(\boldsymbol{u})} \left|\frac{\partial(\Theta', \boldsymbol{u}')}{\partial(\Theta, \boldsymbol{u})}\right|, \tag{4.22}$$

where $P(m|m')$ denotes the probability of proposing to move from m to m', $\left|\frac{\partial(\Theta', \boldsymbol{u}')}{\partial(\Theta, \boldsymbol{u})}\right|$ is the Jacobian, and $p(\Theta, m|\boldsymbol{y})$ is the joint distribution of Θ and the model m: $p(\Theta_m, m|\boldsymbol{y}, \boldsymbol{z}) \propto p(\boldsymbol{y}, \boldsymbol{z}|\Theta_m, m)p(\Theta_m|m)p(m)$.

In our implementation, each group j has an equal chance of being selected and flipped within each iteration, $P(m'|m) = P(m|m')$. Therefore, the A term simplifies to

$$\begin{aligned} A_{j':\text{REL}\to\text{IND}} =& \frac{p(\Theta', m'|\boldsymbol{y}, \boldsymbol{z})P(m|m')q'(\boldsymbol{u}')}{p(\Theta, m|\boldsymbol{y}, \boldsymbol{z})P(m'|m)q(\boldsymbol{u})} \left|\frac{\partial(\Theta', \boldsymbol{u}')}{\partial(\Theta, \boldsymbol{u})}\right| \\ =& \frac{p(\Theta', m'|\boldsymbol{y}, \boldsymbol{z})}{p(\Theta, m|\boldsymbol{y}, \boldsymbol{z})q(\boldsymbol{u})} \\ =& \frac{f(\boldsymbol{y}, \boldsymbol{z}|\Theta', m')p(\Theta'|m')p(m')}{f(\boldsymbol{y}, \boldsymbol{z}|\Theta, m)p(\Theta|m)p(m)q(u)}. \end{aligned} \tag{4.23}$$

The reverse move from the model where estimation of group j' is changed from IND to REL can be described by the inverse of the stated acceptance probability such that $\min(1, A_{j':\text{IND}\to\text{REL}}) = \min(1, A^{-1}_{j':\text{REL}\to\text{IND}})$.

The final step in 2b is to simulate a single value $v \sim \text{Unif}(0, 1)$ and accept the move to the proposed state if $v \leq \min(1, A)$. If the proposed move is not accepted, the original model m continues to the next iteration with its original parameter values.

4.3 Simulation studies to establish small sample properties

In this section, we present the results of a small simulation study to evaluate the small sample properties of the model averaging approach discussed in Section 4.2.2. Our simulation study will mimic the data collected in CENIC-p1; we consider 5 dose-levels, with dose-level 5 treated as the normal nicotine condition, dose-levels 1-4 treated as varying levels of reduced nicotine content conditions and $n_j = 100$ for all dose-levels. Data for the control condition are drawn from a single component, $\mathcal{N}(3, 1)$. Data for the four treatment conditions are simulated from a two-component mixture distribution with true probability

of non-compliance, p_j, equal to 0.70, which is approximately equal to the estimated probability of non-compliance for the 0.4 mg/g group reported in Nardone et al. (2016) and Boatman et al. (2017). Data for the non-compliant component were drawn from a $\mathcal{N}(3,1)$ distribution for all dose-levels. Data for the compliant components were drawn from a normal distribution with $\tau = 1$ and a mean that varied as a function of dose-level. For the remainder of this section, we specify the mean for the reduced nicotine content groups by the effect size, defined as ES=$\frac{3-\mu_j}{\sigma}$. That is, we specify the effect size, which in turn specifies μ_j. For purposes of estimation, the hypothesized dose-response relationship for the REL model assumes reductions of 98.1%, 95.0%, 86.5%, and 63.2% in the mean of the biomarker for groups 1 through 4 relative to the control condition, corresponding to ES of 4, 3, 2, and 1 for groups 1 through 4, respectively. For all j dose-levels, assume "non-informative" prior specifications of $\alpha_j = \beta_j = 1$ for the beta prior on p_j, $a_j = b_j = 0.001$ for the gamma prior on τ_j, and $s_{\mu_j} = s_{\theta_j} = 0.00001$ for the normal priors on μ_j and θ_j.

We consider six scenarios for the true relationship between the mean in the compliant component and the dose-level. The first scenario assumes that the hypothesized relationship in the REL model is correct, in which case ES=(4,3,2,1) and $\mu_{j1} = (-1, 0, 1, 2)$ for dose-levels 1-4, respectively. The second scenario assumes that dose-levels 1 through 3 follow the hypothesized relationship but dose-level 4 does not: ES=(4,3,2,0.5) and $\mu_{j1} = (-1, 0, 1, 2.5)$. The third scenario assumes that dose-levels 1 and 2 follow the hypothesized relationship but dose-levels 3 and 4 do not: ES=(4,3,1,0.5) and $\mu_{j1} = (-1, 0, 2, 2.5)$. The fourth scenario assumes that only dose-level 1 follows the relationship: ES=(4,1.5,1,0.5) and $\mu_{j1} = (-1, 1.5, 2, 2.5)$. The fifth scenario assumes that the effect size is less than the hypothesized effect size for all dose-levels: ES=(2,1.5,1,0.5) and $\mu_{j1} = (1, 1.5, 2, 2.5)$. Finally, the sixth scenario assumes that the effect size is more than the hypothesized effect size for all dose-levels: ES=(6,4.5,3,1.5) and $\mu_{j1} = (-3, -1.5, 0, 1.5)$.

1,000 simulated studies were completed for each scenario. Simulation results are presented for the IND model, the REL model with our hypothesized relationship, and for our RJMCMC approach that averages over the IND, REL, and intermediate approaches. Two different model priors are chosen for the RJMCMC simulations: assuming each model specification is equally likely (RJ) with $p(I_j = 1) = 0.5$ for each group j, versus favoring models assuming IND for each level (RJ_{95}) with $p(I_j = 1) = 0.95$ for each group j. Within

each MCMC, a chain with 10,000 total iterations was used, with the first 1,000 iterations excluded for the burn-in period. The performance of each model is summarized by bias, standard deviation (SD), and mean square error (MSE) within each dose-level j for the proportion compliant (p_c), the mean of the compliant component (μ_1), the mean of the non-compliant component (μ_2), their shared standard deviation (σ), and the 95th percentile of the distribution of the biomarker on the original scale (i.e., e^y), which is included as this is a quantity of interest for the analysis of the data from CENIC-p1. We note that the SD for the REL, RJ, and RJ_{95} models are presented as a ratio relative to the SD for the IND model in order to illustrate the gain in efficiency due to borrowing. Additionally, the RJMCMC results also provide the mean (standard deviation) of the proportion of iterations spent in states where the hypothesized relationship is true for each dose level.

Simulation results for Scenario 1 are presented in Table 4.1. In Scenario 1, the assumed relationship is correct for all dose-levels and we expect that model averaging will lead to increased efficiency compared to the model that estimates all components. All four models are mostly unbiased across the posterior estimates, with the exception of the proportion compliant and the non-compliant component mean for the group with ES=1. Additionally, the IND approach shows greater bias in the estimation of the compliant component mean for the group with ES=2 and its corresponding 95th percentile of the TNE distribution. Both RJMCMC models strongly favor the REL model with each level spending more than 84% of the chain in states that assume the hypothesized relationship. The REL model is more efficient than IND, as is evident by the SD ratio values below 1, and, because the two RJMCMC models strongly favor REL, they are also much more efficient than IND. For example, we observe a 50% to 80% decrease in the standard deviation of the mean of the compliant component and as much as an 85% decrease in the MSE of the 95th percentile.

The results in Tables 4.2-4.6 (Scenarios 2-6) demonstrate that model averaging down-weights the proportion of the chain estimated by the REL approach if a dose-level is not appropriately defined by the hypothesized relationship. However, large effect sizes are needed to downweight the influence of the REL approach. For example, when there was a one ES difference between the two and hypothesized mean for the compliant component, the RJ model picked models that assume the hypothesized relationship approximately 75%

of the time, while the RJ_{95} picked models that assume the hypothesized relationship approximately 45% of the time (Tables 4.3-4.5). This demonstrates that calibration of model priors in the RJMCMC algorithm is important and can have a direct impact on the proportion of time spent in a given state. The direction of misspecification (i.e., effect sizes greater or less than the hypothesized relationship) influences the amount of downweighting of the REL approach. This can be be observed in Table 4.6 (Scenario 6), which has effect sizes that are larger than the hypothesized relationship, where the RJ and RJ_{95} models were less likely to pick models that assumed the hypothesized relationship than in Scenario 5, even though the absolute difference of the ES from the hypothesized mean relationship was the same for both scenarios. Interestingly, the true value for one group does not seem to have a large impact on whether borrowing from the REL model occurs. For example, the proportion of time a group spends in REL when misspecified is fairly constant across other scenarios, even though different groups are misspecified in each scenario.

		$\hat{p}_c$			$\hat{\mu}_1$			$\hat{\mu}_2$			$\hat{\sigma}$			e^{μ_1} 95$^{\text{th}}$ Percentile			Proportion
Model	ES	Bias	SD*	MSE	Bias	SD*	MSE	Bias	SD*	MSE	Bias	SD*	MSE	Bias	SD*	MSE	in REL
	4	0.00	0.02	0.00	0.00	0.20	0.04	-0.01	0.13	0.02	0.02	0.08	0.01	0.21	0.61	0.42	-
	3	0.01	0.04	0.00	0.06	0.30	0.09	-0.03	0.16	0.03	0.06	0.13	0.02	3.47	6.53	54.75	-
IND	2	0.06	0.09	0.01	0.16	0.41	0.19	0.00	0.18	0.03	0.09	0.12	0.02	14.51	13.01	379.89	-
	1	0.17	0.13	0.05	-0.05	0.40	0.16	0.30	0.29	0.17	-0.04	0.09	0.01	4.93	12.43	178.90	-
	0	-	-	-	0.00	0.10	0.01	-	-	-	0.00	0.07	0.01	-	-	-	-
	4	0.00	0.93	0.00	0.01	0.42	0.01	-0.01	0.99	0.02	0.02	0.98	0.01	0.14	0.58	0.15	-
	3	0.00	0.80	0.00	0.01	0.29	0.01	-0.03	0.98	0.03	0.04	0.83	0.01	0.69	0.19	2.05	-
REL	2	-0.01	0.62	0.00	0.01	0.21	0.01	-0.04	0.93	0.03	0.04	0.89	0.01	1.80	0.25	13.51	-
	1	0.11	0.64	0.02	0.01	0.21	0.01	0.22	0.59	0.07	-0.06	1.04	0.01	-1.82	0.53	46.49	-
	0	-	-	-	0.01	0.85	0.01	-	-	-	0.00	1.00	0.01	-	-	-	-
	4	0.00	0.94	0.00	0.01	0.44	0.01	-0.01	0.99	0.02	0.02	0.98	0.01	0.14	0.59	0.15	0.98 (0.06)
	3	0.00	0.78	0.00	0.01	0.32	0.01	-0.03	0.97	0.03	0.04	0.83	0.01	0.77	0.22	2.73	0.98 (0.05)
RJ	2	-0.01	0.59	0.00	0.02	0.24	0.01	-0.04	0.92	0.03	0.04	0.90	0.01	2.34	0.29	20.11	0.97 (0.05)
	1	0.12	0.65	0.02	0.01	0.25	0.01	0.22	0.61	0.08	-0.06	1.03	0.01	-1.58	0.53	45.45	0.97 (0.05)
	0	-	-	-	0.01	0.85	0.01	-	-	-	0.00	1.00	0.01	-	-	-	-
	4	0.00	0.94	0	0.01	0.51	0.01	-0.01	1	0.02	0.02	0.98	0.01	0.14	0.62	0.16	0.91 (0.12)
	3	0.00	0.77	0.00	0.02	0.43	0.02	-0.03	0.97	0.03	0.04	0.85	0.01	1.10	0.37	7.05	0.90 (0.12)
RJ_{95}	2	0.00	0.60	0.00	0.05	0.41	0.03	-0.03	0.91	0.03	0.04	0.92	0.01	4.18	0.46	53.12	0.84 (0.13)
	1	0.13	0.69	0.02	0.00	0.46	0.03	0.24	0.68	0.10	-0.06	1.03	0.01	-0.85	0.59	54.77	0.85 (0.12)
	0	-	-	-	0.01	0.86	0.01	-	-	-	0.00	1.00	0.01	-	-	-	-

Table 4.1: Scenario 1 simulation results where all dose-levels follow the assumed relationship. Bias and MSE reported for proportion compliant ($\hat{p}_c$), compliant mean ($\hat{\mu}_1$), non-compliant mean($\hat{\mu}_2$), shared standard deviation ($\hat{\sigma}$), and the estimated 95th percentile for $\exp(\mu_1)$ for IND model, REL model, and model averaging between the two approaches with equal model priors (RJ) and priors favoring IND (RJ_{95}). Estimated standard deviation (SD*) provided for the IND approach with the ratio of the $\frac{SD}{SD_{IND}}$ for REL, RJ, and RJ_{95}.

		$\hat{p}_c$			$\hat{\mu}_1$			$\hat{\mu}_2$			$\hat{\sigma}$			e^{μ_1} 95$^{\text{th}}$ Percentile			Proportion
Model	ES	Bias	SD*	MSE	Bias	SD*	MSE	Bias	SD*	MSE	Bias	SD*	MSE	Bias	SD*	MSE	in REL
	4	0.00	0.02	0.00	0.00	0.20	0.04	-0.01	0.13	0.02	0.02	0.09	0.01	0.21	0.61	0.41	-
	3	0.01	0.04	0.00	0.06	0.29	0.09	-0.03	0.16	0.03	0.06	0.13	0.02	3.45	6.42	53.08	-
IND	2	0.06	0.09	0.01	0.16	0.42	0.20	0.00	0.19	0.04	0.09	0.12	0.02	14.46	12.92	376.00	-
	0.5	0.20	0.14	0.06	-0.29	0.35	0.21	0.47	0.32	0.33	-0.1	0.08	0.02	-14.83	12.50	376.26	-
	0	-	-	-	0.00	0.10	0.01	-	-	-	0.00	0.07	0.01	-	-	-	-
	4	0.00	0.93	0.00	0.02	0.42	0.01	-0.01	0.99	0.02	0.02	0.98	0.01	0.16	0.60	0.16	-
	3	0.00	0.81	0.00	0.02	0.29	0.01	-0.03	0.98	0.03	0.04	0.83	0.01	0.75	0.20	2.18	-
REL	2	-0.01	0.61	0.00	0.02	0.21	0.01	-0.04	0.91	0.03	0.04	0.89	0.01	1.98	0.25	14.54	-
	0.5	0.03	0.61	0.01	-0.48	0.25	0.24	0.26	0.47	0.09	-0.13	1.03	0.02	-30.65	0.42	966.94	-
	0	-	-	-	0.02	0.85	0.01	-	-	-	0.00	1	0.01	-	-	-	-
	4	0.00	0.93	0.00	0.02	0.45	0.01	-0.01	0.99	0.02	0.02	0.98	0.01	0.16	0.60	0.16	0.98 (0.06)
	3	0.00	0.80	0.00	0.02	0.32	0.01	-0.03	0.98	0.03	0.04	0.83	0.01	0.81	0.23	2.80	0.98 (0.05)
RJ	2	-0.01	0.59	0.00	0.03	0.24	0.01	-0.04	0.90	0.03	0.04	0.90	0.01	2.45	0.29	20.17	0.97 (0.05)
	0.5	0.05	0.61	0.01	-0.46	0.30	0.22	0.29	0.52	0.11	-0.13	1.03	0.02	-29.47	0.48	904.06	0.95 (0.09)
	0	-	-	-	0.02	0.85	0.01	-	-	-	0.00	1	0.01	-	-	-	-
	4	0.00	0.94	0.00	0.01	0.51	0.01	-0.01	0.99	0.02	0.02	0.98	0.01	0.15	0.64	0.17	0.91 (0.12)
	3	0.00	0.79	0.00	0.03	0.43	0.02	-0.03	0.97	0.03	0.04	0.85	0.01	1.13	0.36	6.64	0.90 (0.12)
RJ_{95}	2	0.00	0.59	0.00	0.05	0.39	0.03	-0.03	0.88	0.03	0.04	0.92	0.01	4.35	0.46	54.54	0.84 (0.13)
	0.5	0.09	0.72	0.02	-0.41	0.47	0.20	0.35	0.71	0.17	-0.13	1.03	0.02	-26.27	0.65	757.01	0.78 (0.17)
	0	-	-	-	0.02	0.87	0.01	-	-	-	0.00	1.00	0.01	-	-	-	-

Table 4.2: Scenario 2 simulation results where dose-level 4 does not follow the assumed relationship. Bias and MSE reported for proportion compliant ($\hat{p}_c$), compliant mean ($\hat{\mu}_1$), non-compliant mean($\hat{\mu}_2$), shared standard deviation ($\hat{\sigma}$), and the estimated 95th percentile for $\exp(\mu_1)$ for IND model, REL model, and model averaging between the two approaches with equal model priors (RJ) and priors favoring IND (RJ_{95}). Estimated standard deviation (SD*) provided for the IND approach with the ratio of the $\frac{SD}{SD_{IND}}$ for REL, RJ, and RJ_{95}.

		$\hat{p}_c$			$\hat{\mu}_1$			$\hat{\mu}_2$			$\hat{\sigma}$			e^{μ_1} 95$^{\text{th}}$ Percentile			Proportion
Model	ES	Bias	SD*	MSE	Bias	SD*	MSE	Bias	SD*	MSE	Bias	SD*	MSE	Bias	SD*	MSE	in REL
	4	0.00	0.02	0.00	0.00	0.20	0.04	-0.01	0.13	0.02	0.02	0.09	0.01	0.21	0.63	0.44	-
	3	0.01	0.04	0.00	0.06	0.29	0.09	-0.03	0.16	0.03	0.06	0.13	0.02	3.41	6.27	50.91	-
IND	1	0.16	0.13	0.04	-0.06	0.42	0.18	0.30	0.29	0.18	-0.04	0.09	0.01	4.39	12.46	174.58	-
	0.5	0.19	0.14	0.06	-0.3	0.35	0.21	0.47	0.32	0.32	-0.1	0.08	0.02	-14.93	12.45	377.74	-
	0	-	-	-	0.00	0.10	0.01	-	-	-	0.00	0.07	0.01	-	-	-	-
	4	0.00	0.93	0.00	0.04	0.43	0.01	-0.01	0.99	0.02	0.02	0.98	0.01	0.19	0.59	0.17	-
	3	0.00	0.81	0.00	0.04	0.30	0.01	-0.03	0.97	0.03	0.04	0.83	0.01	0.84	0.21	2.40	-
REL	1	-0.18	0.30	0.04	-0.96	0.21	0.93	-0.12	0.39	0.03	-0.03	0.91	0.01	-23.87	0.18	574.96	-
	0.5	0.04	0.62	0.01	-0.46	0.25	0.22	0.26	0.47	0.09	-0.13	1.03	0.02	-30.15	0.43	937.94	-
	0	-	-	-	0.04	0.87	0.01	-	-	-	0.00	1	0.01	-	-	-	-
	4	0.00	0.93	0.00	0.03	0.46	0.01	-0.01	0.99	0.02	0.02	0.98	0.01	0.18	0.59	0.17	0.98 (0.07)
	3	0.00	0.80	0.00	0.03	0.34	0.01	-0.03	0.97	0.03	0.04	0.83	0.01	0.89	0.24	3.04	0.98 (0.05)
RJ	1	-0.08	0.93	0.02	-0.71	0.75	0.60	0.02	0.79	0.05	-0.04	0.93	0.01	-16.66	0.74	363.57	0.77 (0.24)
	0.5	0.05	0.63	0.01	-0.45	0.31	0.21	0.29	0.55	0.12	-0.13	1.04	0.02	-29.14	0.49	886.92	0.95 (0.09)
	0	-	-	-	0.03	0.89	0.01	-	-	-	0.00	1	0.01	-	-	-	-
	4	0.00	0.93	0.00	0.02	0.52	0.01	-0.01	0.99	0.02	0.02	0.98	0.01	0.16	0.61	0.17	0.91 (0.12)
	3	0.00	0.79	0.00	0.03	0.44	0.02	-0.03	0.97	0.03	0.04	0.85	0.01	1.15	0.38	7.14	0.90 (0.12)
RJ_{95}	1	0.04	1.14	0.03	-0.39	0.97	0.32	0.17	1.02	0.12	-0.05	0.95	0.01	-6.78	1.02	206.56	0.44 (0.26)
	0.5	0.09	0.72	0.02	-0.41	0.49	0.20	0.34	0.72	0.17	-0.13	1.03	0.02	-26.27	0.65	755.77	0.79 (0.17)
	0	-	-	-	0.02	0.89	0.01	-	-	-	0.00	1.00	0.01	-	-	-	-

Table 4.3: Scenario 3 simulation results where dose-levels 3 and 4 do not follow the assumed relationship. Bias and MSE reported for proportion compliant ($\hat{p}_c$), compliant mean ($\hat{\mu}_1$), non-compliant mean($\hat{\mu}_2$), shared standard deviation ($\hat{\sigma}$), and the estimated 95th percentile for $\exp(\mu_1)$ for IND model, REL model, and model averaging between the two approaches with equal model priors (RJ) and priors favoring IND (RJ_{95}). Estimated standard deviation (SD*) provided for the IND approach with the ratio of the $\frac{SD}{SD_{IND}}$ for REL, RJ, and RJ_{95}.

		$\hat{p}_c$			$\hat{\mu}_1$			$\hat{\mu}_2$			$\hat{\sigma}$			e^{μ_1} 95$^{\text{th}}$ Percentile			Proportion
Model	ES	Bias	SD*	MSE	Bias	SD*	MSE	Bias	SD*	MSE	Bias	SD*	MSE	Bias	SD*	MSE	in REL
	4	0.00	0.02	0.00	0.00	0.20	0.04	-0.01	0.13	0.02	0.02	0.09	0.01	0.21	0.62	0.43	-
	1.5	0.12	0.12	0.03	0.11	0.41	0.18	0.13	0.23	0.07	0.04	0.10	0.01	14.17	12.89	367.02	-
IND	1	0.16	0.14	0.04	-0.06	0.42	0.18	0.30	0.3	0.18	-0.04	0.09	0.01	4.37	12.49	175.13	-
	0.5	0.20	0.14	0.06	-0.3	0.35	0.21	0.47	0.32	0.33	-0.1	0.08	0.02	-14.82	12.44	374.27	-
	0	-	-	-	0.00	0.10	0.01	-	-	-	0.00	0.07	0.01	-	-	-	-
	4	0.00	0.92	0.00	0.06	0.46	0.01	-0.01	0.98	0.02	0.02	0.98	0.01	0.24	0.62	0.21	-
	1.5	-0.23	0.19	0.06	-1.44	0.23	2.09	-0.32	0.47	0.12	0.11	0.84	0.02	-16.48	0.08	272.66	-
REL	1	-0.18	0.30	0.03	-0.94	0.22	0.90	-0.11	0.39	0.03	-0.03	0.91	0.01	-23.63	0.19	563.73	-
	0.5	0.04	0.65	0.01	-0.44	0.27	0.21	0.27	0.49	0.10	-0.13	1.03	0.02	-29.46	0.45	899.40	-
	0	-	-	-	0.06	0.93	0.01	-	-	-	0.01	1	0.01	-	-	-	-
	4	0.00	0.93	0.00	0.04	0.49	0.01	-0.01	0.99	0.02	0.02	0.98	0.01	0.20	0.62	0.19	0.98 (0.06)
	1.5	-0.05	1.28	0.03	-0.67	1.47	0.82	-0.07	1.11	0.07	0.06	0.95	0.01	-1.74	1.02	174.71	0.55 (0.31)
RJ	1	-0.07	0.96	0.02	-0.68	0.77	0.57	0.04	0.85	0.06	-0.04	0.92	0.01	-16.02	0.80	357.47	0.76 (0.25)
	0.5	0.06	0.65	0.01	-0.44	0.31	0.20	0.30	0.57	0.12	-0.13	1.03	0.02	-28.75	0.51	866.84	0.95 (0.09)
	0	-	-	-	0.04	0.94	0.01	-	-	-	0.00	1	0.01	-	-	-	-
	4	0.00	0.94	0.00	0.02	0.54	0.01	-0.01	0.99	0.02	0.02	0.98	0.01	0.17	0.63	0.18	0.91 (0.12)
	1.5	0.05	1.18	0.02	-0.19	1.30	0.32	0.06	1.07	0.06	0.04	0.97	0.01	8.00	1.04	242.01	0.24 (0.22)
RJ_{95}	1	0.04	1.12	0.03	-0.38	0.96	0.31	0.18	0.98	0.12	-0.05	0.95	0.01	-6.66	1.00	200.63	0.44 (0.26)
	0.5	0.09	0.74	0.02	-0.40	0.51	0.20	0.35	0.72	0.18	-0.13	1.03	0.02	-25.99	0.67	744.54	0.79 (0.17)
	0	-	-	-	0.03	0.94	0.01	-	-	-	0.00	1.00	0.01	-	-	-	-

Table 4.4: Scenario 4 simulation results where dose-levels 2, 3, and 4 do not follow the assumed relationship. Bias and MSE reported for proportion compliant ($\hat{p}_c$), compliant mean ($\hat{\mu}_1$), non-compliant mean($\hat{\mu}_2$), shared standard deviation ($\hat{\sigma}$), and the estimated 95th percentile for exp(μ_1) for IND model, REL model, and model averaging between the two approaches with equal model priors (RJ) and priors favoring IND (RJ_{95}). Estimated standard deviation (SD*) provided for the IND approach with the ratio of the $\frac{SD}{SD_{IND}}$ for REL, RJ, and RJ_{95}.

		$\hat{p}_c$			$\hat{\mu}_1$			$\hat{\mu}_2$			$\hat{\sigma}$			e^{μ_1} 95$^{\text{th}}$ Percentile			Proportion
Model	ES	Bias	SD*	MSE	Bias	SD*	MSE	Bias	SD*	MSE	Bias	SD*	MSE	Bias	SD*	MSE	in REL
	2	0.06	0.09	0.01	0.15	0.39	0.18	-0.01	0.18	0.03	0.10	0.12	0.02	14.41	12.64	367.49	-
	1.5	0.12	0.12	0.03	0.12	0.41	0.18	0.14	0.23	0.07	0.04	0.10	0.01	14.26	12.82	367.68	-
IND	1	0.16	0.14	0.04	-0.07	0.43	0.19	0.30	0.3	0.18	-0.04	0.09	0.01	4.22	12.54	175.04	-
	0.5	0.19	0.14	0.06	-0.29	0.34	0.20	0.47	0.32	0.33	-0.1	0.08	0.02	-14.86	12.51	377.29	-
	0	-	-	-	0.00	0.10	0.01	-	-	-	0.00	0.07	0.01	-	-	-	-
	2	-0.25	0.17	0.06	-1.91	0.26	3.67	-0.49	0.55	0.25	0.26	0.66	0.07	-10.79	0.04	116.70	-
	1.5	-0.23	0.20	0.05	-1.41	0.26	2.01	-0.32	0.47	0.11	0.11	0.85	0.02	-16.28	0.09	266.15	-
REL	1	-0.18	0.32	0.03	-0.91	0.24	0.84	-0.11	0.39	0.02	-0.03	0.92	0.01	-23.21	0.20	544.83	-
	0.5	0.06	0.68	0.01	-0.41	0.30	0.18	0.29	0.51	0.11	-0.13	1.03	0.02	-28.20	0.49	833.09	-
	0	-	-	-	0.09	1.03	0.02	-	-	-	0.01	1.01	0.01	-	-	-	-
	2	-0.03	1.4	0.02	-0.52	1.9	0.84	-0.14	1.34	0.08	0.13	1.05	0.03	5.43	0.94	170.84	0.35 (0.31)
	1.5	-0.06	1.23	0.02	-0.68	1.45	0.81	-0.08	1.07	0.07	0.06	0.97	0.01	-2.07	0.98	162.24	0.56 (0.31)
RJ	1	-0.07	0.97	0.02	-0.67	0.76	0.56	0.04	0.83	0.06	-0.04	0.93	0.01	-15.99	0.79	352.75	0.76 (0.25)
	0.5	0.06	0.66	0.01	-0.42	0.33	0.19	0.30	0.53	0.12	-0.13	1.03	0.02	-28.19	0.51	835.49	0.95 (0.08)
	0	-	-	-	0.06	1.05	0.01	-	-	-	0.01	1.01	0.01	-	-	-	-
	2	0.03	1.13	0.01	-0.06	1.34	0.28	-0.05	1.12	0.04	0.10	1.01	0.03	11.62	0.97	286.01	0.11 (0.15)
	1.5	0.05	1.18	0.02	-0.20	1.32	0.33	0.06	1.07	0.06	0.04	0.99	0.01	7.73	1.04	238.01	0.24 (0.22)
RJ_{95}	1	0.04	1.09	0.02	-0.40	0.95	0.33	0.17	0.95	0.11	-0.05	0.96	0.01	-7.09	0.99	203.47	0.45 (0.26)
	0.5	0.09	0.75	0.02	-0.40	0.52	0.19	0.35	0.72	0.18	-0.13	1.03	0.02	-25.79	0.67	734.85	0.79 (0.17)
	0	-	-	-	0.04	1.04	0.01	-	-	-	0.00	1.00	0.01	-	-	-	-

Table 4.5: Scenario 5 simulation results where no dose-level follows the assumed relationship. Bias and MSE reported for proportion compliant ($\hat{p}_c$), compliant mean ($\hat{\mu}_1$), non-compliant mean($\hat{\mu}_2$), shared standard deviation ($\hat{\sigma}$), and the estimated 95th percentile for $\exp(\mu_1)$ for IND model, REL model, and model averaging between the two approaches with equal model priors (RJ) and priors favoring IND (RJ_{95}). Estimated standard deviation (SD*) provided for the IND approach with the ratio of the $\frac{SD}{SD_{IND}}$ for REL, RJ, and RJ_{95}.

Model	ES	$\hat{p}_c$ Bias	$\hat{p}_c$ SD*	$\hat{p}_c$ MSE	$\hat{\mu}_1$ Bias	$\hat{\mu}_1$ SD*	$\hat{\mu}_1$ MSE	$\hat{\mu}_2$ Bias	$\hat{\mu}_2$ SD*	$\hat{\mu}_2$ MSE	$\hat{\sigma}$ Bias	$\hat{\sigma}$ SD*	$\hat{\sigma}$ MSE	e^{μ_1} 95$^{\text{th}}$ Percentile Bias	e^{μ_1} 95$^{\text{th}}$ Percentile SD*	e^{μ_1} 95$^{\text{th}}$ Percentile MSE	Proportion in REL
	6	0.00	0.00	0.00	0.00	0.19	0.03	-0.01	0.12	0.01	0.01	0.07	0.01	0.02	0.06	0.00	-
	4.5	0.00	0.01	0.00	0.01	0.19	0.04	0.00	0.13	0.02	0.01	0.08	0.01	0.11	0.32	0.12	-
IND	3	0.01	0.04	0.00	0.06	0.30	0.09	-0.03	0.16	0.03	0.06	0.13	0.02	3.25	6.18	48.73	-
	1.5	0.12	0.12	0.03	0.11	0.43	0.19	0.13	0.22	0.07	0.04	0.10	0.01	14.45	13.03	378.69	-
	0	-	-	-	0.00	0.10	0.01	-	-	-	0.00	0.07	0.01	-	-	-	-
	6	0.02	2.72	0.00	1.40	0.91	2.00	0.02	0.98	0.01	0.30	1.84	0.11	1.68	13.41	3.54	-
	4.5	0.03	1.48	0.00	0.90	0.87	0.84	0.05	0.99	0.02	0.16	1.62	0.04	3.02	5.68	12.44	-
REL	3	0.03	0.82	0.00	0.40	0.57	0.19	0.02	0.98	0.03	0.09	1.00	0.02	4.76	0.61	37.10	-
	1.5	-0.01	0.65	0.01	-0.10	0.39	0.04	0.01	0.70	0.02	-0.01	0.97	0.01	-1.07	0.43	32.15	-
	0	-	-	-	-0.60	1.67	0.39	-	-	-	0.18	1.73	0.05	-	-	-	-
	6	0.00	1.01	0.00	0.00	1.00	0.03	-0.01	1.00	0.01	0.01	1.00	0.01	0.02	1.00	0.00	0.00 (0.00)
	4.5	0.00	1.01	0.00	0.01	1.03	0.04	0.00	1.00	0.02	0.01	1.01	0.01	0.13	1.21	0.17	0.00 (0.03)
RJ	3	0.04	1.20	0.00	0.40	1.57	0.37	-0.03	1.03	0.03	0.14	1.36	0.05	9.24	1.48	169.07	0.48 (0.38)
	1.5	0.18	0.73	0.04	0.46	0.33	0.23	0.14	0.77	0.05	0.07	0.98	0.01	21.00	0.74	533.41	0.96 (0.07)
	0	-	-	-	-0.02	1.08	0.01	-	-	-	0.00	1.00	0.01	-	-	-	-
	6	0.00	1.01	0.00	0.00	1.00	0.03	-0.01	1.00	0.01	0.01	1.00	0.01	0.02	1.00	0.00	0.00 (0.00)
	4.5	0.00	1.00	0.00	0.01	1.01	0.04	0.00	1.00	0.02	0.01	1.00	0.01	0.12	1.02	0.12	0.00 (0.01)
RJ_{95}	3	0.02	1.13	0.00	0.22	1.36	0.21	-0.03	1.00	0.03	0.09	1.18	0.03	5.76	1.21	88.66	0.26 (0.29)
	1.5	0.17	0.79	0.04	0.39	0.54	0.21	0.14	0.79	0.05	0.06	0.98	0.01	19.67	0.76	484.76	0.83 (0.15)
	0	-	-	-	-0.01	1.04	0.01	-	-	-	0.00	1.00	0.01	-	-	-	-

Table 4.6: Scenario 6 simulation results the dose-level relationship underestimates the compliant means. Bias and MSE reported for proportion compliant ($\hat{p}_c$), compliant mean ($\hat{\mu}_1$), non-compliant mean($\hat{\mu}_2$), shared standard deviation ($\hat{\sigma}$), and the estimated 95th percentile for $\exp(\mu_1)$ for IND model, REL model, and model averaging between the two approaches with equal model priors (RJ) and priors favoring IND (RJ_{95}). Estimated standard deviation (SD*) provided for the IND approach with the ratio of the $\frac{SD}{SD_{IND}}$ for REL, RJ, and RJ_{95}.

4.4 Application to a regulatory tobacco clinical trial

This section provides a case study of our model averaging approach applied to the data from CENIC-p1. As mentioned in Section 4.1, previous research has identified TNE cut-offs that can be used to identify non-compliance among subjects randomized to the 0.4 mg/g groups. However, the investigators are also interested in understanding the effect of nicotine reduction among smokers that complied to the intervention in the intermediate nicotine content groups where biomarker thresholds for identifying non-compliance are not currently available.

Our analysis will focus on total nicotine equivalents (TNE; a biomarker of nicotine exposure that measures most nicotine metabolites) as our biomarker for identifying non-compliance. For the purposes of this analysis, we will analyze TNE on the log scale and assume that all mixture components for the log-transformed biomarker values are normally distributed. CENIC-p1 included two control conditions: a usual brand (UB) condition, who received their preferred brand of commercially available cigarettes, and a 15.8 mg/g study cigarette (roughly equivalent to the nicotine content of commercial cigarettes). In addition, CENIC-p1 included five experimental conditions: a 5.2 mg/g group, a 2.4 mg/g group, a 1.3 mg/g group, a 0.4 mg/g group, and a 0.4 mg/g condition with elevated tar yield (0.4 mg/g (HT) group) to understand the impact of tar yield on the effect of nicotine reduction. For the purposes of this analysis, we combine the two controls conditions and the two 0.4 mg/g groups because these groups have the same nicotine content, respectively, which is the primary factor influencing the distribution of the biomarkers. The proposed relationship from Section 4.2.1 is used with $w_j = (0.4, 1.3, 2.4, 5.2, 15.8)$, corresponding to a 97.5%, 91.8%, 85.0%, and 66.7% reduction (effect size of -3.7, -2.5, -1.9, and -1.1) in the average biomarker value for the VLNC group, 1.3 mg/g group, 2.4 mg/g group, and 5.2 mg/g group, respectively, relative to the combined usual brand/15.8 mg/g control group.

Table 4.7 provides summary statistics at week 6 by treatment group for all participants and restricted to those who self-reported compliance. Figure 4.1 provides histograms of log(TNE) by treatment group for self-reported compliers at trial completion (week 6 for CENIC-p1). A bimodal distribution can be observed for the 1.3 mg/g group and the combined VLNC group whereas a bimodal pattern is not easily identified for the 2.4 and 5.2 mg/g groups. This suggests that, while it may not be difficult to identify the compliant

and non-compliant subjects in the 1,3 mg/g and VLNC groups, there may not be any truly compliant subjects for the 2.4 and 5.2 mg/g groups, or, more likely, the distributions of log(TNE) for compliant and non-compliant subjects have sufficient overlap to make distinct bimodal patterns difficult to identify.

Covariate	UB	15.8	5.2	2.4	1.3	0.4 HT	0.4
All Subjects:	(N=118)	(N=119)	(N=122)	(N=119)	(N=119)	(N=123)	(N=119)
CPD Available	113 (95.8%)	111 (93.3%)	114 (93.4%)	107 (89.9%)	110 (92.4%)	116 (94.3%)	109 (91.6%)
TNE Data Available	109 (92.4%)	108 (90.8%)	107 (87.7%)	107 (89.9%)	109 (91.6%)	113 (91.9%)	107 (89.9%)
Study CPD	21.9 (13.0)	20.9 (12.4)	19.9 (14.6)	14.7 (10.5)	14.7 (10.7)	13.9 (9.65)	13.5 (10.1)
Non-Study CPD	0.28 (1.4)	0.41 (1.2)	0.92 (2.47)	1.83 (3.97)	1.63 (4.16)	1.94 (5.05)	1.35 (3.45)
TNE (nmol/mL)	59.5 (53.9)	49.9 (33.0)	39.6 (33.6)	34.8 (33.3)	35.2 (32.3)	36.1 (39.8)	34.7 (32.1)
Self-Reported Compliers:	(N=102)	(N=80)	(N=74)	(N=57)	(N=64)	(N=67)	(N=69)
CPD Available	102 (100.0%)	80 (100.0%)	74 (100.0%)	57 (100.0%)	64 (100.0%)	67 (100.0%)	69 (100.0%)
TNE Data Available	100 (98.0%)	78 (97.5%)	69 (93.2%)	56 (98.2%)	64 (100.0%)	65 (97.0%)	68 (98.6%)
Study CPD	22.7 (13.1)	21.3 (12.6)	22.2 (15.6)	16.3 (10.9)	17.0 (10.8)	14.9 (9.82)	15.5 (10.7)
Non-Study CPD	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
TNE (nmol/mL)	59.3 (53.4)	48.0 (31.9)	33.2 (29.6)	30.6 (36.2)	30.8 (31.5)	27.3 (36.8)	30.0 (33.7)

Table 4.7: Summary statistics at week 6 by CENIC-p1 groups for all subjects and restricted to those self-reporting compliance at week 6. Mean (sd) for continuous measures, N (%) for categorical measures.

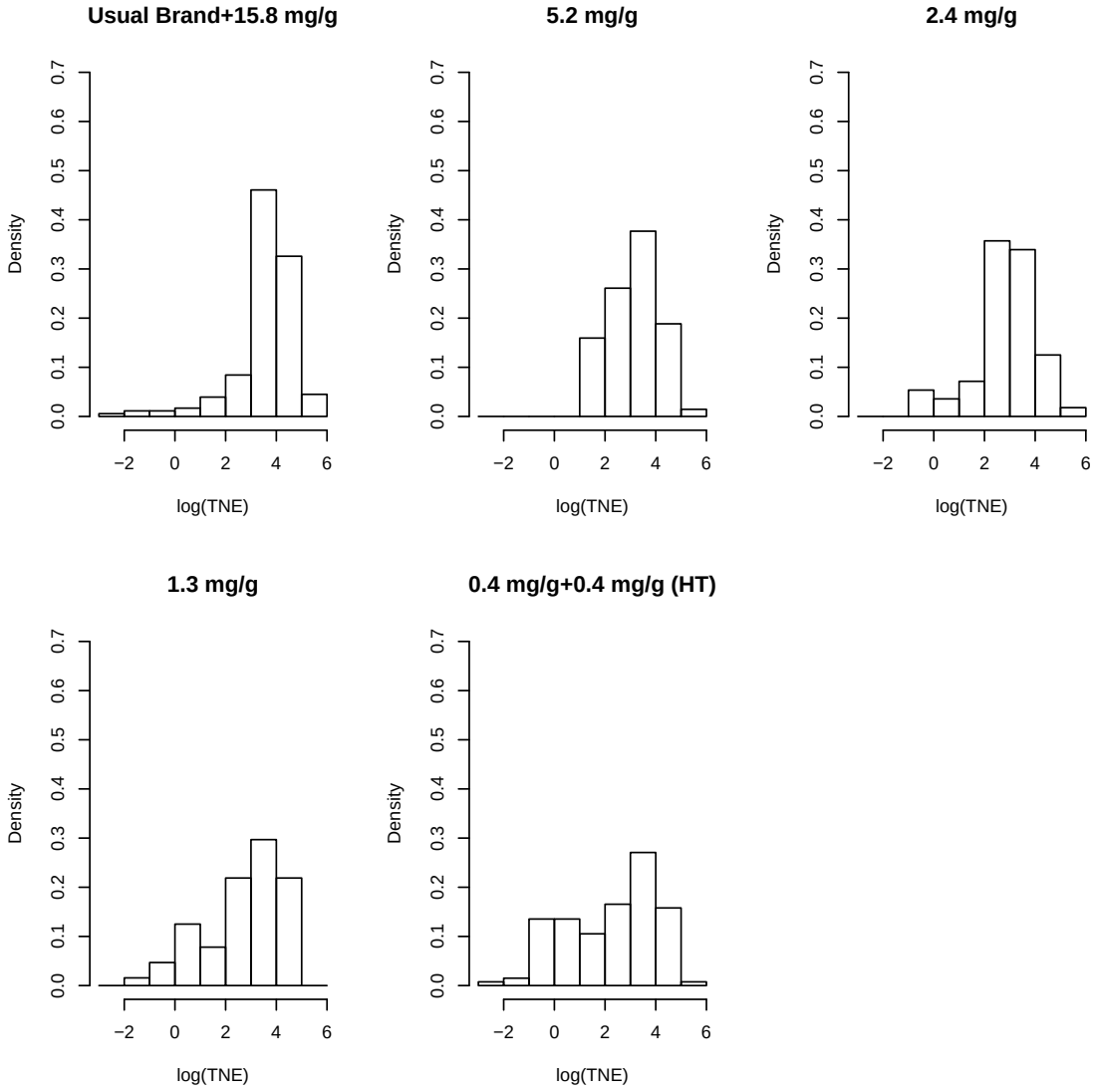

Figure 4.1: Histograms for distribution of self-reported compliers of log(TNE) in each CENIC-p1 treatment grouping at week 6.

For all j dose-levels, assume "non-informative" prior specifications of $\alpha_j = \beta_j = 1$ for the beta prior on p_j, $a_j = b_j = 0.001$ for the gamma prior on τ_j, and $s_{\mu_j} = s_{\theta_j} = 0.00001$ for the normal priors on μ_j and θ_j. The REL and IND models were fit on three chains with different, overdispersed initial values of length 20,000 with a burn-in of 2,000 iterations.

Chain length, burn-in, and MCMC convergence of the parameters were established by Gelman-Rubin diagnostic values below 1.05 (Gelman and Rubin, 1992), autocorrelation plots with a lag up to 50 iterations, and trace plots. The RJMCMC models were fit on three chains with a length of 400,000 with a burn-in of 200,000 iterations. Convergence is challenging for reversible jump algorithms since other methods for MCMC diagnostics are not applicable due to the changing model space (Green and Hastie, 2009), but utilizing longer chains and discarding the first half of the observations provide a greater opportunity to explore the total model space of 16 potential models.

The results of applying both approaches and the RJMCMC algorithm assuming $p(I_j) = 0.5$ for each group j are presented in Table 4.8. Presented are the posterior means and 95% highest posterior density (HPD) intervals for the proportion compliant ($\hat{p}_c$), the compliant component mean for log(TNE) ($\hat{\mu}_1$), the non-compliant component mean for log(TNE) ($\hat{\mu}_2$), the shared standard deviation between the two components ($\hat{\sigma}$), and the 90th and 95th percentile of the biomarker distribution for compliant subjects. The last quantities have been suggested as thresholds for identifying non-compliance in trials of reduced nicotine content cigarettes (Denlinger et al., 2016).

Results for the IND and REL models are similar, in general, suggesting that the hypothesized relationship between nicotine dose and biomarkers of nicotine exposure may be reasonable. As a result, our RJMCMC algorithm chooses the relationship over 90% of the time for all dose-levels. Figure 4.2 presents histograms of log(TNE) by group and the fitted distributions from the RJMCMC algorithm. We see that our approach results in a reasonable fit to the data, with the possible exception of the 2.4 mg/g group, where it is difficult to identify the two components. The gains in efficiency from using the REL approach as compared to the IND approach can be seen in more precise estimates of the 90th and 95th percentiles across all groups, with reductions in the width of the HPD intervals ranging from 37-55% and 24-51% for the 90th and 95th percentile, respectively. In our simulation study, we observed that substantial overlap in the mixture components can lead to bias in the estimated marginal probability of compliance, which may explain why the estimated probability of compliance is higher for the 5.2 mg/g group than the other groups. Similarly, the 2.4 mg/g group lacks a clear bimodal distribution which makes estimation for the IND approach challenging and may introduce bias.

Boatman et al. (2017) showed that the probability of a subject being compliant, C_{ij}, can be determined as a function of the observed biomarker values and written as a function of the mixture components:

$$P(C_{ij}|y_{ij}) = \frac{(1-p_j)\mathcal{N}(y_{ij}|\mu_j, \tau_j)}{(1-p_j)\mathcal{N}(y_{ij}|\mu_j, \tau_j) + p_j\mathcal{N}(y_{ij}|\mu_j + \theta_j, \tau_j)}.$$

Figure 4.3 presents the estimated probability of compliance as a function of the biomarkers from the IND and RJMCMC models. With the exception of the 2.4 mg/g group, the curves for the RJMCMC approach are similar to the IND approach, suggesting that any bias due to borrowing is minimal and outweighed by the increased precision, as noted in Table 4.8. There is a substantial difference between the two curves for the 2.4 mg/g group, but it should be noted that both curves are below the curve for the 0.4 mg/g group for low biomarker values, which is inconsistent with our understanding of the underlying mechanism. Further work is needed to understand the discrepancy and to determine the validity of these results.

Group	Approach	$\hat{p}_c$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\sigma}$	Percentile for μ_1 90th	Percentile for μ_1 95th	Proportion in REL
	IND	0.36 (0.27,0.45)	0.20 (-0.12,0.51)	3.44 (3.21,3.67)	0.92 (0.78,1.05)	4.03 (2.62,5.54)	5.63 (3.67,8.03)	
VLNC	REL	0.34 (0.25,0.43)	-0.09 (-0.25,0.06)	3.38 (3.15,3.60)	0.93 (0.80,1.08)	3.03 (2.40,3.71)	4.26 (3.23,5.39)	
	RJ	0.34 (0.26,0.43)	-0.07 (-0.28,0.14)	3.39 (3.16,3.62)	0.93 (0.80,1.07)	3.10 (2.35,3.95)	4.35 (3.17,5.72)	0.935
	IND	0.27 (0.11,0.44)	0.77 (-0.01,1.48)	3.37 (2.95,3.76)	0.94 (0.70,1.25)	8.21 (2.53,15.45)	11.86 (3.57,23.10)	
1.3 mg/g	REL	0.31 (0.16,0.45)	1.09 (0.93,1.24)	3.45 (3.07,3.80)	0.93 (0.71,1.20)	9.95 (6.83,14.12)	14.09 (8.74,21.46)	
	RJ	0.31 (0.16,0.46)	1.08 (0.92,1.25)	3.44 (3.05,3.80)	0.93 (0.71,1.21)	9.97 (6.61,14.42)	14.15 (8.48,22.00)	0.990
	IND	0.15 (0.01,0.35)	0.40 (-1.03,2.26)	3.10 (2.74,3.46)	1.00 (0.77,1.29)	8.51 (0.31,29.45)	12.72 (0.36,45.17)	
2.4 mg/g	REL	0.27 (0.04,0.48)	1.70 (1.54,1.85)	3.21 (2.73,3.67)	1.10 (0.83,1.38)	23.03 (14.46,32.89)	34.81 (19.39,53.59)	
	RJ	0.26 (0.04,0.47)	1.58 (0.22,1.94)	3.20 (2.74,3.66)	1.09 (0.82,1.38)	21.65 (1.61,32.60)	32.68 (2.11,53.04)	0.901
	IND	0.43 (0.13,0.75)	2.35 (1.91,3.02)	3.69 (3.21,4.03)	0.63 (0.43,0.90)	25.41 (12.84,56.96)	32.56 (14.82,76.61)	
5.2 mg/g	REL	0.44 (0.22,0.65)	2.47 (2.32,2.62)	3.71 (3.29,4.05)	0.64 (0.45,0.88)	27.49 (18.82,38.67)	35.02 (21.70,52.25)	
	RJ	0.45 (0.23,0.66)	2.47 (2.31,2.63)	3.72 (3.30,4.06)	0.64 (0.45,0.88)	27.31 (18.39,38.94)	34.74 (21.55,52.93)	0.989
	IND	-	3.57 (3.40,3.75)	-	1.21 (1.09,1.34)	-	-	
UB/15.8 mg/g	REL	-	3.58 (3.43,3.73)	-	1.21 (1.08,1.34)	-	-	
	RJ	-	3.58 (3.42,3.74)	-	1.21 (1.09,1.34)	-	-	-

Table 4.8: Mean (95% HPD interval) of proportion in compliant group ($\hat{p}_c$), compliant component mean log(TNE) ($\hat{\mu}_1$), non-compliant component mean log(TNE) ($\hat{\mu}_2$), shared standard deviation ($\hat{\sigma}$), 90/95th percentile of $\hat{\mu}_1$, and proportion of the chain spent in the REL approach for the RJMCMC.

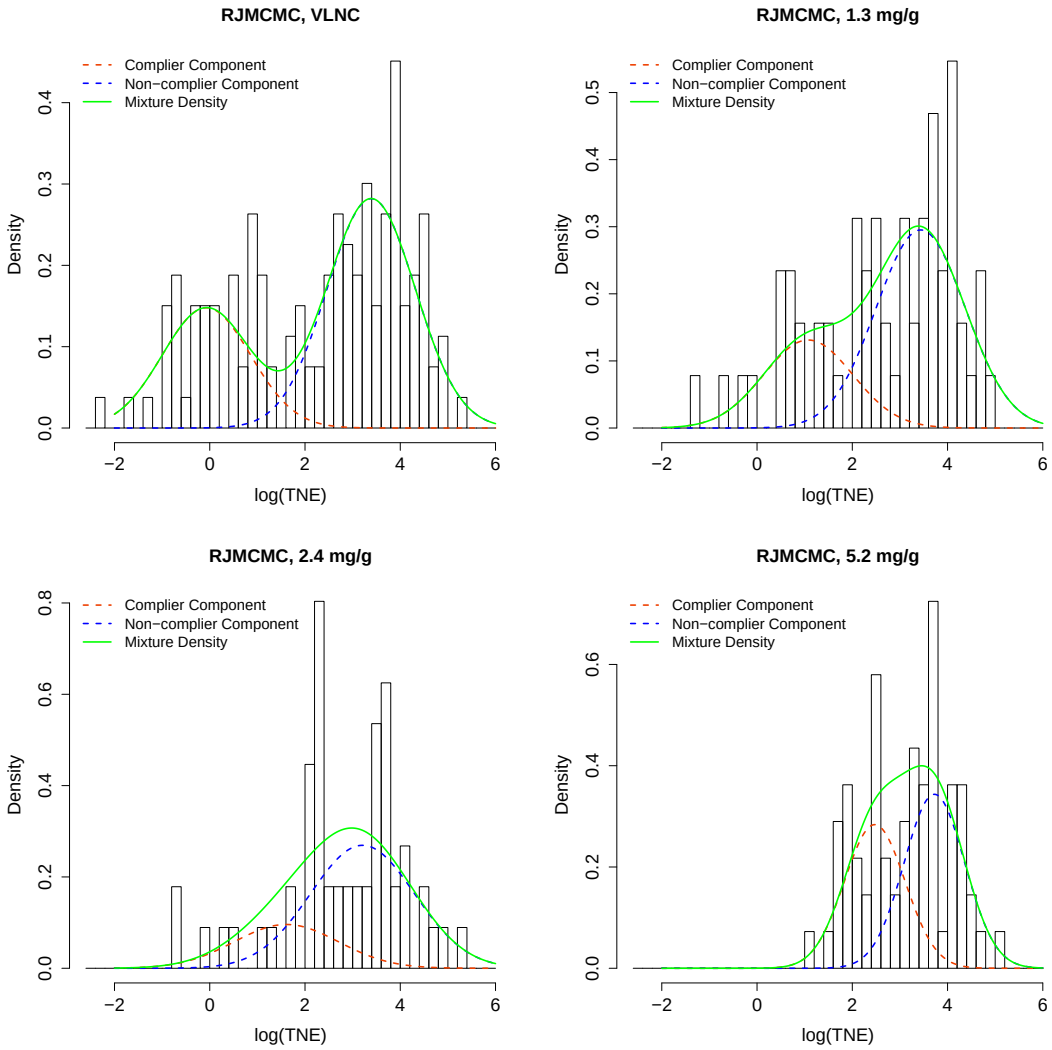

Figure 4.2: Histograms for distribution of self-reported compliers of log(TNE) in each CENIC-p1 treatment arm at week 6 with mixture densities for the RJ approach.

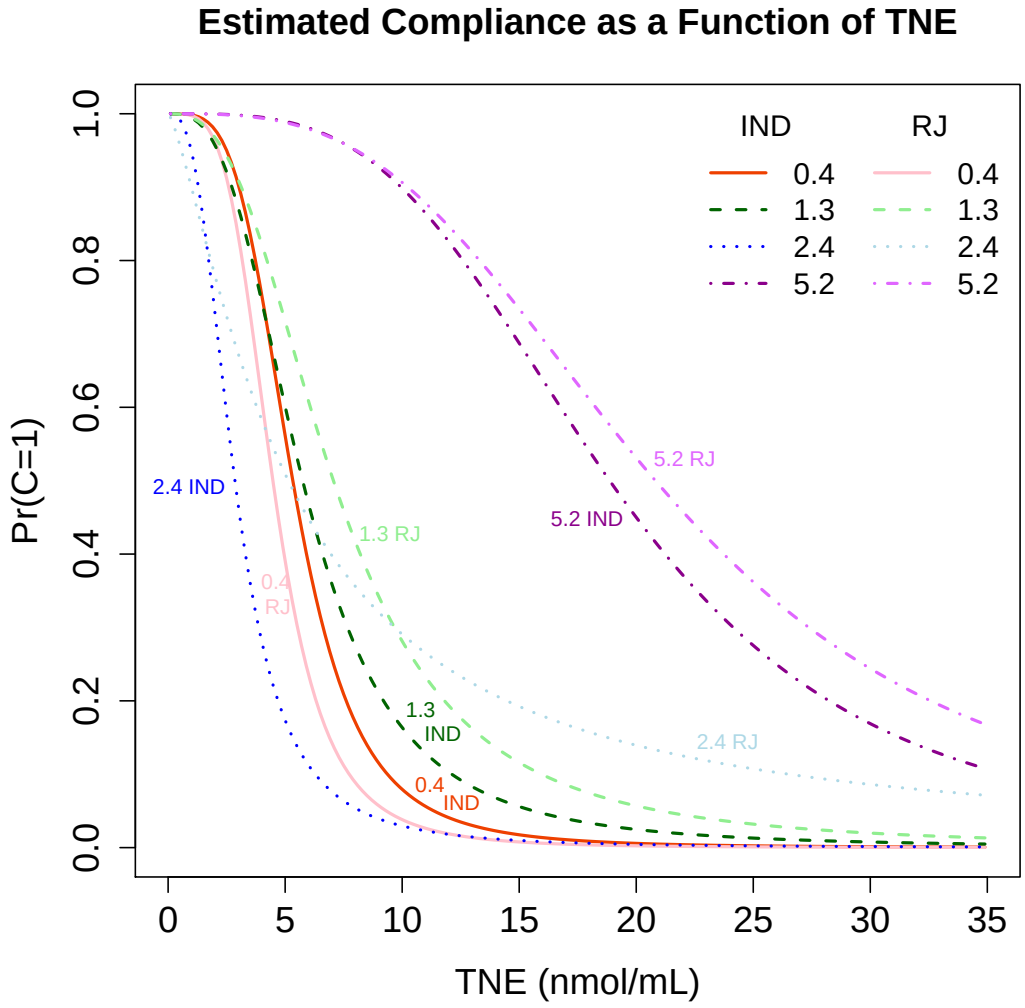

Figure 4.3: Predicted compliance (Pr(C=1)) as a function of observed TNE by study group for IND approach and model averaging with RJMCMC.

4.5 Discussion

The ability to identify non-compliance has important implications for interpreting the results of CENIC-p1. Previous work to identify non-compliance in CENIC-p1 leveraged the availability of auxiliary data to determine cut-offs for identifying non-compliance and to

estimate the probability of compliance conditional on the observed biomarker values for the VLNC group, only (Boatman et al., 2017). However, auxiliary data is not available for the intermediate dose-levels. The method proposed by Boatman et al. (2017) could be used to estimate the probability of non-compliance for the intermediate dose-levels, but the estimates will be inefficient in the absence of auxiliary data from known compliers. We propose a fully Bayesian approach to estimating the mixture components across all groups simultaneously by averaging over the model space, which included different assumptions regarding the association between the nicotine content of the cigarettes and biomarkers of nicotine exposure. Simulation results were encouraging, with model averaging achieving more precise estimates of the mixture components by choosing the state assuming the hypothesized relationship when the relationship was true and downweighting the hypothesized relationship when the relationship was misspecified.

One of the objectives of CENIC-p1 was to collect data that would inform the FDA as they consider new product standards for cigarettes, including a potential reduction in the nicotine content of cigarettes. The intention-to-treat analysis provides an estimate of the effect of an intervention, in practice, which may include non-compliance to the intervention, but it does not reflect the potential future reality where policy choices may lead to previous options disappearing from the market due to regulation (i.e., the FDA using its power to lower nicotine content of cigarettes). This motivates the desire to estimate the causal effect of nicotine reduction, which requires that investigators are able to accurately identify non-compliance to randomized treatment assignment. Previously, investigators were only able to identify non-compliance in the VLNC group, but the method proposed in this chapter provides a framework for identifying non-compliance at other dose-levels, as well. This is particularly important because a number of randomized trials of reduced nicotine content cigarettes are currently under way and some use dose-levels other than the 0.4 mg/g dose.

Mixture models can be challenging to estimate in practice, with problems such as label switching or unstable estimates. In the Bayesian context these can be addressed through assumptions regarding the relationships between components (i.e., forcing one component to have a larger mean than the other) and by monitoring the behavior of the resulting MCMC chains to ensure convergence to a steady state. While our Gaussian mixture model is for a finite, two-component mixture model, our approach could be extended to

any finite number of components. For example, our analysis only considered subjects that self-report compliance, in which case there are two groups: subjects that honestly self-report compliance and subjects that self-report compliance but were actually non-compliant. Alternately, we could analyze all subjects, which would result in a third mixture component for subjects that honestly self-report non-compliance.

The proposed Bayesian framework represents a flexible approach to identifying non-compliance in regulatory tobacco trials, but could be applied to other settings as well. For instance, the approach may be beneficial in therapeutic clinical trials where multiple doses are considered, but pharmacokinetic data are only available for one dose. In addition, future work will extend our proposed modeling framework to a regression setting where biomarkers can be modeled as a function of other covariates of interest that may impact the observed biomarker values, such as cigarettes per day. Additionally, the results of this chapter could be used to estimate graphical compliance networks, which can be combined with causal inference techniques to estimate causal effect for the intermediate dose levels.

Chapter 5

Conclusion

5.1 Summary of developments

In this thesis we developed multi-source exchangeability models, a general Bayesian framework for integrating multiple, potentially non-exchangeable, supplemental data sources into the analysis of a primary source. MEMs yield source-specific smoothing parameters that can be estimated by the data to facilitate a dynamic multi-resolution smoothed estimator that is asymptotically consistent while reducing the dimensionality of the prior space.

The general MEM framework, as developed in Chapter 2, was first applied to the context of Gaussian-distributed data with a known mean and unknown variance. The asymptotic consistency of MEM posterior weights, and consequently the marginal posterior distributions, were presented along with extensive simulation studies to exhibit the small sample properties. An application to estimate the reduction in cigarettes smoked per day in a regulatory tobacco clinical trial demonstrated the potential increase in efficiency that results from incorporating supplemental sources into the analysis of a clinical trial.

Chapter 3 extended the MEM framework to binary outcome data in the context of a proposed multi-source adaptive platform design. Motivated by the recent Ebola epidemic, the MEM framework is incorporated into the standard platform trial design to enable the incorporation of information from previous segments into the analysis of the current segment to improve efficiency. Maintaining a fixed randomization ratio may induce an imbalance of information between study groups when supplemental information is used

during data analysis. To address this potential for information imbalance, an adaptive randomization scheme was presented which adapts the allocation ratio for future participants and targets balanced information across treatment groups within a segment. We demonstrated via simulation that our proposed multi-source adaptive platform design using MEMs resulted in increased power and reasonable type-I error rates as compared to a standard design with no borrowing.

We returned to the Gaussian setting in Chapter 4 and considered the situation where there are multiple proposed models for estimating the components of mixture distributions in the context of identifying non-compliance in a regulatory tobacco clinical trial, and then applied the RJMCMC algorithm to induce model averaging over multiple candidate models. Simulation studies indicated that the RJMCMC algorithm properly weights the proposed models and results in more efficient estimates of the mixture component parameters when borrowing is appropriate. The application to CENIC-p1 was able to more efficiently estimate thresholds for identifying non-compliance at intermediate dose-levels, which were not previously available.

5.2 Significance of the work

While a number of methods for adaptively incorporating supplemental information into data analysis have been proposed in the literature, existing approaches typically rely on the data to inform a single parameter, which determines the extent of influence or shrinkage from all sources, risking considerable bias or minimal borrowing in the presence of heterogeneous supplemental data sources. The MEM framework is a general Bayesian hierarchical approach which effectuates source-specific smoothing parameters that can be estimated from the data. By estimating source-specific smoothing parameters we are able to accommodate multiple, potentially non-exchangeable sources without pre-specifying the smoothing parameters or relying on a single parameter to inform borrowing.

The multi-source adaptive platform design proposed in Chapter 3 addresses some of the concerns over the "traditional" drug development process which was seen by some as too time consuming and unable to quickly respond to a potentially dynamic, fast-changing disease outbreak. Simulations demonstrated increased power with reasonable inflation of the type-I error rate as compared to a standard platform design, which only

considered contemporaneous controls. Our proposed multi-source adaptive platform trial was strongly motivated by the recent Ebola outbreak in West Africa, but future infectious disease epidemics are inevitable and our proposed design represents an advancement in the tools available to respond to future epidemics.

The model averaging approach described in Chapter 4 represents an intuitive approach to borrowing information across multiple sources of information in the presence of a hypothesized relationship across sources. If the relationship is true, incorporating our biological understanding of the relationship between the groups can increase efficiency, while model averaging allows the flexibility to downweight the information from other groups when the relationship does not hold. Additionally, posterior model weights can provide evidence for the appropriateness of the hypothesized relationship separately for each group.

Finally, the use of effective supplemental sample size and posterior model weights provides a convenient way to convey the amount of supplemental information incorporated into a primary data analysis. Further, the fact that the posterior weights are constrained to be positive and sum to one provides an intuitive measure which can promote a greater understanding of how supplemental information is utilized in posterior calculations that may not be as apparent in other approaches. This can be viewed as a beneficial feature of our proposed framework when working with collaborators who may be wary of incorporating supplemental information or who want to be informed of how influential each supplemental source is on the analysis.

5.3 Future work and considerations

Throughout this dissertation, we only considered models that evaluate exchangeability based on a single endpoint, but source inclusion probabilities could also be determined by considering multiple endpoints, simultaneously. In practice, clinical trials in a common disease area will often have multiple endpoints in common. For example, the CENIC-p1 study in Chapters 2 and 4 is part of a larger network of Tobacco Centers of Regulatory Science (TCORS), a joint initiative of the FDA's Center for Tobacco Products (CTP) and the NIH. These centers are planning to conduct multiple trials relating to nicotine reduction and are actively working to identify a common set of outcome measures for evaluating the impact of tobacco product regulation.

Expanding MEMs to multiple endpoints is methodologically interesting in that different types of endpoints may be collected across multiple trials. For example, endpoints could be related to binary, count, or continuous data and all could be beneficial for evaluating exchangeability. In accounting for multiple endpoints one could evaluate each endpoint individually and then utilize the separate source inclusion probabilities, but jointly modeling endpoints to evaluate exchangeability could provide stronger evidence that sources are exchangeable rather than evaluating endpoints one-by-one. Accounting for these correlated endpoints could be achieved through many different approaches. One approach would be to induce correlation on the outcomes via copula models and use the resulting density to calculate the posterior model weights (Nelsen, 2006; Embrechts et al., 2001; Durrleman et al., 2000). If the multiple endpoints are all Gaussian distributed, a Gaussian copula would provide an intuitive structure. Alternatively, if there are non-Gaussian endpoints or a mixture of endpoints with different distributions, more flexible copula classes, such as Archimedean copulas, could account for these different distributions while enabling the ability to model the dependence of the endpoints through one parameter. An alternate approach would be to use generalized linear mixed models for each outcome where dependence across the outcomes is accounted for by specifying a joint distribution of the random effects in place of fitting separate univariate models for each outcome (Komárek and Komárková, 2013).

References

Adebamowo, C., Bah-Sow, O., Binka, F., Bruzzone, R., Caplan, A., Delfraissy, J.-F., Heymann, D., Horby, P., Kaleebu, P., Tamfum, J.-J. M., et al. (2014). Randomised controlled trials for ebola: practical and ethical issues. *Lancet* **384,** 1423.

Benowitz, N. L., Nardone, N., Hatsukami, D. K., and Donny, E. C. (2015). Biochemical estimation of noncompliance with smoking of very low nicotine content cigarettes. *Cancer Epidemiology and Prevention Biomarkers* **24,** 331–335.

Berry, D. A. and Eick, S. G. (1995). Adaptive assignment versus balanced randomization in clinical trials: A decision analysis. *Statistics in Medicine* **14,** 231–246.

Boatman, J. A., Vock, D. M., Koopmeiners, J. S., and Donny, E. C. (2017). Estimating causal effects from a randomized clinical trial when noncompliance is measured with error. *Biostatistics* .

Browne, W. J. and Draper, D. (2006). A comparison of bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis* **1,** 473–514.

Chen, M.-H., Ibrahim, J. G., Lam, P., Yu, A., and Zhang, Y. (2011). Bayesian design of noninferiority trials for medical devices using historical data. *Biometrics* **67,** 1163–1170.

Chib, S. (1996). Calculating posterior distributions and modal estimates in markov mixture models. *Journal of Econometrics* **75,** 79–97.

Daniels, M. J. (1999). A prior for the variance in hierarchical models. *The Canadian Journal of Statistics* **27,** 567–578.

De Santis, F. (2006). Power priors and their use in clinical trials. *The American Statistician* **60,** 122–129.

Denlinger, R. L., Smith, T. T., Murphy, S. E., Koopmeiners, J. S., Benowitz, N. L., Hatsukami, D. K., Pacek, L. R., Colino, C., Cwalina, S. N., and Donny, E. C. (2016). Nicotine and anatabine exposure from very low nicotine content cigarettes. *Tobacco Regulatory Science* **2,** 186–203.

Dodd, L. E., Proschan, M. A., Neuhaus, J., Koopmeiners, J. S., Neaton, J., Beigel, J. D., Barrett, K., Lane, H. C., and Davey Jr., R. T. (2016). Design of a randomized controlled trial for ebola virus disease medical countermeasures: PREVAIL II, the ebola MCM study. *The Journal of Infectious Diseases* **213,** 1906–1913.

Doi, S., Barendregt, J., and Mozurkewich, E. (2011). Meta-analysis of heterogeneous clinical trials: An empirical example. *Contemporary Clinical Trials* **32,** 288–298.

Donny, E. C., Denlinger, R. L., Tidey, J. W., Koopmeiners, J. K., et al. (2015). Randomized trial of reduced-nicotine standard for cigarettes. *New England Journal of Medicine* **373,** 1340–1349.

Durrleman, V., Nikeghbali, A., and Roncalli, T. (2000). Which copula is the right one?

Eicher, T. S., Papageorgiou, C., and Raftery, A. E. (2011). Default priors and predictive performance in bayesian model averaging, with application to growth determinants. *Journal of Applied Econometrics* **26,** 30–55.

Embrechts, P., Lindskog, F., and McNeil, A. (2001). Modelling dependence with copulas. *Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich* .

Fernández, C., Ley, E., and Steel, M. F. (2001). Benchmark priors for bayesian model averaging. *Journal of Econometrics* **100,** 381–427.

French, J. L., Thomas, N., and Wang, C. (2012). Using historical data with bayesian methods in early clinical trial monitoring. *Statistics in Biopharmaceutical Research* **4,** 384–394.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1,** 515–534.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science* pages 457–472.

Goodman, S. N. and Sladky, J. T. (2005). A bayesian approach to randomized controlled trials in children utilizing information from adults: the case of guillain-barre. *Clinical Trials* **2,** 305–310.

Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika* **82,** 711–732.

Green, P. J. and Hastie, D. I. (2009). Reversible jump mcmc. *Genetics* **155,** 1391–1403.

Hatsukami, D. K., Hertsgaard, L. A., Vogel, R. I., Jensen, J. A., Murphy, S. E., Hecht, S. S., Carmella, S. G., al'Absi, M., Joseph, A. M., and Allen, S. S. (2013). Reduced nicotine content cigarettes and nicotine patch. *Cancer Epidemiology, Biomarkers and Prevention* **22,** 1015–1024.

Hatsukami, D. K., Kotlyar, M., Hertsgaard, L. A., Zhang, Y., Carmella, S. G., Jensen, J. A., Allen, S. S., Shields, P. G., Murphy, S. E., Stepanov, I., and Hecht, S. S. (2010). Reduced nicotine content cigarettes: Effects on toxicant exposure, dependence, and cessation. *Addiction* **105,** 343–355.

Hobbs, B. P., Carlin, B. P., Mandrekar, S. J., and Sargent, D. J. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historic information in clinical trials. *Biometrics* **67,** 1047–1056.

Hobbs, B. P., Carlin, B. P., and Sargent, D. J. (2013). Adaptive adjustment of the randomization ratio using historical control data. *Clinical Trials* **10,** 430–440.

Hobbs, B. P., Sargent, D. J., and Carlin, B. P. (2012). Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Analysis* **7,** 639–674.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science* **14,** 382–417.

Ibrahim, J. G. and Chen, M.-H. (2000). Power prior distributions for regression models. *Statistical Science* **15,** 46–60.

Ippolito, G., Lanini, S., Brouqui, P., Di Caro, A., Vairo, F., Fusco, F. M., Krishna, S., Capobianchi, M. R., Kyobe-Bosa, H., Puro, V., et al. (2016). Non-randomised ebola trialslessons for optimal outbreak research. *The Lancet Infectious Diseases* **16,** 407–408.

Jasra, A., Holmes, C. C., and Stephens, D. A. (2005). Markov chain monte carlo methods and the label switching problem in bayesian mixture modeling. *Statistical Science* pages 50–67.

Kaizer, A. M., Koopmeiners, J. S., and Hobbs, B. P. (2017). Bayesian hierarchical modeling based on multi-source exchangeability. *Biostatistics* .

Kass, R. E. and Natarajan, R. (2006). A default conjugate prior for variance components in generalized linear mixed models. *Bayesian Analysis* **1,** 535–542.

Komárek, A. and Komárková, L. (2013). Clustering for multivariate continuous and discrete longitudinal data. *The Annals of Applied Statistics* **7,** 177–200.

Madigan, D., York, J., and Allard, D. (1995). Bayesian graphical models for discrete data. *International Statistical Review* **63,** 215–232.

Marin, J.-M., Mengersen, K., and Robert, C. P. (2005). Bayesian modelling and inference on mixtures of distributions. *Handbook of statistics* **25,** 459–507.

Morita, S., Thall, P. F., and Müller, P. (2008). Determining the effective sample size of a parametric prior. *Biometrics* **64,** 595–602.

Murray, T. A., Hobbs, B. P., Lystig, T. C., and Carlin, B. P. (2014). Semiparametric bayesian commensurate survival model for post-market medical device surveillance with non-exchangeable historical data. *Biometrics* **70,** 185–191.

Nardone, N., Donny, E. C., Hatsukami, D. K., Koopmeiners, J. S., Murphy, S. E., Strasser, A. A., Tidey, J. W., Vandrey, R., and Benowitz, N. L. (2016). Estimations and predictors of non-compliance in switchers to reduced nicotine content cigarettes. *Addiction* **111,** 2208–2216.

Natarajan, R. and Kass, R. E. (2000). Reference bayesian methods for generalized linear mixed models. *Journal of the American Statistical Association* **95,** 227–237.

Neelon, B. and O'Malley, A. J. (2010). Bayesian analysis using power priors with application to pediatric quality of care. *Journal of Biometrics and Biostatistics* **1,**.

Nelsen, R. B. (2006). *An Introduction to Copulas.* Springer, 2 edition.

Neuenschwander, B., Capkun-Niggli, G., Branson, M., and Spiegelhalter, D. J. (2010). Summarizing historic information on controls in clinical trials. *Clinical Trials* **7,** 5–18.

Ning, J. and Huang, X. (2010). Response-adaptive randomization for clinical trials with adjustment for covariate imbalance. *Statistics in Medicine* **29,** 1761–1768.

Pennello, G. and Thompson, L. (2008). Experience with reviewing bayesian medical device trials. *Journal of Biopharmaceutical Statistics* **18,** 81–115.

Pocock, S. J. (1976). The combination of randomized and historical controls in clinical trials. *Journal of Chronic Diseases* **29,** 175–188.

PREVAIL II Writing Group, for the Multi-National PREVAIL II Study Team (2016). A randomized, controlled trial of zmapp for ebola virus infection. *The New England Journal of Medicine* **375,** 1448.

Proschan, M. A., Dodd, L. E., and Price, D. (2016). Statistical considerations for a trial of ebola virus disease therapeutics. *Clinical Trials* **13,** 39–48.

R Core Team (2013). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology* **25,** 111–164.

Raftery, A. E., Madigan, D., and Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* **92,** 179–191.

Renfro, L. and Sargent, D. (2016). Statistical controversies in clinical research: basket trials, umbrella trials, and other master protocols: a review and examples. *Annals of Oncology* **28,** 34–43.

Rietbergen, C., Klugkist, I., Janssen, K., Moons, K., and Hoijtink, H. (2011). Incorporation of historical data in the analysis of randomized therapeutic trials. *Contemporary Clinical Trials* **32,** 848–855.

Rupprecht, L. E., Koopmeiners, J. S., Dermody, S. S., Oliver, J. A., Al'Absi, M., Benowitz, N. L., Denlinger-Apte, R., Drobes, D. J., Hatsukami, D., McClernon, F. J., et al. (2017). Reducing nicotine exposure results in weight gain in smokers randomised to very low nicotine content cigarettes. *Tobacco Control* pages e43–e48.

Schieffelin, J. S., Shaffer, J. G., Goba, A., Gbakie, M., Gire, S. K., Colubri, A., Sealfon, R. S., Kanneh, L., Moigboi, A., Momoh, M., et al. (2014). Clinical illness and outcomes in patients with ebola in sierra leone. *New England Journal of Medicine* **371,** 2092–2100.

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

Spiegelhalter, D. J. (2001). Bayesian methods for cluster randomized trials with continuous responses. *Statistics in Medicine* **20,** 435–452.

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, volume 13. John Wiley & Sons.

Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62,** 795–809.

Thall, P., Fox, P., and Wathen, J. (2015). Statistical controversies in clinical research: Scientific and ethical problems with adaptive randomization in comparative clinical trials. *Annals of Oncology* **26,** 1621–1628.

Thall, P. F. and Wathen, J. K. (2007). Practical bayesian adaptive randomisation in clinical trials. *European Journal of Cancer* **43,** 859–866.

Tidey, J. W., Pacek, L. R., Koopmeiners, J. S., Vandrey, R., Nardone, N., Drobes, D. J., Benowitz, N. L., Dermody, S. S., Lemieux, A., Denlinger, R. L., et al. (2017). Effects of 6-week use of reduced-nicotine content cigarettes in smokers with and without elevated depressive symptoms. *Nicotine & Tobacco Research* **19,** 59–67.

U.S. Food and Drug Administration (2001). Guidance for industry: E 10 choice of control group and related issues in clinical trials.

Whitehead, J., Valdés-Márquez, E., Johnson, P., and Graham, G. (2008). Bayesian sample size for exploratory clinical trials incorporating historical data. *Statistics in Medicine* **27,** 2307–2327.

WHO (2016). Ebola virus disease. `http://apps.who.int/mediacentre/factsheets/fs103/en/index.html`. Accessed: 2017-01-04.

Williams, R. J., Tse, T., DiPiazza, K., and Zarin, D. A. (2015). Terminated trials in the clinicaltrials.gov results database: Evaluation of availability of primary outcome data and reasons for termination. *PLoS One* **10,**.

Yin, G., Chen, N., and Lee, J. J. (2012). Phase II trial design with bayesian adaptive randomization and predictive probability. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **61,** 219–235.

Zhou, X., Liu, S., Kim, E. S., Herbst, R. S., and Lee, J. J. (2008). Bayesian adaptive design for targeted therapy development in lung cancer–a step toward personalized medicine. *Clinical Trials* **5,** 181–193.

Appendix A

Proof of theorem from Chapter 2

A.1 Proofs for Convergence of Model Weights

Proof of Theorem 1: WLOG assume $n = n_1 = \ldots = n_h$. First, note that we can re-write the model-specific posterior inclusion probabilities as follows:

$$\omega_k = p(\mathbf{\Omega}_k|D) = \frac{p(D|\mathbf{\Omega}_k)\pi(\mathbf{\Omega}_k)}{\sum_{j=I}^{K} p(D|\mathbf{\Omega}_j)\pi(\mathbf{\Omega}_j)} \tag{A.1}$$

$$= \left[1 + \sum_{j \neq k} \frac{p(D|\mathbf{\Omega}_j)\pi(\mathbf{\Omega}_j)}{p(D|\mathbf{\Omega}_k)\pi(\mathbf{\Omega}_k)}\right]^{-1}. \tag{A.2}$$

We will show that $\omega_k \to 1$ when $\mathbf{\Omega}_k = \mathbf{\Omega}_{k^*}$ and $\omega_k \to 0$ when $\mathbf{\Omega}_k \neq \mathbf{\Omega}_{k^*}$ as $n \to \infty$. This can be demonstrated by showing that the ratios of the marginal likelihoods in (A.2) converge to 0 or ∞, as desired, and by showing that $\frac{\pi(\mathbf{\Omega}_j)}{\pi(\mathbf{\Omega}_k)}$ converges to a constant for the two priors under consideration for the MEM framework for all $j \neq k$.

A.1.1 Convergence of the marginal likelihoods

When $\mathbf{\Omega}_k = \mathbf{\Omega}_{k^*}$ we must show that all $\frac{p(D|\mathbf{\Omega}_j)}{p(D|\mathbf{\Omega}_k)}$ terms in (A.2) converge to 0 as $n \to \infty$. To demonstrate the convergence of the model weight when $\mathbf{\Omega}_k \neq \mathbf{\Omega}_{k^*}$ we must show that at least one term of $\frac{p(D|\mathbf{\Omega}_j)}{p(D|\mathbf{\Omega}_k)}$ in (A.2) converges to ∞ and therefore $\omega_k \to 0$ as $n \to \infty$. As long as one term converges to ∞, the convergence of the other terms is trivial since it can

be shown that $\frac{p(D|\mathbf{\Omega}_j)}{p(D|\mathbf{\Omega}_k)}$ must converge within the range of 0 to ∞.

It is helpful to first describe the convergence of the integrated marginal likelihoods. Rewriting (3.5) we have

$$\begin{aligned} p(D|\mathbf{\Omega}_k) = & \frac{(\sqrt{2\pi})^{(H+1)-\sum_{h=1}^{H}\{s_{h,k}\}}}{\sqrt{\left(\frac{1}{v}+\sum_{i=1}^{H}\left\{\frac{s_{i,k}}{v_i}\right\}\right)\left(\prod_{j=1}^{H}\left\{\left[\frac{1}{v_j}\right]^{1-s_{j,k}}\right\}\right)}} \\ & \times \prod_{l=1}^{H}\prod_{l<r}^{H}\left\{\exp\left(-\frac{n}{2}\left[\frac{s_{l,k}(\bar{x}-\bar{x}_l)^2}{\sigma+\sigma_l+\sigma\sigma_l(\sum_{m\neq l}\{s_{m,k}\sigma_m^{-1}\})}\right]\right)\right. \\ & \left.\times\exp\left(-\frac{n}{2}\left[\frac{s_{l,k}s_{r,k}(\bar{x}_l-\bar{x}_r)^2}{\sigma_l+\sigma_r+\sigma_l\sigma_r(\sigma^{-1}+\sum_{p\neq l,r}\{s_{p,k}\sigma_p^{-1}\})}\right]\right)\right\} \end{aligned} \tag{A.3}$$

First we will describe the convergence of the exponential terms. WLOG, consider the first exponential term in (A.3) for any model l and define the following:

$$d = 1+\frac{\sigma\sigma_l}{\sigma+\sigma_l}\left(\sum_{m\neq l}\{s_{m,k}\sigma_m^{-1}\}\right), \tag{A.4}$$

$$d_g = \sigma+\sigma_l, \tag{A.5}$$

$$g = \bar{x}-\bar{x}_l, \tag{A.6}$$

$$g_\mu = \mu-\mu_l. \tag{A.7}$$

Using this notation we can demonstrate the convergence of these exponential terms in the integrated marginal likelihoods. Since g is the sum of normally distributed random variables it can be shown that $\sqrt{n}(g-g_\mu)\sim\mathcal{N}(0,d_g),\forall n$. By adding and subtracting g_μ we see that

$$\exp\left(-\frac{n}{2}\frac{g^2}{dd_g}\right) = \exp\left(-\frac{n}{2}\frac{(g-g_\mu+g_\mu)^2}{dd_g}\right) \tag{A.8}$$

$$= \exp\left\{-\frac{1}{2}\left[\frac{(\sqrt{n}(g-g_\mu))^2}{dd_g}+\frac{ng_\mu^2}{dd_g}+2\sqrt{n}g_\mu\frac{\sqrt{n}(g-g_\mu)}{dd_g}\right]\right\}. \tag{A.9}$$

We can examine the convergence of the terms (A.9) to better understand their convergence behavior. For the first term, $\frac{(\sqrt{n}(g-g_\mu))^2}{dd_g} \sim \Gamma(k=\frac{1}{2}, \theta=\frac{2}{d}), \forall n$. For the remaining terms, if $\mu_l = \mu$, then $g_\mu = 0$ which implies $\frac{ng_\mu^2}{dd_g} + 2\sqrt{n}g_\mu\frac{\sqrt{n}(g-g_\mu)}{dd_g} = 0$. If $\mu_l \neq \mu$, then $g_\mu \neq 0$ which implies $\lim_{n\to\infty} \frac{ng_\mu^2}{dd_g} + 2\sqrt{n}g_\mu\frac{\sqrt{n}(g-g_\mu)}{dd_g} \to \infty$, which can be seen by noting that:

$$\frac{ng_\mu^2}{dd_g} + 2\sqrt{n}g_\mu\frac{\sqrt{n}(g-g_\mu)}{dd_g} = ng_\mu^2\left(\frac{1}{dd_g} + 2\frac{\sqrt{n}(g-g_\mu)}{\sqrt{n}g_\mu dd_g}\right)$$

The overall convergence of the three terms in (A.9) depend on two cases. If $\mu_l = \mu$, then as $n \to \infty$ the only term remaining will be the gamma random variable. If $\mu_l \neq \mu$, then as $n \to \infty$ the dominance of $-\frac{n}{2}\frac{g_\mu^2}{dd_g}$ sends the entire exponential term to 0.

For the the integrated marginal likelihood in (A.3) we then see that the rate of convergence depends on whether or not all sources included in the model are exchangeable with the primary source. If $\mu_h = \mu$ for all sources in $\mathbf{\Omega}_k$, then each exponential term in the marginal likelihood converges to a gamma random variable and the convergence depends on the fractional component such that the overall rate of convergence is $\lim_{n\to\infty} p(D|\mathbf{\Omega}_k) = \mathcal{O}\left(\left[\sqrt{n^{(H+1)-\sum_{h=1}^{H}\{s_{h,k}\}}}\right]^{-1}\right)$. If at least one source is not exchangeable in $\mathbf{\Omega}_k$, then at least one exponential term converges to 0 and the overall rate of convergence is $\lim_{n\to\infty} p(D|\mathbf{\Omega}_k) = \mathcal{O}\left(exp(-n)\right)$. Both cases ultimately converge to 0, but at different rates.

A.1.2 Convergence of the ratio of marginal likelihoods

If $\mathbf{\Omega}_k = \mathbf{\Omega}_{k^*}$ we must show that all $\frac{p(D|\mathbf{\Omega}_j)}{p(D|\mathbf{\Omega}_k)}$ terms in (A.2) converge to 0 as $n \to \infty$ so that $\lim_{n\to\infty} p(\mathbf{\Omega}_k|D) = 1$. This can be exhaustively shown by considering two cases. First, $\mathbf{\Omega}_j$ from (A.2) only contains a subset of the exchangeable sources which implies

$$\lim_{n\to\infty} \frac{p(D|\Omega_j)}{p(D|\Omega_k)} = \mathcal{O}\left(\sqrt{n^{\sum_{h=1}^{H}\{s_{h,j}\}-\sum_{h=1}^{H}\{s_{h,k}\}}}\right), \tag{A.10}$$

and converges to 0. Alternatively, $\mathbf{\Omega}_j$ may be a model which contains nonexchangeable sources. In this case,

$$\lim_{n\to\infty} \frac{p(D|\Omega_j)}{p(D|\Omega_k)} = \mathcal{O}\left(\exp(-n)\right), \tag{A.11}$$

and converges to 0. Therefore if $\mathbf{\Omega}_k = \mathbf{\Omega}_{k^*}$, all terms in (A.2) converge to 0 and $\lim_{n\to\infty} p(\mathbf{\Omega}_k|D) = 1$.

If $\mathbf{\Omega}_k \neq \mathbf{\Omega}_{k^*}$ we must show that at least one term in (A.2) converges to ∞ so that it sends the posterior model weight to 0. Similar to before, we can show this holds for any model by considering two cases. First, consider the case where $\mathbf{\Omega}_k$ only contains a subset of the exchangeable sources in $\mathbf{\Omega}_{k^*}$. Examining the convergence behavior of comparing these models results in

$$\lim_{n\to\infty} \frac{p(D|\Omega_{k^*})}{p(D|\Omega_k)} = \mathcal{O}\left(\sqrt{n^{\sum_{h=1}^{H}\{s_{h,k^*}\}-\sum_{h=1}^{H}\{s_{h,k}\}}}\right), \tag{A.12}$$

which will approach ∞ as $n \to \infty$ since $\mathbf{\Omega}_{k^*}$ includes a larger number of exchangeable sources than $\mathbf{\Omega}_k$. Alternatively, $\mathbf{\Omega}_k$ may contain nonexchangeable sources which, when compared to $\mathbf{\Omega}_{k^*}$, results in convergence behavior of

$$\lim_{n\to\infty} \frac{p(D|\Omega_{k^*})}{p(D|\Omega_k)} = \mathcal{O}\left(\exp(n)\right), \tag{A.13}$$

which will approach ∞ as $n \to \infty$. Therefore, if $\mathbf{\Omega}_k \neq \mathbf{\Omega}_{k^*}$, at least one term will converge to ∞ in (A.2) with the all other terms converging between 0 and ∞, therefore $\lim_{n\to\infty} p(\mathbf{\Omega}_k|D) = 0$.

A.1.3 Convergence of the priors

The results of the previous sub-section show that the ratio of marginal likelihoods in (A.2) will converge appropriately. Recall the MEMs assume the supplemental sources are independent such that $\pi(\mathbf{\Omega}_k) = \pi(S_1 = s_{1,k}, ..., S_H = s_{H,k}) = \pi(S_1 = s_{1,k}) \times \cdots \times \pi(S_H = s_{H,k})$ so that priors are placed are the sources rather than the model alone. This results in ratios of source priors on inclusion and exclusion, e.g. $\frac{\pi(S_h=0)}{\pi(S_h=1)}$ or $\frac{\pi(S_h=1)}{\pi(S_h=0)}$. To show that the convergence of these priors do not impact the convergence results already shown for marginal likelihoods, we will demonstrate our proposed priors converge to constant values and therefore do not change the rates of convergence proved previously.

First, we know that $\frac{\pi_e(S_h=0)}{\pi_e(S_h=1)}$ or $\frac{\pi_e(S_h=1)}{\pi_e(S_h=0)}$ is a fixed constant and we can therefore conclude that the posterior source-specific inclusion probabilities will be consistent for π_e. Similarly, we know that $\frac{\pi_n(S_h=0)}{\pi_n(S_h=1)}$ or $\frac{\pi_n(S_h=1)}{\pi_n(S_h=0)}$ converges to a constant, which confirms that

the posterior source-specific inclusion probabilities will be consistent for π_n. The last result for π_n can be seen by noting that the largest term in both $\pi_n(S_h = 0)$ and $\pi_n(S_h = 1)$ is $\sqrt{n^H}$, which implies that the ratio will converge to a constant.

Since the terms in (A.2) all converge as desired, MEM model weights are asymptotically consistent.

Appendix B

Additional simulation results for Chapter 2

B.1 Simulation Operating Characteristics

This section provides additional results for the simulation study presented in Section 4. Figure B.1 shows bias at varying levels of μ. It can observed that MEMs are incorporating information in a region around supplemental means whereas CP struggles to identify what supplemental information to integrate. In the presence of heterogeneity of supplementary sources the SHM shows a preference for not borrowing information and has little bias as a consequence. Figure B.2 shows MSE represents similar trends with the MEMs incorporating information in regions around supplemental source means whereas the CP and SHM struggle to incorporate supplemental data when they are heterogeneous. Finally, Figure B.3 represents the coverage of 95% HPD interval for each scenario. These figures indicate that higher coverage is attained at locations where borrowing occurs, but the location of borrowing is not necessarily in the region of a supplemental source for CP in scenarios with heterogeneous supplementary data.

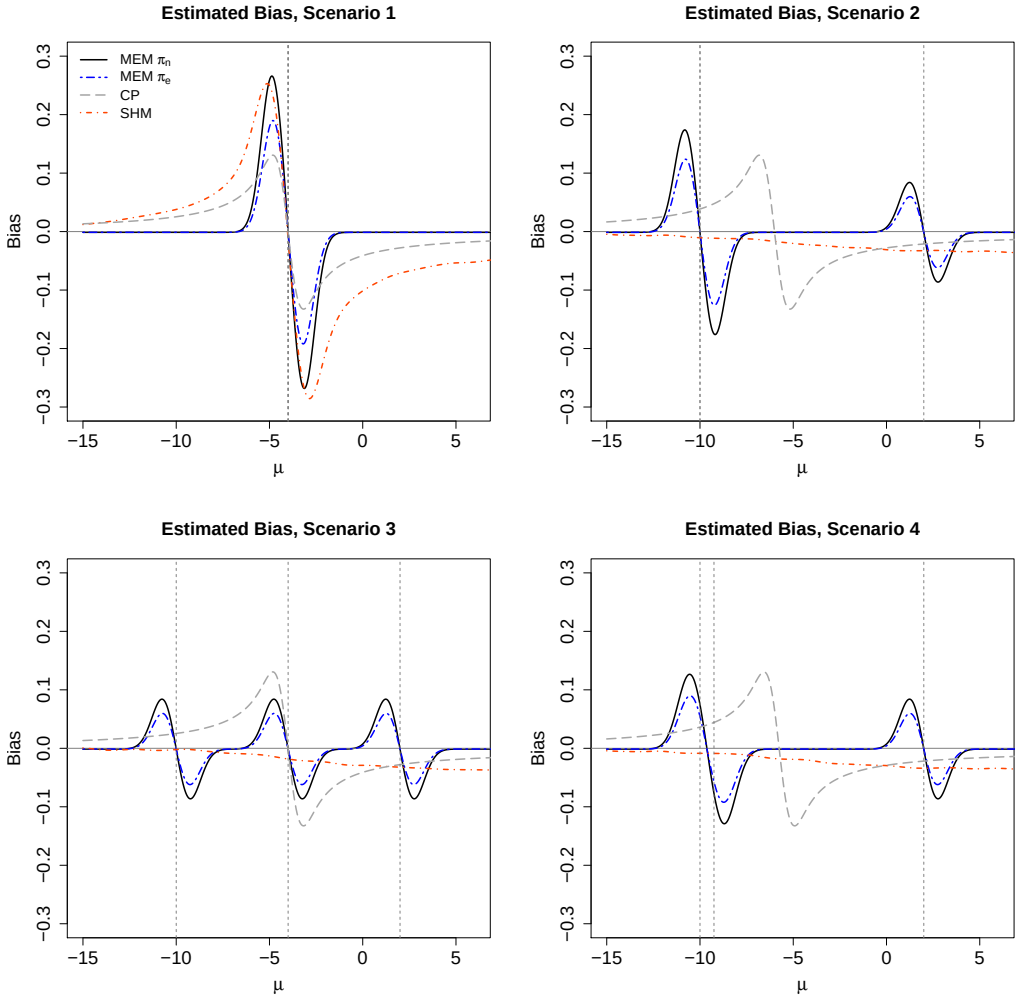

Figure B.1: Bias for Each Scenario

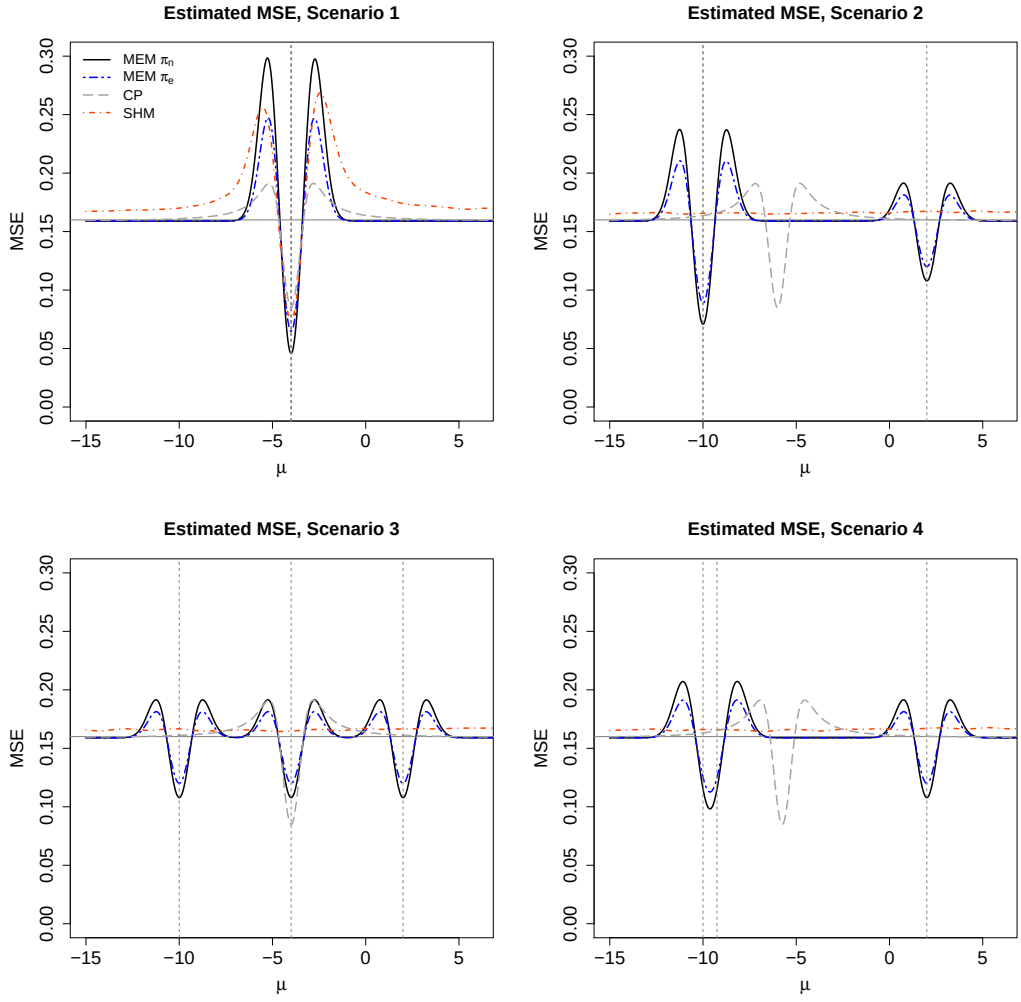

Figure B.2: MSE for Each Scenario

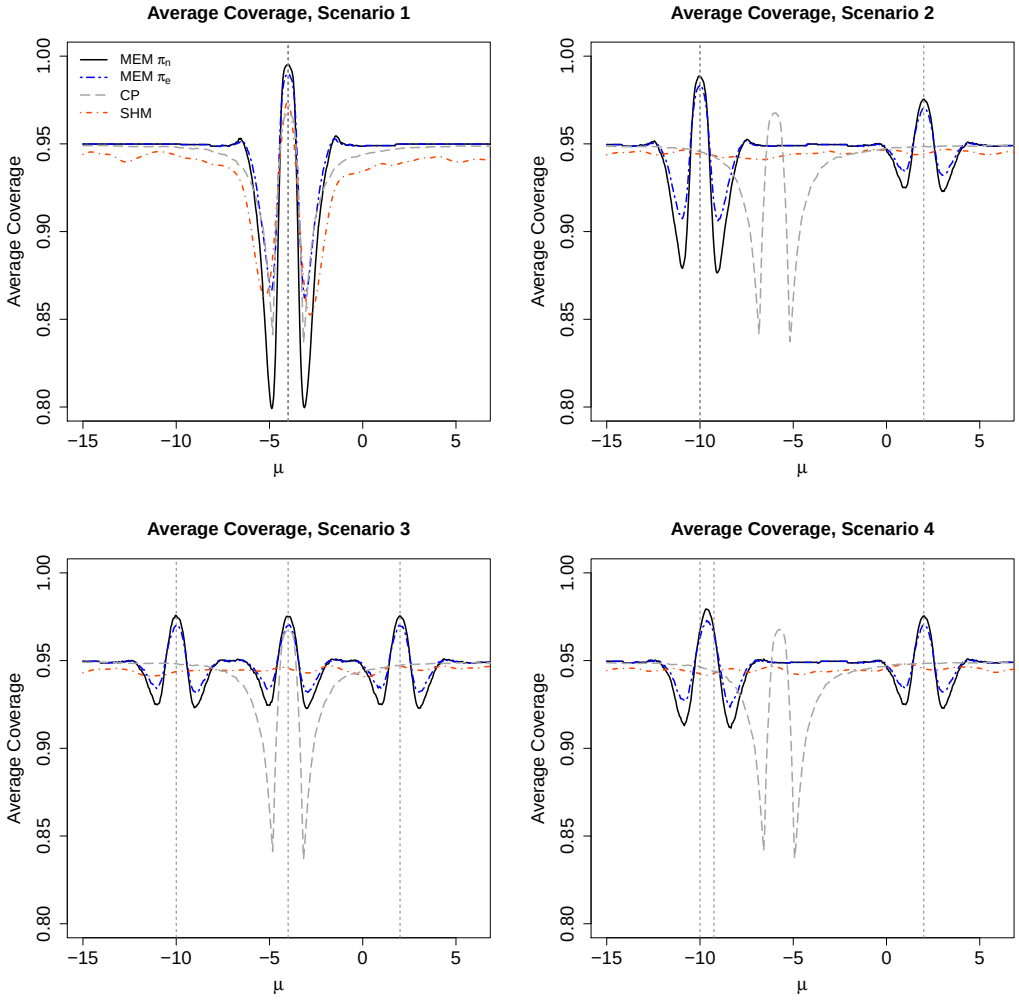

Figure B.3: Coverage for Each Scenario

Appendix C

Additional simulation results for Chapter 3

C.1 Additional simulation results for proposed multi-source adaptive platform design

The results presented in the paper demonstrated the performance of the posterior probability thresholds optimized for performance under the constant underlying mortality scenario. In the supplementary materials we provide results for these scenarios under the posterior probability thresholds calibrated to maintain the power of the constant mortality while reducing the inflation of the type-I error rates of the varying underlying mortality. We also provide results which look at some scenarios where a stronger effective treatment is included at RR=0.5 as compared to the results for RR=0.7. Finally, results are presented which explore how operating characteristics change as various parameters such as the burn-in period or $c = 0.20$ for π_{EB_c}.

C.1.1 Both PP thresholds

Tables C.1 and C.2 examine the performance using the same parameters from the manuscript of $n_{burn} = 60$ and $B = 5$ with adaptive randomization and interim monitoring, but use the

thresholds calibrated to minimize the inflation of the type-I error rate in the varying underlying mortality case of Table C.2 while maintaining a power in the constant underlying mortality case of Table C.1 for the MEMs similar to that of PREVAIL II (P-II).

In Table C.1 we see that the type-I error rates are now much lower than PREVAIL II across segments 2-5 when supplemental information is available to integrate to the analysis as compared to the results for the posterior probability thresholds calibrated to maintain an average type-I error rate across scenarios of 0.025. However, this reduction in type-I error rates results in a reduction of power such that performance for $\pi_{\text{EB}_{10}}$ and π_e is similar to P-II, whereas POOL now has less power than P-II. All MEMs still randomize more patients to the treatment arm and see an increase in the mean proportion surviving in the non-null segments compared to P-II.

The varying underlying mortality scenario in Table C.2 shows similar results to the case with posterior probability thresholds calibrated to the constant scenario. There is an inflation of the average type-I error within a segment for each MEM compared to P-II. However the inflation is less than that observed when the thresholds are optimized for the constant mortality case and still demonstrate an improvement in power to detect an effective treatment. Again, more patients are randomized to the treatment arm in segments 2-5 and there is an increase in the proportion surviving the non-null segments.

The results of calibrating the posterior probability thresholds to both scenarios for underlying mortality represents similar overall trends to operating characteristics as seen in the calibration of thresholds to just the constant case's average type-I error rate. However, π_e does have more acceptable levels of inflation and may be appropriate if one desires to maximize the number of patients randomized to the treatment arm while maintaining operating characteristics. $\pi_{\text{EB}_{10}}$ also demonstrates desirable characteristics as before, but POOL still represents an undesirable trade-off wherein there is drastically increased power compared to P-II in the varying mortality case but with very high type-I error rates.

		Probability Reject in Segment					Mean (sd)	Mean (sd)	Mean (sd)
	Trt	1	2	3	4	5	N	Prop Trt	Prop Surv
P-II	S0	0.032	0.028	0.029	0.031	0.029	996 (25.58)	0.5 (0)	0.600 (0.017)
	S2	0.032	0.432	0.028	0.029	0.028	988 (38.41)	0.5 (0)	0.659 (0.036)
	S3	0.032	0.028	0.431	0.029	0.030	988 (38.89)	0.5 (0)	0.659 (0.036)
	S4	0.032	0.028	0.029	0.434	0.028	988 (38.79)	0.5 (0)	0.659 (0.036)
	S5	0.032	0.028	0.029	0.031	0.441	988 (38.78)	0.5 (0)	0.659 (0.036)
$\pi_{\text{EB}_{10}}$	S0	0.027	0.023	0.018	0.018	0.014	998 (14.97)	0.656 (0.028)	0.600 (0.017)
	S2	0.027	0.446	0.017	0.017	0.013	990 (32.02)	0.643 (0.032)	0.669 (0.035)
	S3	0.027	0.023	0.435	0.018	0.014	990 (31.46)	0.638 (0.036)	0.676 (0.035)
	S4	0.027	0.023	0.018	0.452	0.016	990 (31.13)	0.641 (0.033)	0.682 (0.035)
	S5	0.027	0.023	0.018	0.018	0.453	991 (31.02)	0.655 (0.028)	0.686 (0.035)
π_e	S0	0.027	0.013	0.009	0.007	0.006	998 (14.72)	0.798 (0.020)	0.600 (0.017)
	S2	0.027	0.429	0.005	0.004	0.004	990 (32.14)	0.787 (0.032)	0.683 (0.035)
	S3	0.027	0.013	0.418	0.004	0.003	989 (32.92)	0.785 (0.038)	0.696 (0.035)
	S4	0.027	0.013	0.009	0.415	0.004	988 (34.12)	0.786 (0.034)	0.701 (0.035)
	S5	0.027	0.013	0.009	0.007	0.452	986 (35.64)	0.794 (0.022)	0.701 (0.035)
POOL	S0	0.027	0.004	0.002	0.004	0.002	998 (15.65)	0.825 (0.007)	0.600 (0.017)
	S2	0.027	0.342	0.001	0.001	0.001	984 (37.27)	0.817 (0.023)	0.691 (0.036)
	S3	0.027	0.004	0.258	0.001	0.001	981 (39.39)	0.814 (0.027)	0.701 (0.036)
	S4	0.027	0.004	0.002	0.354	0.001	980 (39.72)	0.814 (0.027)	0.702 (0.036)
	S5	0.027	0.004	0.002	0.004	0.306	978 (40.58)	0.821 (0.012)	0.702 (0.036)

Table C.1: Constant underlying mortality scenario results using posterior probability thresholds calibrated for both constant and varying mortality cases with RR=0.7.

		Probability Reject in Segment					Mean (sd)	Mean (sd)	Mean (sd)
	Trt	1	2	3	4	5	N	Prop Trt	Prop Surv
P-II	S0	0.028	0.030	0.032	0.027	0.025	997 (22.73)	0.5 (0)	0.580 (0.017)
	S2	0.028	0.764	0.027	0.025	0.027	972 (51.38)	0.5 (0)	0.482 (0.041)
	S3	0.028	0.030	0.555	0.026	0.026	984 (42.33)	0.5 (0)	0.591 (0.038)
	S4	0.028	0.030	0.032	0.386	0.026	990 (35.35)	0.5 (0)	0.693 (0.035)
	S5	0.028	0.030	0.032	0.027	0.235	994 (27.80)	0.5 (0)	0.804 (0.029)
$\pi_{\text{EB}_{10}}$	S0	0.027	0.035	0.034	0.038	0.038	998 (16.16)	0.542 (0.015)	0.580 (0.017)
	S2	0.027	0.773	0.028	0.039	0.043	971 (48.00)	0.556 (0.019)	0.487 (0.038)
	S3	0.027	0.035	0.561	0.033	0.043	983 (39.20)	0.55 (0.018)	0.597 (0.036)
	S4	0.027	0.035	0.034	0.425	0.035	990 (31.19)	0.545 (0.016)	0.699 (0.032)
	S5	0.027	0.035	0.034	0.038	0.272	994 (24.99)	0.542 (0.015)	0.807 (0.027)
π_e	S0	0.027	0.059	0.085	0.084	0.092	994 (24.12)	0.658 (0.04)	0.580 (0.017)
	S2	0.027	0.785	0.036	0.085	0.137	960 (53.46)	0.683 (0.041)	0.500 (0.038)
	S3	0.027	0.059	0.627	0.037	0.109	972 (47.21)	0.667 (0.041)	0.610 (0.036)
	S4	0.027	0.059	0.085	0.481	0.058	979 (41.91)	0.658 (0.040)	0.710 (0.032)
	S5	0.027	0.059	0.085	0.084	0.361	985 (37.07)	0.655 (0.039)	0.815 (0.027)
POOL	S0	0.027	0.161	0.535	0.406	0.392	895 (45.15)	0.764 (0.034)	0.574 (0.023)
	S2	0.027	0.985	0.030	0.232	0.495	856 (49.39)	0.739 (0.033)	0.510 (0.047)
	S3	0.027	0.161	0.946	0.040	0.259	874 (37.51)	0.733 (0.024)	0.619 (0.048)
	S4	0.027	0.161	0.535	0.777	0.103	879 (40.85)	0.741 (0.031)	0.717 (0.041)
	S5	0.027	0.161	0.535	0.406	0.738	857 (47.80)	0.753 (0.035)	0.822 (0.034)

Table C.2: Varying underlying mortality scenario results using posterior probability thresholds calibrated for both constant and varying mortality cases with RR=0.7.

C.1.2 RR=0.5 results

Tables C.3 and C.4 examine the operating characteristics if the actual RR=0.5 instead of RR=0.7. Since $\pi_{\text{EB}_{10}}$ appears to have the more desirable traits of the explored MEM priors, we include both cases with the posterior probability thresholds calibrated for RR=0.7 under both the constant mortality only ($\pi_{\text{EB}_{10}}$-C) and attempting to control the trade-off between the constant mortality type-I error and varying mortality power ($\pi_{\text{EB}_{10}}$-B).

Under the assumption of constant underlying mortality in Table C.3 we see both MEMs

result in increases of power relative to P-II and similar or lower average type-I error rates across the scenarios. There are also more patients randomized to the treatment arm and corresponding increases in proportion surviving the non-null arms.

In Table C.4 we see gains in power for each segment with acceptable levels of inflation to the type-I error rate. We also can note that the trials are terminating earlier for the MEMs than the P-II cases as seen by the marginally lower median N for the overall trial.

These results suggest that all methods perform more robustly when the effectiveness of the treatment is stronger (i.e., there is larger power and comparable type-I error rates). Utilizing MEMs still show improvements by improving power while reducing the inflation of the type-I error rate in the varying mortality case, by improving the proportion assigned to the treatment arm, and by increasing the survivorship of patients.

		Probability Reject in Segment					Mean (sd)	Mean (sd)	Mean (sd)
	Trt	1	2	3	4	5	N	Prop Trt	Prop Surv
P-II	S0	0.032	0.028	0.029	0.031	0.029	996 (25.58)	0.5 (0)	0.600 (0.017)
	S2	0.032	0.876	0.027	0.029	0.026	961 (55.20)	0.5 (0)	0.698 (0.041)
	S3	0.032	0.028	0.876	0.027	0.026	961 (55.94)	0.5 (0)	0.697 (0.041)
	S4	0.032	0.028	0.029	0.875	0.025	960 (56.08)	0.5 (0)	0.697 (0.041)
	S5	0.032	0.028	0.029	0.031	0.880	960 (56.67)	0.5 (0)	0.697 (0.041)
$\pi_{\text{EB}_{10}}$-C	S0	0.027	0.026	0.026	0.030	0.026	998 (14.97)	0.655 (0.029)	0.600 (0.017)
	S2	0.027	0.905	0.027	0.030	0.027	959 (51.55)	0.627 (0.03)	0.711 (0.039)
	S3	0.027	0.026	0.922	0.033	0.031	958 (51.87)	0.609 (0.032)	0.720 (0.040)
	S4	0.027	0.026	0.026	0.943	0.035	957 (52.36)	0.615 (0.030)	0.729 (0.040)
	S5	0.027	0.026	0.026	0.030	0.948	957 (52.67)	0.646 (0.030)	0.736 (0.041)
$\pi_{\text{EB}_{10}}$-B	S0	0.027	0.023	0.018	0.018	0.014	998 (14.97)	0.656 (0.028)	0.600 (0.017)
	S2	0.027	0.894	0.018	0.018	0.016	959 (51.57)	0.628 (0.03)	0.711 (0.039)
	S3	0.027	0.023	0.893	0.020	0.018	958 (51.91)	0.61 (0.033)	0.721 (0.040)
	S4	0.027	0.023	0.018	0.909	0.021	957 (52.38)	0.616 (0.031)	0.729 (0.040)
	S5	0.027	0.023	0.018	0.018	0.912	956 (52.73)	0.646 (0.029)	0.736 (0.041)

Table C.3: Constant underlying mortality scenario results with RR=0.5 for the PREVAIL II design (P-II) and $\pi_{\text{EB}_{10}}$ with posterior probability thresholds calibrated under the constant mortality only (-C) or both constant and varying mortality (-B).

		Probability Reject in Segment					Mean (sd)	Mean (sd)	Mean (sd)
	Trt	1	2	3	4	5	N	Prop Trt	Prop Surv
P-II	S0	0.028	0.030	0.032	0.027	0.025	997 (22.73)	0.5 (0)	0.580 (0.017)
	S2	0.028	0.994	0.026	0.025	0.025	912 (59.09)	0.5 (0)	0.540 (0.052)
	S3	0.028	0.030	0.945	0.027	0.025	945 (59.83)	0.5 (0)	0.636 (0.046)
	S4	0.028	0.030	0.032	0.824	0.025	968 (51.98)	0.5 (0)	0.727 (0.038)
	S5	0.028	0.030	0.032	0.027	0.589	986 (37.84)	0.5 (0)	0.826 (0.030)
$\pi_{\text{EB}_{10}}$-C	S0	0.027	0.040	0.048	0.058	0.061	998 (16.20)	0.543 (0.016)	0.580 (0.017)
	S2	0.027	0.995	0.037	0.055	0.075	910 (53.01)	0.568 (0.020)	0.546 (0.048)
	S3	0.027	0.040	0.964	0.044	0.069	941 (55.00)	0.554 (0.017)	0.644 (0.042)
	S4	0.027	0.040	0.048	0.889	0.052	964 (49.76)	0.545 (0.016)	0.735 (0.034)
	S5	0.027	0.040	0.048	0.058	0.699	983 (38.10)	0.542 (0.015)	0.831 (0.027)
$\pi_{\text{EB}_{10}}$-B	S0	0.027	0.035	0.034	0.038	0.038	998 (16.16)	0.542 (0.015)	0.580 (0.017)
	S2	0.027	0.995	0.026	0.034	0.047	910 (53.01)	0.568 (0.02)	0.546 (0.048)
	S3	0.027	0.035	0.948	0.027	0.041	941 (55.00)	0.554 (0.017)	0.644 (0.042)
	S4	0.027	0.035	0.034	0.842	0.030	964 (49.76)	0.545 (0.015)	0.735 (0.034)
	S5	0.027	0.035	0.034	0.038	0.616	983 (38.10)	0.541 (0.015)	0.831 (0.027)

Table C.4: Varying underlying mortality scenario results with RR=0.5 for the PREVAIL II design (P-II) and $\pi_{\text{EB}_{10}}$ with posterior probability thresholds calibrated under the constant mortality only (-C) or both constant and varying mortality (-B).

C.1.3 Exploring different parameter values

To explore the sensitivity of our choices for various parameters we present a selection of results assuming posterior probability thresholds calibrated for the constant case only. We explore what would happen if we decrease the burn-in period to 30 patients while maintaining similarly sized blocks ($\mathcal{B}$=6) or if we increase the burn-in period to 90 patients while maintaining similarly sized blocks ($\mathcal{B}$=4). We explore increasing the constrained EB approach to 0.20 from 0.10 with calibrated posterior probability thresholds of (0.975,0.96375,0.96125,0.95626,0.95) to examine the operating characteristics under a slightly less conservative prior. We also explore the $\pi_{\text{EB}_{10}}$ with no AR but maintaining interim monitoring (IM) and also with no IM while maintaining AR. The results for $\pi_{\text{EB}_{10}}$ from the main manuscript are provided for ease of comparison.

We see in Table C.5 for the constant underlying mortality that the operating characteristics for $\pi_{\text{EB}_{10}}$ are not greatly altered by the changes of burn-in period. In fact, as n_{burn} approaches the maximum value of 200, the results will begin to converge to the values seen for $\pi_{\text{EB}_{10}}$ with no AR since larger burn-in periods result in fewer patients being adaptively randomized. $\pi_{\text{EB}_{20}}$ results in slight increases of power and comparable type-I error rates, but more patients are randomized to the treatment arm since more supplemental information is potentially integrated into the analysis. Finally, when RR=0.7, $\pi_{\text{EB}_{10}}$ with no IM shoes similar performance to $\pi_{\text{EB}_{10}}$ with IM and AR since the segments rarely terminate early due to the moderate strength of the non-null treatment.

Extremely similar results are seen in Table C.6 for the varying underlying mortality. In this case we see similar levels of inflation to the type-I error with comparable power. In general, we see comparisons of performance of $\pi_{\text{EB}_{10}}$ to the alternative parameter values which follow similar patterns as seen in Table C.5.

		Probability Reject in Segment					Mean (sd)	Mean (sd)	Mean (sd)
	Trt	1	2	3	4	5	N	Prop Trt	Prop Surv
$\pi_{\text{EB}_{10}}$	S0	0.027	0.026	0.026	0.030	0.026	998 (14.97)	0.655 (0.029)	0.600 (0.017)
	S2	0.027	0.470	0.025	0.028	0.024	990 (32.03)	0.642 (0.032)	0.669 (0.035)
	S3	0.027	0.026	0.509	0.031	0.025	990 (31.49)	0.634 (0.035)	0.676 (0.035)
	S4	0.027	0.026	0.026	0.548	0.028	990 (31.09)	0.638 (0.032)	0.682 (0.035)
	S5	0.027	0.026	0.026	0.030	0.560	991 (30.96)	0.654 (0.029)	0.686 (0.036)
$\pi_{\text{EB}_{10}}$ n_{burn}=30	S0	0.028	0.026	0.026	0.029	0.025	998 (15.03)	0.655 (0.028)	0.600 (0.017)
	S2	0.028	0.471	0.024	0.027	0.023	990 (31.11)	0.642 (0.032)	0.669 (0.035)
	S3	0.028	0.026	0.504	0.029	0.026	991 (29.97)	0.635 (0.034)	0.676 (0.035)
	S4	0.028	0.026	0.026	0.543	0.028	991 (29.86)	0.638 (0.031)	0.682 (0.035)
	S5	0.028	0.026	0.026	0.029	0.560	991 (30.14)	0.654 (0.029)	0.687 (0.035)
$\pi_{\text{EB}_{10}}$ n_{burn}=90	S0	0.027	0.026	0.025	0.028	0.024	998 (14.74)	0.656 (0.029)	0.600 (0.017)
	S2	0.027	0.473	0.023	0.027	0.022	990 (30.89)	0.642 (0.032)	0.669 (0.035)
	S3	0.027	0.026	0.508	0.028	0.025	990 (30.40)	0.635 (0.036)	0.675 (0.035)
	S4	0.027	0.026	0.025	0.550	0.028	990 (30.98)	0.637 (0.033)	0.681 (0.035)
	S5	0.027	0.026	0.025	0.028	0.558	990 (30.91)	0.654 (0.030)	0.686 (0.035)
$\pi_{\text{EB}_{20}}$	S0	0.027	0.031	0.027	0.026	0.025	998 (14.97)	0.726 (0.035)	0.600 (0.017)
	S2	0.027	0.522	0.025	0.024	0.020	990 (32.12)	0.709 (0.041)	0.674 (0.035)
	S3	0.027	0.031	0.551	0.025	0.023	990 (31.86)	0.700 (0.044)	0.684 (0.036)
	S4	0.027	0.031	0.027	0.584	0.025	990 (31.23)	0.704 (0.041)	0.692 (0.035)
	S5	0.027	0.031	0.027	0.026	0.619	990 (31.61)	0.724 (0.036)	0.697 (0.036)
$\pi_{\text{EB}_{10}}$ No AR	S0	0.029	0.027	0.029	0.037	0.035	999 (12.73)	0.5 (0)	0.600 (0.017)
	S2	0.029	0.476	0.027	0.035	0.034	991 (29.19)	0.5 (0)	0.660 (0.035)
	S3	0.029	0.027	0.514	0.036	0.034	992 (28.15)	0.5 (0)	0.660 (0.035)
	S4	0.029	0.027	0.029	0.566	0.034	992 (27.89)	0.5 (0)	0.660 (0.035)
	S5	0.029	0.027	0.029	0.037	0.591	992 (27.82)	0.5 (0)	0.661 (0.035)
$\pi_{\text{EB}_{10}}$ No IM	S0	0.025	0.025	0.025	0.028	0.026	1000 (0.00)	0.656 (0.028)	0.600 (0.017)
	S2	0.025	0.470	0.025	0.026	0.021	1000 (0.00)	0.646 (0.028)	0.669 (0.033)
	S3	0.025	0.025	0.505	0.028	0.024	1000 (0.00)	0.638 (0.030)	0.676 (0.033)
	S4	0.025	0.025	0.025	0.547	0.028	1000 (0.00)	0.641 (0.028)	0.682 (0.033)
	S5	0.025	0.025	0.025	0.028	0.554	1000 (0.00)	0.656 (0.028)	0.687 (0.034)

Table C.5: Constant underlying mortality scenario results assuming RR=0.7 with $\pi_{\text{EB}_{10}}$ results provided for direct comparison with increased and decreased values for n_{burn}, increased c to 0.20, no AR, and interim monitoring (IM) only for AR without possibility of terminating early.

		Probability Reject in Segment					Mean (sd)	Mean (sd)	Mean (sd)
	Trt	1	2	3	4	5	N	Prop Trt	Prop Surv
$\pi_{\text{EB}_{10}}$	S0	0.027	0.040	0.048	0.058	0.061	998 (16.20)	0.543 (0.016)	0.580 (0.017)
	S2	0.027	0.781	0.042	0.063	0.070	971 (47.98)	0.557 (0.019)	0.487 (0.038)
	S3	0.027	0.040	0.637	0.052	0.069	983 (39.13)	0.551 (0.018)	0.597 (0.036)
	S4	0.027	0.040	0.048	0.511	0.057	990 (31.19)	0.546 (0.016)	0.699 (0.032)
	S5	0.027	0.040	0.048	0.058	0.354	994 (24.99)	0.542 (0.016)	0.807 (0.027)
$\pi_{\text{EB}_{10}}$ n_{burn}=30	S0	0.025	0.039	0.048	0.060	0.063	998 (16.66)	0.544 (0.016)	0.580 (0.017)
	S2	0.025	0.780	0.042	0.063	0.074	971 (47.49)	0.558 (0.019)	0.488 (0.037)
	S3	0.025	0.039	0.635	0.051	0.068	983 (38.69)	0.552 (0.018)	0.597 (0.035)
	S4	0.025	0.039	0.048	0.510	0.059	990 (32.16)	0.548 (0.017)	0.699 (0.032)
	S5	0.025	0.039	0.048	0.060	0.363	994 (25.55)	0.544 (0.016)	0.807 (0.027)
$\pi_{\text{EB}_{10}}$ n_{burn}=90	S0	0.024	0.041	0.047	0.060	0.062	998 (14.97)	0.544 (0.016)	0.580 (0.017)
	S2	0.024	0.783	0.041	0.062	0.069	973 (45.90)	0.557 (0.019)	0.487 (0.037)
	S3	0.024	0.041	0.636	0.052	0.068	984 (36.73)	0.551 (0.018)	0.597 (0.036)
	S4	0.024	0.041	0.047	0.512	0.058	991 (29.32)	0.547 (0.017)	0.698 (0.032)
	S5	0.024	0.041	0.047	0.060	0.353	994 (23.06)	0.543 (0.016)	0.807 (0.027)
$\pi_{\text{EB}_{20}}$	S0	0.027	0.062	0.074	0.082	0.095	997 (17.49)	0.572 (0.027)	0.580 (0.017)
	S2	0.027	0.825	0.058	0.092	0.121	969 (49.26)	0.595 (0.031)	0.490 (0.037)
	S3	0.027	0.062	0.676	0.065	0.112	981 (40.89)	0.584 (0.029)	0.600 (0.035)
	S4	0.027	0.062	0.074	0.554	0.085	988 (33.39)	0.577 (0.028)	0.702 (0.032)
	S5	0.027	0.062	0.074	0.082	0.425	992 (27.30)	0.572 (0.027)	0.809 (0.027)
$\pi_{\text{EB}_{10}}$ No AR	S0	0.025	0.036	0.043	0.054	0.054	998 (15.12)	0.5 (0)	0.580 (0.017)
	S2	0.025	0.772	0.038	0.054	0.058	975 (44.76)	0.5 (0)	0.482 (0.039)
	S3	0.025	0.036	0.619	0.048	0.055	986 (36.32)	0.5 (0)	0.592 (0.037)
	S4	0.025	0.036	0.043	0.495	0.052	991 (29.08)	0.5 (0)	0.694 (0.034)
	S5	0.025	0.036	0.043	0.054	0.342	995 (22.49)	0.5 (0)	0.804 (0.029)
$\pi_{\text{EB}_{10}}$ No IM	S0	0.025	0.039	0.049	0.058	0.061	1000 (0.00)	0.543 (0.016)	0.580 (0.017)
	S2	0.025	0.778	0.045	0.067	0.069	1000 (0.00)	0.561 (0.021)	0.489 (0.033)
	S3	0.025	0.039	0.632	0.054	0.068	1000 (0.00)	0.553 (0.020)	0.598 (0.032)
	S4	0.025	0.039	0.049	0.510	0.057	1000 (0.00)	0.548 (0.018)	0.699 (0.031)
	S5	0.025	0.039	0.049	0.058	0.354	1000 (0.00)	0.543 (0.016)	0.807 (0.027)

Table C.6: Varying underlying mortality scenario results assuming RR=0.7 with $\pi_{\text{EB}_{10}}$ results provided for direct comparison with increased and decreased values for n_{burn}, increased c to 0.20, no AR, and interim monitoring (IM) only for AR without possibility of terminating early.

C.1.4 Two effective therapeutics

Results for trials with two effect therapeutics are presented below. Simulations with an effective drug in the first segment were excluded since each approach would have identical results given the fact that no supplemental information is available until the second segment. Posterior probability thresholds calibrated for the constant mortality case in Table 3.2 used.

Operating characteristics and trial properties are extremely similar to results for simulations with only one effective treatment for Tables C.7 and C.8. In the constant case we similar type-I error rates between PREVAIL II and the MEM priors, but we see increased power to detect the effective therapeutic compared to PREVAIL II. In the varying case we see minor inflation to the type-1 error rates for $\pi_{\text{EB}_{10}}$ compared to PREVAIL II, but π_e and POOL have unacceptable levels. Power is, again, increased for all MEM priors compared to PREVAIL II.

		Probability Reject in Segment					Mean (sd)	Mean (sd)	Mean (sd)
	Trt	1	2	3	4	5	N	Prop Trt	Prop Surv
P-II	S0	0.032	0.028	0.029	0.031	0.029	996 (25.58)	0.5 (0)	0.600 (0.017)
	2,3	0.032	0.432	0.368	0.029	0.028	982 (45.85)	0.5 (0)	0.683 (0.036)
	2,4	0.032	0.432	0.028	0.371	0.027	982 (45.84)	0.5 (0)	0.677 (0.043)
	2,5	0.032	0.432	0.028	0.029	0.378	982 (46.03)	0.5 (0)	0.677 (0.043)
	3,4	0.032	0.028	0.431	0.368	0.027	982 (45.99)	0.5 (0)	0.698 (0.059)
	3,5	0.032	0.028	0.431	0.029	0.378	981 (46.69)	0.5 (0)	0.699 (0.059)
	4,5	0.032	0.028	0.029	0.434	0.376	982 (46.07)	0.5 (0)	0.692 (0.072)
$\pi_{\text{EB}_{10}}$	S0	0.027	0.026	0.026	0.030	0.026	998 (14.97)	0.655 (0.029)	0.600 (0.017)
	2,3	0.027	0.470	0.409	0.029	0.025	984 (39.63)	0.628 (0.034)	0.696 (0.034)
	2,4	0.027	0.470	0.025	0.449	0.026	984 (39.23)	0.629 (0.033)	0.686 (0.042)
	2,5	0.027	0.470	0.025	0.028	0.453	984 (39.35)	0.641 (0.032)	0.686 (0.042)
	3,4	0.027	0.026	0.509	0.440	0.027	985 (38.65)	0.624 (0.034)	0.710 (0.057)
	3,5	0.027	0.026	0.509	0.031	0.444	985 (38.51)	0.634 (0.035)	0.710 (0.057)
	4,5	0.027	0.026	0.026	0.548	0.437	985 (38.10)	0.637 (0.032)	0.700 (0.072)
π_e	S0	0.027	0.027	0.026	0.033	0.037	998 (14.63)	0.797 (0.021)	0.600 (0.017)
	2,3	0.027	0.556	0.423	0.016	0.019	985 (38.56)	0.776 (0.039)	0.714 (0.034)
	2,4	0.027	0.556	0.017	0.519	0.020	984 (39.38)	0.777 (0.039)	0.698 (0.040)
	2,5	0.027	0.556	0.017	0.020	0.584	983 (39.97)	0.783 (0.033)	0.698 (0.041)
	3,4	0.027	0.027	0.611	0.481	0.020	985 (38.47)	0.776 (0.042)	0.727 (0.056)
	3,5	0.027	0.027	0.611	0.021	0.549	984 (39.17)	0.781 (0.039)	0.727 (0.056)
	4,5	0.027	0.027	0.026	0.694	0.514	984 (39.59)	0.783 (0.034)	0.710 (0.070)
POOL	S0	0.027	0.022	0.030	0.036	0.040	998 (15.38)	0.825 (0.009)	0.600 (0.017)
	2,3	0.027	0.581	0.440	0.010	0.016	978 (43.19)	0.812 (0.030)	0.722 (0.036)
	2,4	0.027	0.581	0.011	0.531	0.016	976 (43.88)	0.812 (0.030)	0.704 (0.040)
	2,5	0.027	0.581	0.011	0.017	0.609	975 (44.89)	0.815 (0.025)	0.704 (0.040)
	3,4	0.027	0.022	0.682	0.490	0.014	975 (43.90)	0.809 (0.034)	0.733 (0.057)
	3,5	0.027	0.022	0.682	0.018	0.569	974 (45.01)	0.812 (0.030)	0.733 (0.057)
	4,5	0.027	0.022	0.030	0.731	0.544	974 (44.20)	0.813 (0.028)	0.715 (0.069)

Table C.7: Constant underlying mortality scenario results, RR=0.7, assuming two effective therapeutics.

		Probability Reject in Segment					Mean (sd)	Mean (sd)	Mean (sd)
	Trt	1	2	3	4	5	N	Prop Trt	Prop Surv
	S0	0.028	0.030	0.032	0.027	0.025	997 (22.73)	0.5 (0)	0.580 (0.017)
	2,3	0.028	0.764	0.400	0.026	0.027	966 (57.19)	0.5 (0)	0.592 (0.048)
	2,4	0.028	0.764	0.027	0.284	0.027	969 (54.91)	0.5 (0)	0.632 (0.049)
P-II	2,5	0.028	0.764	0.027	0.025	0.185	971 (52.69)	0.5 (0)	0.680 (0.048)
	3,4	0.028	0.030	0.555	0.310	0.026	979 (47.71)	0.5 (0)	0.723 (0.052)
	3,5	0.028	0.030	0.555	0.026	0.198	982 (44.11)	0.5 (0)	0.768 (0.045)
	4,5	0.028	0.030	0.032	0.386	0.211	988 (37.79)	0.5 (0)	0.801 (0.047)
	S0	0.027	0.040	0.048	0.058	0.061	998 (16.20)	0.543 (0.016)	0.580 (0.017)
	2,3	0.027	0.781	0.494	0.052	0.074	963 (54.13)	0.564 (0.021)	0.598 (0.045)
	2,4	0.027	0.781	0.042	0.422	0.060	966 (52.25)	0.559 (0.019)	0.640 (0.046)
$\pi_{\text{EB}_{10}}$	2,5	0.027	0.781	0.042	0.063	0.311	968 (50.59)	0.556 (0.019)	0.686 (0.044)
	3,4	0.027	0.040	0.637	0.429	0.062	978 (44.27)	0.553 (0.018)	0.731 (0.050)
	3,5	0.027	0.040	0.637	0.052	0.322	980 (42.17)	0.550 (0.018)	0.775 (0.043)
	4,5	0.027	0.040	0.048	0.511	0.320	987 (34.64)	0.546 (0.016)	0.808 (0.046)
	S0	0.027	0.102	0.165	0.213	0.262	995 (23.33)	0.665 (0.044)	0.580 (0.017)
	2,3	0.027	0.854	0.624	0.134	0.338	951 (58.23)	0.695 (0.042)	0.617 (0.042)
	2,4	0.027	0.854	0.094	0.684	0.214	951 (59.05)	0.686 (0.043)	0.659 (0.043)
π_e	2,5	0.027	0.854	0.094	0.236	0.658	953 (59.32)	0.687 (0.041)	0.703 (0.042)
	3,4	0.027	0.102	0.768	0.610	0.209	965 (51.27)	0.673 (0.042)	0.750 (0.044)
	3,5	0.027	0.102	0.768	0.134	0.641	965 (52.54)	0.670 (0.042)	0.791 (0.037)
	4,5	0.027	0.102	0.165	0.717	0.563	975 (45.06)	0.664 (0.043)	0.822 (0.042)
	S0	0.027	0.342	0.729	0.561	0.736	926 (54.93)	0.785 (0.038)	0.576 (0.023)
	2,3	0.027	0.997	0.749	0.209	0.603	866 (46.84)	0.741 (0.031)	0.646 (0.038)
	2,4	0.027	0.997	0.171	0.928	0.365	853 (52.80)	0.722 (0.048)	0.704 (0.034)
POOL	2,5	0.027	0.997	0.171	0.486	0.922	846 (55.09)	0.737 (0.036)	0.748 (0.034)
	3,4	0.027	0.342	0.986	0.704	0.374	875 (40.03)	0.732 (0.029)	0.771 (0.032)
	3,5	0.027	0.342	0.986	0.199	0.930	854 (50.74)	0.728 (0.032)	0.808 (0.030)
	4,5	0.027	0.342	0.729	0.947	0.747	876 (47.42)	0.748 (0.035)	0.839 (0.033)

Table C.8: Varying underlying mortality scenario results, RR=0.7, assuming two effective therapeutics.

Appendix D

R Code for Chapter 4

This appendix provides the R code for the functions used for both MCMC approaches in Chapter 4 and the reversible-jump MCMC algorithm for model averaging over the two approaches. To avoid repetition, the following list provides descriptions for a majority of the arguments for the subsequent functions:

```
#grp: a vector identifying which group each data value belongs to
#hp: list of hyperparameter values (for mu, theta, tau, and p)
#hpj: list of hyperparameter values (for mu, theta, tau, and p)
#j: number of groups (the final Jth group should be UB/15.8 or whatever
   reference group is defined)
#M: number of iterations for Gibbs sampler
#model: a vector of indicators of the current model we are in for each
   level j (e.g., for CENIC-p1 it would be 0=no relationship assumed, 1=
   relationship between UB/15.8 and level j assumed)
#muj: current estimate for mu parameters for compliant group components
#nburn: number of iterations to discard from beginning of chain
#nic: nicotine level for group j (used for proposed relationship, e.g.,
   0.4, 1.3, 2.4, 5.2, 15.8)
#nj: sample size of each group in vector by group
#notmix: indicator if last group should all be assumed compliant, default
   is TRUE
```

```
#pvecj: corresponding list to y for probability of each person belonging
   to non-compliant group within group j
#seed: set seed to replicate results
#tauj: current estimate for shared estimate of precision in mixture
   distribution within each group j
#thetaj: current estimate for theta parameter for non-compliant group
   components
#yj: data in a list where each item/vector includes all observations (y_i)
    for group j (should be ordered smallest to largest within each group)
```

A special note about the "notmix" argument needs to be mentioned. Since the provided functions are meant to be generally written, you can specify notmix=TRUE if you desire to use the Gibbs sampler which does not assume a relationship to estimate the mixture components for each level j, except you are assuming that level J is entirely compliant (not a mixture). For the RJMCMC this is set to equal TRUE in order to be match the restriction for the Gibbs sampler of the relationship where the control has a fixed compliance status (e.g., all compliant).

D.1 Functions to implement Gibbs samplers without model averaging

D.1.1 Gibbs sampler for approach not assuming a relationship

```
gibbs_normalmixture_all <- function(yj,nj,j,pvecj,muj,thetaj,tauj,hpj,
   notmix=T){
###This function provides the code for updating the Gibbs sampler for a
   mixture of two normal distributions where all components are updated
   across j-levels
```

```
zj <- lapply(1:j, function(x) c(0,rbinom(n=(nj[x]-2), size=1, pvecj
   [[x]][2:(nj[x]-1)]),1) ) #generate new latent indicators from
   Bernoulli with smallest observation fixed as compliant and
   largest fixed as non-compliant
if(notmix==T){ zj[[j]] <- rep(0, length(zj[[j]])) } #if highest
   group/level is only compliant, convert all indicators to 0

ncj <- sapply(1:j, function(x) sum(zj[[x]]==0) )
nnj <- sapply(1:j, function(x) sum(zj[[x]]==1) )

ycj <- lapply(1:j, function(x) yj[[x]][which(zj[[x]]==0)] )
ynj <- lapply(1:j, function(x) yj[[x]][which(zj[[x]]==1)] )

###calculate component means and variances for each group
#compliant components
ycm <- sapply(1:j, function(x) mean(ycj[[x]]) )
ycv <- sapply(1:j, function(x) if( length(ycj[[x]])==1 ){ 0 }else{
   var(ycj[[x]]) }) #place variance of 0 if only one observation to
    avoid errors
#non-compliant components
ynm <- sapply(1:j, function(x) if( length(ynj[[x]])==0 ){0}else{
   mean(ynj[[x]]) })
ynv <- sapply(1:j, function(x) if( length(ynj[[x]]) == 0 | length(
   ynj[[x]]) == 1 ){ 0 }else{ var(ynj[[x]]) }) #set variance to 0
   if only one observation to avoid errors later on

# sampling p: probability of membership in group 2 (non-compliers)
pj <- rbeta(j, nnj + hpj$p.alpha, ncj + hpj$p.beta)

# sampling tau
shape.val <- (nj/2) + hpj$tau.a
```

```
        rate.val <- 0.5 * ((ncj-1)*ycv + (nnj-1)*ynv + ncj*(ycm-muj)^2 +
            nnj*(ynm-(muj+thetaj))^2 + 2*hpj$tau.b )
        tauj <- rgamma(j,shape=shape.val,rate=rate.val)
        sigmaj <- 1/sqrt(tauj)

        # sampling mu
        m.mean <- -( tauj * (nnj*(thetaj - ynm) - ncj*ycm) ) / ( nj*tauj +
            hpj$mu.prec )
        m.prec <- ( nj*tauj + hpj$mu.prec )
        muj <- rnorm(j, mean=m.mean, sd=sqrt(1/m.prec))

        # sampling theta
        t.mean <- -( nnj*tauj*(muj - ynm) ) / (nnj*tauj + hpj$theta.prec)
        t.prec <- (nnj*tauj + hpj$theta.prec)
        thetaj <- abs( rnorm(j, mean=t.mean, sd=sqrt(1/t.prec)))
        if(notmix==T){ thetaj[j] <- 0 } #replace largest group theta with 0
             if notmix is TRUE

        ret <- list(mu_comp=muj, mu_nonc=muj+thetaj, mu=muj, theta=thetaj,
            tau=tauj, sigma=sigmaj, z=zj, p=pj)

}
```

D.1.2 Gibbs sampler for approach assuming a relationship

```
gibbs_normalmixture_meanrel <- function(yj,nj,j,pvecj,muj,thetaj,tauj,hpj,
    nic){
###This function provides the code for updating the Gibbs sampler for a
    mixture of two normal distributions where the compliant group means are
     in relationship with the UB/15.8 group

        dj <- log(nic/nic[length(nic)]) #proposed relationship for CENIC-p1
```

```
zj <- lapply(1:j, function(x) c(0,rbinom(n=(nj[x]-2), size=1, pvecj
    [[x]][2:(nj[x]-1)]),1) ) #generate new latent indicators from
    Bernoulli with smallest observation fixed as compliant and
    largest fixed as non-compliant
zj[[j]] <- rep(0, length(zj[[j]])) #highest group/level is only
    compliant, convert all indicators to 0

ncj <- sapply(1:j, function(x) sum(zj[[x]]==0) )
nnj <- sapply(1:j, function(x) sum(zj[[x]]==1) )

ycj <- lapply(1:j, function(x) yj[[x]][which(zj[[x]]==0)] )
ynj <- lapply(1:j, function(x) yj[[x]][which(zj[[x]]==1)] )

###calculate component means and variances for each group
#compliant components
ycm <- sapply(1:j, function(x) mean(ycj[[x]]) )
ycv <- sapply(1:j, function(x) if( length(ycj[[x]])==1 ){ 0 }else{
    var(ycj[[x]]) }) #place variance of 0 if only one observation to
     avoid errors
#non-compliant components
ynm <- sapply(1:j, function(x) if( length(ynj[[x]])==0 ){0}else{
    mean(ynj[[x]]) })
ynv <- sapply(1:j, function(x) if( length(ynj[[x]]) == 0 | length(
    ynj[[x]]) == 1 ){ 0 }else{ var(ynj[[x]]) }) #set variance to 0
    if only one observation to avoid errors later on

# sampling p: probability of membership in group 2 (non-compliers)
pj <- rbeta(j, nnj + hpj$p.alpha, ncj + hpj$p.beta)

# sampling tau
```

```
        shape.val <- (nj/2) + hpj$tau.a
        rate.val <- 0.5 * ((ncj-1)*ycv + (nnj-1)*ynv + ncj*(ycm - (dj + muj
            [j]))^2 + nnj*(ynm-(dj+muj[j]+thetaj))^2 + 2*hpj$tau.b )
        tauj <- rgamma(j,shape=shape.val,rate=rate.val)
        sigmaj <- 1/sqrt(tauj)

        # sampling mu_J
        g <- tauj*(nj*dj + nnj*thetaj - ncj*ycm - nnj*ynm)
        H1 <- sum(tauj*nj) + hpj$mu.prec[j]
        H2 <- sum(g[1:(j-1)]) - tauj[j]*nj[j]*ycm[j]
        m.mean <- -H2/H1
        m.prec <- H1
        muJ <- rnorm(1, mean=m.mean, sd=sqrt(1/m.prec))
        muj <- muJ+dj #muj determined by proposed relationship

        # sampling theta
        t.mean <- -( nnj*tauj*(dj + muj[j] - ynm) ) / (nnj*tauj + hpj$theta
            .prec)
        t.prec <- (nnj*tauj + hpj$theta.prec)
        thetaj <- abs( rnorm(j, mean=t.mean, sd=sqrt(1/t.prec)))
        thetaj[j] <- 0 #replace largest group theta with 0 since everyone
            is compliant

        ret <- list(mu_comp=muj, mu_nonc=muj+thetaj, mu=muj, theta=thetaj,
            tau=tauj, sigma=sigmaj, z=zj, p=pj)

}
```

D.1.3 Implement Gibbs samplers for posterior estimation

```
mix2normals = function(type,y,grp,nic=NULL,inits,hp,M,nburn,notmix=T,seed)
    {
```

```
###Gibbs samplers for mixture of 2 normal distributions assuming equal
   variance
#type: Gibbs sampler to use:
##'all' estimates all parameters
##'rel' assumes relationship between UB/15.8 group and compliant group
   means for other nicotine levels
#inits: list of initial values to use for mu, theta, tau, p [note: tau is
   precision]
##if one value given, it is assumed to be the same across all groups;
   otherwise give vector with value for each group!!!

set.seed(seed)

       yj <- lapply( split(y, grp), sort) #split data and sort in one step
       j <- length(yj) #number of groups
       nj <- sapply( yj, length) #calculate number of observations within
           each group

       draws=matrix(NA,M,7*j)
       colnames(draws) <- c(paste0('p',1:j),paste0('mu_comp',1:j),paste0('
           mu_nonc',1:j),paste0('sigma',1:j),paste0('c80_',1:j),paste0('c
           90_',1:j),paste0('c95_',1:j))

       #initial values (if only one given, assume it is initial value for
           each level/group)
       if( length(inits$mu_comp)==1 ){ mu_compj <- rep(inits$mu_comp, j) }
           else{ mu_compj <- inits$mu_comp }
       if( length(inits$mu_nonc)==1 ){ mu_noncj <- rep(inits$mu_nonc, j) }
           else{ mu_noncj <- inits$mu_nonc }
       if( length(inits$tau)==1 ){ tauj <- rep(inits$tau, j) }else{ tauj
           <- inits$tau }
```

```
if( length(inits$p)==1 ){ pj <- rep(inits$p, j) }else{ pj <- inits$
   p }
if( (length(mu_compj)!=j | length(mu_noncj)!=j | length(tauj)!=j |
   length(pj)!=j) == TRUE ){ stop('Please either give inits of
   length 1 or length j. At least one is not equal to 1 or j!') } #
   break from function if inits are not of correct length

#Convert hyperparameter values to be for each group if separate
   values not provided for each group (if only one given, assume it
    is used for each level/group)
if( length(hp$tau.a)==1 ){ hp$tau.a <- rep(hp$tau.a,j) }
if( length(hp$tau.b)==1 ){ hp$tau.b <- rep(hp$tau.b,j) }
if( length(hp$p.alpha)==1 ){ hp$p.alpha <- rep(hp$p.alpha,j) }
if( length(hp$p.beta)==1 ){ hp$p.beta <- rep(hp$p.beta,j) }
if( length(hp$mu.prec)==1 ){ hp$mu.prec <- rep(hp$mu.prec,j) }
if( length(hp$theta.prec)==1 ){ hp$theta.prec <- rep(hp$theta.prec,
   j) }
if( (length(hp$tau.a)!=j | length(hp$tau.b)!=j | length(hp$p.alpha)
   !=j | length(hp$p.beta)!=j | length(hp$mu.prec)!=j | length(hp$
   theta.prec)!=j) == TRUE ){ stop('Please give hp of length 1 or
   length j. At least one is not equal to 1 or j!')}

#intialize latent indicator variable
zj <- lapply(1:j, function(x) c(0, rep(NA, nj[x]-2), 1) ) #
   initialize latent indicator for each group
zj.sum <- lapply(1:j, function(x) rep(0, nj[x]) ) #list keeping
   track of number of iterations after burn-in where y_ij is
   classified as k=1

for (iter in 1:M){
```

```
        pvecj <- lapply(1:length(yj), function(x) pj[x]*dnorm(yj[[x
            ]], mean=mu_noncj[x], sd=sqrt(1/tauj[x])) / ( (1-pj[x])*
            dnorm(yj[[x]], mean=mu_compj[x], sd=sqrt(1/tauj[x])) +
            pj[x]*dnorm(yj[[x]], mean=mu_noncj[x], sd=sqrt(1/tauj[x
            ])) ) )
        if(notmix==T){ pvecj[[j]] <- rep(0, nj[j] ) } #if notmix=
            TRUE, set all values as 0 to indicate membership in
            compliant group

        if(type=='all'){gibbs.step <- gibbs_normalmixture_all(yj=yj,
            nj=nj,pvecj=pvecj,muj=mu_compj,thetaj=mu_noncj-mu_compj,
            tauj=tauj,hpj=hp,notmix=notmix,j=j)}
        if(type=='rel'){gibbs.step <- gibbs_normalmixture_meanrel(yj
            =yj,nj=nj,pvecj=pvecj,muj=mu_compj,thetaj=mu_noncj-mu_
            compj,tauj=tauj,hpj=hp,j=j,nic=nic)}

        mu_compj <- gibbs.step$mu_comp
        mu_noncj <- gibbs.step$mu_nonc
        tauj <- gibbs.step$tau
        sigmaj <- gibbs.step$sigma
        zj <- gibbs.step$z
        pj <- gibbs.step$p

        if(iter > nburn){
               zj.sum <- lapply(1:j, function(x) zj[[x]] + zj.sum[[x
                   ]]) #add this iterations zj values to sum
        }

        draws[iter,] <- c(pj,mu_compj,mu_noncj,sigmaj)
}
```

```
        ret <- list( samp=draws[-(1:nburn),], zj.sum=zj.sum )

        return(ret)

}
```

D.2 Reversible-jump MCMC algorithm functions

D.2.1 Gibbs sampler to flexibly estimate with either approach for RJMCMC

```
gibbs_normalmixture_rjmcmc <- function(yj,nj,j,pvecj,muj,thetaj,tauj,hpj,
    model,nic){
###This function provides the code for updating the Gibbs sampler for a
    mixture of two normal distributions where all components are updated
    across j-levels, where each level may be estimated using either Gibbs
    sampler approach

        j.all <- which( model==0 ) #identify levels estimated using model
            not assuming the relationship
        j.rel <- which( model==1 ) #identify levels estimated using model
            assuming the relationship
        dj <- log(nic/nic[length(nic)])

        zj <- lapply(1:j, function(x) c(0,rbinom(n=(nj[x]-2), size=1, pvecj
            [[x]][2:(nj[x]-1)]),1) ) #generate new latent indicators from
            Bernoulli with smallest observation fixed as compliant and
            largest fixed as non-compliant
        zj[[j]] <- rep(0, length(zj[[j]])) #if highest group/level is only
            compliant, convert all indicators to 0

        ncj <- sapply(1:j, function(x) sum(zj[[x]]==0) )
```

```
nnj <- sapply(1:j, function(x) sum(zj[[x]]==1) )

ycj <- lapply(1:j, function(x) yj[[x]][which(zj[[x]]==0)] )
ynj <- lapply(1:j, function(x) yj[[x]][which(zj[[x]]==1)] )

###calculate component means and variances for each group
#compliant components
ycm <- sapply(1:j, function(x) mean(ycj[[x]]) )
ycv <- sapply(1:j, function(x) if( length(ycj[[x]])==1 ){ 0 }else{
    var(ycj[[x]]) }) #place variance of 0 if only one observation to
     avoid errors
#non-compliant components
ynm <- sapply(1:j, function(x) if( length(ynj[[x]])==0 ){0}else{
    mean(ynj[[x]]) })
ynv <- sapply(1:j, function(x) if( length(ynj[[x]]) == 0 | length(
    ynj[[x]]) == 1 ){ 0 }else{ var(ynj[[x]]) }) #place variance of 0
     if only one observation to avoid errors

# sampling p: probability of membership in group 2 (non-compliers)
pj <- rbeta(j, nnj + hpj$p.alpha, ncj + hpj$p.beta)

# sampling tau
shape.val <- (nj/2) + hpj$tau.a
rate.val <- 0.5 * ((ncj-1)*ycv + (nnj-1)*ynv + ncj*(ycm-muj)^2 +
    nnj*(ynm-(muj+thetaj))^2 + 2*hpj$tau.b )
tauj <- rgamma(j,shape=shape.val,rate=rate.val)
sigmaj <- 1/sqrt(tauj)

### sampling mu
#levels estimating compliant mean with own information only
```

```
        m.mean_all <- -( tauj * (nnj*(thetaj - ynm) - ncj*ycm) ) / ( nj*
            tauj + hpj$mu.prec )
        m.prec_all <- ( nj*tauj + hpj$mu.prec )
        muj_all <- rnorm( length(j.all) , mean=m.mean_all[j.all], sd=sqrt
            (1/m.prec_all[j.all]))

        #groups calculating compliant mean based on relationship to mu_J
        g <- tauj[j.rel]*(nj[j.rel]*dj[j.rel] + nnj[j.rel]*thetaj[j.rel] -
            ncj[j.rel]*ycm[j.rel] - nnj[j.rel]*ynm[j.rel])
        H1 <- sum(tauj[j.rel]*nj[j.rel]) + hpj$mu.prec[j]
        H2 <- sum(g[0:(length(j.rel)-1)]) - tauj[j]*nj[j]*ycm[j]
        m.mean_rel <- -H2/H1
        m.prec_rel <- H1
        muJ <- rnorm(1, mean=m.mean_rel, sd=sqrt(1/m.prec_rel))
        muj_rel <- muJ+dj[j.rel]

        #combine muj estimates from both approaches
        muj[j.all] <- muj_all
        muj[j.rel] <- muj_rel

        # sampling theta
        t.mean <- -( nnj*tauj*(muj - ynm) ) / (nnj*tauj + hpj$theta.prec)
        t.prec <- (nnj*tauj + hpj$theta.prec)
        thetaj <- abs( rnorm(j, mean=t.mean, sd=sqrt(1/t.prec)))
        thetaj[j] <- 0 #replace largest group theta with 0 if notmix is
            TRUE

        ret <- list(mu_comp=muj, mu_nonc=muj+thetaj, mu=muj, theta=thetaj,
            tau=tauj, sigma=sigmaj, z=zj, p=pj)

}
```

D.2.2 RJMCMC update step

```
updatemodel <- function(yj,nj,j,pj,tauj,zj,muj,thetaj,model,hpj,prob,nic){
###Function to complete the "reversible jump" portion of the RJMCMC to
   move between our two proposed models
#pj: current estimate for proportion in non-compliant group for each level
    j
#zj: latent variable classifying y_i subject in compliant or non-compliant
    group
#prob: prior probability for level j using all components model (model 0)

       ncj <- sapply(1:j, function(x) sum(zj[[x]]==0) )
       nnj <- sapply(1:j, function(x) sum(zj[[x]]==1) )

       ycj <- sapply(1:j, function(x) yj[[x]][which(zj[[x]]==0)] )
       ynj <- sapply(1:j, function(x) yj[[x]][which(zj[[x]]==1)] )

       ###calculate component means and variances for each group
       #compliant components
       ycm <- sapply(1:j, function(x) mean(ycj[[x]]) )
       ycv <- sapply(1:j, function(x) if( length(ycj[[x]])==1 ){ 0 }else{
           var(ycj[[x]]) }) #place variance of 0 if only one observation to
            avoid errors
       #non-compliant components
       ynm <- sapply(1:j, function(x) if( length(ynj[[x]])==0 ){0}else{
           mean(ynj[[x]]) })
       ynv <- sapply(1:j, function(x) if( length(ynj[[x]])==0 | length(ynj
           [[x]])==1 ){ 0 }else{ var(ynj[[x]]) }) #place variance of 0 if
           only one observation to avoid errors

       ###Identify current model state and propose new variables for other
            model
```

```
#Determine which level to propose different model for
r <- sample(1:(j-1),1)
r.state <- model[r]

muj_prop <- muj #create proposal muj to update with value below
   depending on current r.state
model_prop <- model; if(r.state==0){model_prop[r] <- 1}else{model_
   prop[r] <- 0}
m.mean_r <- (-( tauj * (nnj*(thetaj - ynm) - ncj*ycm) ) / ( nj*tauj
    + hpj$mu.prec ))[r]
m.prec_r <- ( nj*tauj + hpj$mu.prec )[r]

if(r.state==0){ #model 0 is all components, so propose values under
    relationship assumption
      muj_prop[r] <- muj[j] + log(nic[r]/nic[length(nic)])
}else{ #model 1 is relationship, so propose values under all
   components model
      muj_prop[r] <- rnorm(1, mean=m.mean_r, sd=sqrt(1/m.prec_r))
}

modprob.num <- prod( ( prob^abs(model_prop-1) * (1-prob)^model_prop
   )[1:(j-1)] )
modprob.den <- prod( ( prob^abs(model-1) * (1-prob)^model)[1:(j-1)]
    )

###Calculate acceptance probability
#Calculate log-likelihoods
loglikj <- sapply(1:j, function(x) sum( log( (1-pj[x])*dnorm(yj[[x
   ]], mean=muj[x], sd=sqrt(1/tauj[x])) + pj[x]*dnorm(yj[[x]], mean
   =muj[x]+thetaj[x], sd=sqrt(1/tauj[x])) ) ) )
loglik <- sum(loglikj)
```

```
newloglikj <- sapply(1:j, function(x) sum( log( (1-pj[x])*dnorm(yj
    [[x]], mean=muj_prop[x], sd=sqrt(1/tauj[x])) + pj[x]*dnorm(yj[[x
    ]], mean=muj_prop[x]+thetaj[x], sd=sqrt(1/tauj[x])) ) ) )
newloglik <- sum(newloglikj)

#if currently in model 0 (est all), look at num being proposed move
     for level r to model 1 (relationship assumed)
if(r.state == 0){
       num <- newloglik + log(dnorm(muj_prop[r], mean=m.mean_r, sd=
           sqrt(1/m.prec_r))) + log(modprob.num)
       den <- loglik + log(dnorm(muj[r], mean=0, sd=sqrt(1/hpj$mu.
           prec[r]))) + log(modprob.den)
}else{ #if currently in model 1 (relationship), look at num being
    proposed move for level r to model 0 (estimate level with level'
    s data)
       num <- newloglik + log(dnorm(muj_prop[r], mean=0, sd=sqrt(1/
           hpj$mu.prec[r]))) + log(modprob.num)
       den <- loglik + log(dnorm(muj[r], mean=m.mean_r, sd=sqrt(1/m
           .prec_r))) + log(modprob.den)
}

###Acceptance probability
A <- min(1, exp(num-den))
u <- runif(1)
if(u <= A){
       if(r.state == 0){model[r] <- 1}else{model[r] <- 0}
       muj <- muj_prop
}

ret <- list(model=model, muj=muj)
```

```
}
```

D.2.3 Function keep track of RJMCMC model state

```
baseKto10 <- function(x,K=2){
###Function to convert base K vector of values into base 10 number
#x: vector with values in base K (e.g., base 2 has 0/1)
#K: base to convert from, default is base 2
       sum(x * K^((length(x)-1):0))
}
```

D.2.4 Implement entire RJMCMC algorithm

```
rjmcmc_2normals = function(y,inits,hp,prob,M,seed,nburn,grp,nic){
###Gibbs samplers for mixture of 2 normal distributions assuming equal
   variance
#inits: list of initial values to use for lambda, theta, tau, p [note: tau
    is precision]
 since proposal distribution is conditional posterior of mu_j)
#prob: prior probability for level j using all components model (model 0)

       set.seed(seed)

       yj <- lapply( split(y, grp), sort) #split data and sort in one step
       j <- length(yj) #number of groups
       nj <- sapply( yj, length) #calculate number of observations within
           each group

       draws=matrix(NA,M,(7*j+1))
       colnames(draws) <- c('model',paste0('p',1:j),paste0('mu_comp',1:j),
           paste0('mu_nonc',1:j),paste0('sigma',1:j),paste0('c80_',1:j),
           paste0('c90_',1:j),paste0('c95_',1:j))
```

```
#initial values (if only one given, assume it is initial value for
    each level/group)
if( length(inits$mu_comp)==1 ){ mu_compj <- rep(inits$mu_comp, j) }
    else{ mu_compj <- inits$mu_comp }
if( length(inits$mu_nonc)==1 ){ mu_noncj <- rep(inits$mu_nonc, j) }
    else{ mu_noncj <- inits$mu_nonc }
if( length(inits$tau)==1 ){ tauj <- rep(inits$tau, j) }else{ tauj
    <- inits$tau }
if( length(inits$p)==1 ){ pj <- rep(inits$p, j) }else{ pj <- inits$
    p }
if( length(inits$model)==1 ){model <- c(rep(inits$model, (j-1)),1)}
    else{ model <- c(inits$model[1:(j-1)],1) } #always assume Jth
    model is 1 for relationship
if( (length(mu_compj)!=j | length(mu_noncj)!=j | length(tauj)!=j |
    length(pj)!=j | length(model)!=j ) == TRUE ){ stop('Please
    either give inits of length 1 or length j. At least one is not
    equal to 1 or j!') } #break from function if inits are not of
    correct length

#Convert hyperparameter values to be for each group if separate
    values not provided for each group (if only one given, assume it
     is used for each level/group)
if( length(hp$tau.a)==1 ){ hp$tau.a <- rep(hp$tau.a,j) }
if( length(hp$tau.b)==1 ){ hp$tau.b <- rep(hp$tau.b,j) }
if( length(hp$p.alpha)==1 ){ hp$p.alpha <- rep(hp$p.alpha,j) }
if( length(hp$p.beta)==1 ){ hp$p.beta <- rep(hp$p.beta,j) }
if( length(hp$mu.prec)==1 ){ hp$mu.prec <- rep(hp$mu.prec,j) }
if( length(hp$theta.prec)==1 ){ hp$theta.prec <- rep(hp$theta.prec,
    j) }
```

```
if( (length(hp$tau.a)!=j | length(hp$tau.b)!=j | length(hp$p.alpha)
    !=j | length(hp$p.beta)!=j | length(hp$mu.prec)!=j | length(hp$
    theta.prec)!=j) == TRUE ){ stop('Please give hp of length 1 or
    length j. At least one is not equal to 1 or j!')}

#intialize latent indicator variable
zj <- lapply(1:length(yj), function(x) c(0, rep(NA, nj[x]-2), 1) )
    #initialize latent indicator for each group

zj.sum <- lapply(1:j, function(x) rep(0, nj[x]) ) #list keeping
    track of number of iterations after burn-in where y_ij is
    classified as k=1

for (iter in 1:M){
        #Update MCMC according to model state currently in
        pvecj <- lapply(1:length(yj), function(x) pj[x]*dnorm(yj[[x
            ]], mean=mu_noncj[x], sd=sqrt(1/tauj[x])) / ( (1-pj[x])*
            dnorm(yj[[x]], mean=mu_compj[x], sd=sqrt(1/tauj[x])) +
            pj[x]*dnorm(yj[[x]], mean=mu_noncj[x], sd=sqrt(1/tauj[x
            ])) ) )
        pvecj[[j]] <- rep(0, nj[j] ) #replace group J with all
            compliant indicators

        #update step with Gibbs sampler for given model assumed for
            each level
        gibbs.step <- gibbs_normalmixture_rjmcmc(yj=yj,nj=nj,pvecj=
            pvecj,muj=mu_compj,thetaj=mu_noncj-mu_compj,tauj=tauj,
            hpj=hp,j=j,model=model,nic=nic)

        mu_compj <- gibbs.step$mu_comp
        mu_noncj <- gibbs.step$mu_nonc
```

```
            tauj <- gibbs.step$tau
            sigmaj <- gibbs.step$sigma
            zj <- gibbs.step$z
            pj <- gibbs.step$p

            if(iter > nburn){
                   zj.sum <- lapply(1:j, function(x) zj[[x]] + zj.sum[[x
                      ]]) #add this iterations zj values to sum
            }

            #Complete RJ step to determine if we switch models and
               update
            rjstep <- updatemodel(yj=yj,pj=pj,j=j,nj=nj,tauj=tauj,zj=zj,
               muj=mu_compj,thetaj=mu_noncj-mu_compj,model=model,hpj=hp
               ,prob=prob,nic=nic)
            model <- rjstep$model
            mu_comp <- rjstep$muj

            mod10 <- baseKto10(x=model[1:(j-1)],K=2)

            draws[iter,] <- c(mod10,pj,mu_compj,mu_noncj,sigmaj)
        }

        ret <- list( samp=draws[-(1:nburn),], zj.sum=zj.sum )

        return(ret)

}
```

Appendix E

Acronyms

Care has been taken in this thesis to minimize the use of acronyms, but this cannot always be achieved. This appendix contains a table of some of the more frequently occurring acronyms and their meaning.

Table E.1: Acronyms

Acronym	Meaning
AR	Adaptive randomization
BMA	Bayesian model averaging
CENIC-p1	Center for the Evaluation of Nicotine in Cigarettes, project 1
CP	Commensurate prior
CPD	Cigarettes (smoked) per day
EB	Empirical Bayes(ian)
ESSS	Effective supplemental sample size
EVD	Ebola virus disease
MEM	Multi-source exchangeability model(s)
PREVAIL II	Partnership for Research on Ebola Vaccines in Liberia
SHM	Standard hierarchical model
SOC/oSOC	Standard of care/optimal standard of care
VLNC	Very low nicotine content