Bayesian Methods and Computation for Large Observational Datasets

Bayesian Methods and Computation for Large Observational Datasets

A thesis presented

by

Krista Leigh Watts

to

The Department of Biostatistics

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Biostatistics

Harvard University
Cambridge, Massachusetts

May 2013

Thesis Advisor: Professor Francesca Dominici Krista Leigh Watts

Bayesian Methods and Computation for Large Observational Datasets

Abstract

Much health related research depends heavily on the analysis of a rapidly expanding universe of observational data. A challenge in analysis of such data is the lack of sound statistical methods and tools that can address multiple facets of estimating treatment or exposure effects in observational studies with a large number of covariates. We sought to advance methods to improve analysis of large observational datasets with an end goal of understanding the effect of treatments or exposures on health. First we compared existing methods for propensity score (PS) adjustment, specifically Bayesian propensity scores. This concept had previously been introduced (*McCandless et al.*, 2009) but no rigorous evaluation had been done to evaluate the impact of feedback when fitting the joint likelihood for both the PS and outcome models. We determined that unless specific steps were taken to mitigate the impact of feedback, it has the potential to distort estimates of the treatment effect. Next, we developed a method for accounting for uncertainty in confounding adjustment in the context of multiple exposures. Our method allows us to select confounders based on their association with the joint exposure and the outcome while also accounting for the uncertainty in the confounding adjustment. Finally, we developed two methods to combine heterogenous sources of data for effect estimation, specifically information coming from a primary data source that provides information for treatments, outcomes, and a limited set of measured confounders on a large number of people and smaller supplementary data sources containing a much richer set of covariates. Our methods avoid the need to specify the full joint distribution of all covariates.

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I would like to thank my wonderful husband, Clint. Without you, I would never have even been able to attempt this endeavor. Thank you for providing me so much support over the years.

I would like to thank Francesca Dominici for being the best advisor any student could ever ask for. Your selflessness in giving to your students continues to amaze me. Aside from all of your invaluable advice and guidance, your can-do attitude was instrumental in accomplishing this. In addition, I would like to thank Brent Coull, a committee member and unofficial co-advisor; you always went above and beyond in giving me assistance and advice. I never failed to leave your office with answers - often to questions I didn't even realize I had. Thank you to Sharon-Lise Normand for agreeing to serve on my committee and always having such valuable questions and insight into my research.

I would like to thank Cory Zigler for both your official collaboration on all three of my papers and also for being available any time I wanted to stop by and bounce something off of you. To Yun Wang, this research would not have been possible without the data you provided and the questions you so tirelessly answered.

Finally, I would like to thank my classmates and unofficial collaborators. Matt Cefalu, Beth Smoot, Denis Agniel, Joey Antonelli, Tristan Hayek and Caleb Miles, your insight and assistance was invaluable. And to all of my classmates, I would never have made it to this point without your support and teamwork. Your presence and attitude has made this an unforgettable experience.

Preface

Increasingly, health effects research relies on large, observational datasets. Compared to clinical trials, the analyses of these databases allow us to study a much larger population and to investigate additional questions of interest. However, analysis of these large and complex administrative databases raises several methodological challenges and requires the development of new statistical methods. Comparing the effectiveness of treatment strategies or the health effect of exposure to harmful agents in observational data is challenging in part because people are not randomly assigned to treatment strategies or exposures, which introduces the likely possibility that outcome comparisons are confounded by factors that simultaneously relate to exposure status, treatment choices, and health outcomes.

Here, we attempt to advance existing research by evaluating existing methods and developing new methods for confounding adjustment in large observational datasets. First, in Chapter 1, we look at Bayesian propensity score methods, formally introduced by *McCandless et al.* (2009). Bayesian methods have natural appeal because separate likelihoods for what is normally a two stage procedure can be combined into a single joint likelihood, with estimation of the two stages carried out simultaneously. In theory, this makes more complete use of the data than traditional frequentist propensity score methods. One key feature of joint estimation in this context is "feedback" between the outcome stage and the propensity score stage, meaning that health outcome data contributes to the estimation of the posterior distributions of the propensity score. This has been criticized as violating the principles of objective experimental design (*Rubin*, 2007, 2008). If propensity scores are meant to approximate the design stage of a randomized study, any access to outcome when calculating the propensity score fails to ensure that objective design decisions are completely separate from analysis decisions. However, methods that incorporate outcome information have also been advocated (*Schneeweiss et al.*, 2009, *McCandless et al.*, 2009) We found that a rigorous investigation of exactly how feedback can impact estimation of causal effects was lacking. We provide this rigorous assessment of Bayesian propensity score estimation and demonstrate that model feedback can bias estimates of the causal effect absent strategies to ensure that the propensity score maintains its properties as a balancing score. Much of this was joint work with Corwin Zigler and large portions of the first chapter have been published in the paper titled

Model Feedback in Bayesian Propensity Score Estimation (*Zigler et al.*, 2013). In Chapter 1, we also add the following contributions:

- An analysis of an approximately Bayesian method that "cuts the feedback" from the outcome model to the propensity score model. *Zigler et al.* (2013) compare the joint Bayesian method to an traditional sequential approach; here we add in a comparison to a sequential Bayesian approach as well. This method is originally described by *McCandless et al.* (2010). Here we redefine the method, evaluate it in our simulation study and apply it to a comparative effectiveness analysis of carotid artery stenting versus the more traditional carotid endarterectomy.

- Comparison of the methods' performances in the situation where either the propensity score or outcome model is misspecified. *Zigler et al.* (2013) note that augmenting the propensity score adjustment in the outcome model with adjustment for every covariate that appears in the propensity score model is "akin to those previously developed to yield "doubly robust" estimators" (*Bang and Robins*, 2005, *Little*, 2011) but do not explore whether this model shares the desirable features of a doubly robust estimator. Here we conduct a simulation study to evaluate the performance of the joint Bayesian, sequential Bayesian and traditional sequential approaches in these settings.

In Chapter 2 we develop a method for confounding adjustment in the setting of multiple exposures or treatments. This method is developed in the context of air pollution epidemiology. Currently, most epidemiological studies examine health effects associated with exposure to a single environmental contaminant at a time. However, humans are exposed to many environmental agents at once and therefore epidemiological studies need to change focus to this more realistic setting. One challenge with the transition from a single exposure to multiple exposures is the lack of a formal approach to select which measured confounders should be included in the outcome model. Standard approaches for selecting confounders in the context of a single exposure are not adequate in the context of multiple exposures; the set of confounders of an outcome associated with simultaneous exposure to more than one exposure or treatment cannot be fully characterized

by the confounders of the effect of each individual exposure separately. The key task is to identify confounders that are jointly associated with multiple exposures and the outcome. In this chapter, we will make two contributions. First, we will clarify the difference as to what constitutes a true confounder in multiple exposure settings versus single exposure settings. A true confounder in the multiple exposure setting is any covariate that confounds the relationship between simultaneous exposure to multiple pollutants and the outcome of interest. This could be a covariate that is marginally associated with one or more exposures - and, hence, would also be a confounder in the single exposure setting - or one that is jointly associated with multiple exposures (and might not be a confounder in the single exposure setting). Second, we will develop a statistical framework to adjust for confounding in the presence of multiple exposures while accounting for uncertainty in the confounding adjustment. Recently (*Wang et al.*, 2012) introduced Bayesian Adjustment for Confounding (BAC) as a method to select confounders in the single exposure setting. BAC uses a Bayesian approach to model averaging to estimate the health effect associated with exposure to a single pollutant while acknowledging the uncertainty in the confounder selection. We introduce BAC for multiple exposures (BAC-ME) to extend this framework where selection of confounders is based on simultaneous exposure to multiple pollutants. Our method allows us to select a subset of covariates to include to control for confounding in a linear regression model while protecting against the possibility of eliminating a true confounder. This also helps identify true confounders for future research efforts. We show through simulation studies that it is of paramount importance to include all confounders in the outcome model and that excluding only one true confounder could lead to substantial bias in estimation of the multi pollutant adverse health effect. We also apply our method to a retrospective epidemiological study aimed at estimating the multi pollutant adverse effect on cardiovascular hospitalization associated with a simultaneous change in ozone and $PM_{2.5}$, controlling for weather data and population level characteristics. This work has been submitted for publication (*Bayesian Adjustment for Confounding in the Presence of Multiple Exposures*, Krista Watts, Corwin M. Zigler and Francesca Dominici)

In Chapter 3 we develop two methods to combine data from heterogeneous data sources when the goal is to compare the effect of two treatments or exposures. We look specifically at the setting where we have information coming from a primary data source that provides information for treatments, outcomes, and a limited set of measured confounders on a large number of people and

smaller supplementary data sources containing a much richer set of covariates. Often, important confounders are not measured in the primary data. However, the supplemental data source may contain information on important confounders in a subset of the population. Current methods for combining such data sources for analysis require specifying the joint distribution of all data (*Little and Rubin*, 2002). When the missing covariates are high dimensional, correlated, or contain both continuous and dichotomous or categorical variables, correctly specifying this distribution is nearly impossible. Recently, *McCandless et al.* (2012) suggest a method to use 'conditional propensity scores' to adjust for confounders available only in a supplementary dataset. We propose two methods that build on their work. We conduct a simulation study to show settings when our methods can substantially reduce bias over complete case analysis or 'naive' analysis that adjusts for only the fully measured covariates. We expect to submit this work for publication in the next few weeks (*Propensity Score Methods for Combining Data Sources*, Krista Watts, Corwin M. Zigler, Yun Wang and Francesca Dominici)

Model Feedback in Bayesian Propensity Score Estimation

1.1 Abstract

Methods based on the propensity score comprise one set of valuable tools for comparative effectiveness research and for estimating causal effects more generally. These methods typically consist of two distinct stages: 1) a propensity score stage where a model is fit to predict the propensity to receive treatment (the propensity score), and 2) an outcome stage where responses are compared in treated and untreated units having similar values of the estimated propensity score. Traditional techniques conduct estimation in these two stages separately; estimates from the first stage are treated as fixed and known for use in the second stage. Bayesian methods have natural appeal in these settings because separate likelihoods for the two stages can be combined into a single joint likelihood, with estimation of the two stages carried out simultaneously. One key feature of joint estimation in this context is 'feedback' between the outcome stage and the propensity score stage, meaning that quantities in a model for the outcome contribute information to posterior distributions of quantities in the model for the propensity score. We provide a rigorous assessment of joint Bayesian propensity score estimation to show that model feedback can produce poor estimates of causal effects absent strategies that augment propensity score adjustment with adjustment for individual covariates. We also explore an approximately Bayesian sequential method and show that adjustment for individual covariates is not required to obtain an unbiased estimate of the causal effect. We illustrate this phenomenon with a simulation study and with a comparative effectiveness investigation of carotid artery stenting vs. carotid endarterectomy among 123,286 Medicare beneficiaries hospitalized for stroke in 2006 and 2007.

1.2 Introduction

Propensity scores (PS) are an often used tool for comparing the effectiveness of clinical treatments as they are applied in routine practice (*Rosenbaum and Rubin*, 1983). PS methods are used to estimate causal effects that are not confounded by observed characteristics. Traditionally, estimating causal effects with PS methods is achieved in two stages: 1) a 'PS stage' where a model is fit to predict the receipt of treatment from available covariates, with the predicted values

from this model representing the estimated PS, and 2) an 'outcome stage' whereby outcomes of treated and untreated units are compared among units with similar values of the PS. Typically, the two-stage nature of the problem is accommodated by separate and sequential estimation; a model is fit in the PS stage, then the estimated PS from this model are treated as fixed and known to conduct adjusted comparisons in the outcome stage. In this paper, we are considering a model base approach for both stages and will refer to the PS and outcome models from here forward.

Recently *McCandless et al.* (2009) proposed Bayesian estimation as a means to jointly estimate quantities in the PS and outcome models. One major motivation for Bayesian PS estimation is that jointly estimating quantities in the two models propagates uncertainty in estimation of the PS into estimation of the treatment effect, whereas one conceivable limitation of traditional sequential methods is that they potentially misstate the uncertainty in causal estimates by treating the estimated PS as a known quantity in the outcome stage (*Gelman and Hill*, 2007). The key idea with joint Bayesian PS estimation is that the PS is acknowledged as an unknown quantity, uncertainty about which is integrated out of posterior distributions of quantities in the outcome stage. Aside from providing a more comprehensive account of uncertainty, clear potential lies in incorporating PS methods into the broader literature on Bayesian methodology.

One important feature of joint modeling with Bayesian estimation is that doing so allows 'feedback' between the models. In the PS context, this means that posterior samples of parameters in the PS stage are informed in part by information from the outcome stage, rendering the problem of Bayesian PS estimation substantially more complex than a simple Bayesian analog to well-established procedures. In fact, the notion of estimation and use of the PS in a joint likelihood has generated some controversy. One view is that the PS is meant to approximate the design stage of a randomized study, and that this should be done without any access to the outcome in order to ensure objective design decisions that are completely separate from analysis decisions (*Rubin*, 2007, 2008). Nonetheless, methods that incorporate outcome information have been advocated (*Schneeweiss et al.*, 2009, *McCandless et al.*, 2009). In principle, incorporating feedback in joint

Bayesian estimation entails estimates of the PS themselves that make more complete use of the data, which could improve estimation of causal effects. However, a rigorous investigation of exactly how feedback can impact estimation of causal effects is lacking.

In what follows we illustrate that, in general, model feedback in joint Bayesian estimation can result in biased estimates of the treatment effect. Unlike traditional sequential procedures that estimate the PS based solely on information on how covariates relate to the treatment, we show that joint Bayesian estimation with feedback uses information from the outcome model to construct the PS, and that feedback from this model can distort the nature of the PS and impair its ability to adjust for confounding. We also demonstrate two techniques that can recover the causal effects: changing the nature of the feedback by using outcome models that augment PS adjustment with adjustment for individual covariates, and 'cutting' the feedback by using an approximately Bayesian sequential approach.

Using nationwide data on 123,286 Medicare beneficiaries, we illustrate joint Bayesain PS estimation in a comparative effectiveness investigation regarding the recent increase in the use of carotid artery stenting (CAS) for treatment of carotid artery disease (a primary cause of stroke), as compared to the more established carotid endarterectomy (CEA) procedure. Because these therapies are not randomly applied in clinical practice, we use several clinical characteristics to adjust for confounding when estimating a causal treatment effect. We compare the results of the joint Bayesian analysis and sequential Bayesian analysis both with and without individual covariate adjustment with a traditional sequential approach.

1.3 Propensity Score Estimation

For a binary treatment, $X = 0, 1$, an outcome, Y, and a vector of p covariates $(C_1, C_2, \ldots, C_p)$, *Rosenbaum and Rubin* (1983) defined the PS as the conditional probability of assignment to

treatment $X = 1$, given the covariates. Causal inference with the PS relies on two important features. First, treatment assignment must be assumed strongly ignorable; that is, there must be no unmeasured confounders. Second, by virtue of the fact that the PS reflects the treatment assignment mechanism, the PS enjoys the property of a *balancing score*, resulting in conditional independence between the treatment and the individual covariates, conditional on the score: $X \perp\!\!\!\perp C_1, \ldots, C_p | PS$. This balancing score property combined with the assumption of strongly ignorable treatment assignment allows average comparisons between treated and untreated outcomes at a given value of the PS to serve as an unbiased estimate of the average treatment effect at that value of the PS.

1.3.1 PS and outcome models

PS methods consist of two distinct parts: the estimation of the PS and estimation of the causal effect conditional on the PS. The PS model models the probability that $X = 1$ (given covariates): $g_x(E[X|C]) = C\gamma$, where $g_x(\cdot)$ is a link function, and C is the collection of pretreatment covariates plus an intercept, $C = (1, C_1, C_2, \ldots, C_p)$. Thus, the PS model can be represented with the following likelihood:

$$L(\mathbf{X}|\gamma, \mathbf{C}) = \prod_{i=1}^{n} [g_x^{-1}(C_i\gamma)]^{X_i} [1 - g_x^{-1}(C_i\gamma)]^{1-X_i}, \tag{1.1}$$

where here and throughout, boldface is used for vectors and matrices representing the values for the entire sample, and $i = 1, \ldots, n$ indexes observational units. With this formulation, the values of γ and C_i determine the PS for the i^{th} unit.

Consider a binary outcome, $Y = 0, 1$, but note that results in the following hold for other outcomes. We define a model for the outcome, conditional on the PS: $g_y(E[Y|X, C]) = \xi_0 + \beta X + \xi h(\gamma, C) + C^+\delta$, where $g_y(\cdot)$ is another link function, the deterministic function $h(\gamma, C)$ specifies how the PS enters the outcome model, and the term $C^+\delta$ denotes possible residual adjustment for some subset $C^+ \in C$ in addition to the PS. For example, $h(\gamma, C) = C\gamma$ would specify linear adjust-

ment for the linear predictor term from model (1.1), and $\delta = 0$ would indicate adjustment for the PS only. Alternatively, $h(\gamma, C)$ could specify dummy variables for membership in subclasses defined by q quantiles of the PS, and $\delta \neq 0$ would augment PS adjustment with individual covariate adjustment within subclass. We express the outcome stage likelihood as:

$$L(\mathbf{Y}|\beta, \xi, \mathbf{X}, \mathbf{C}, \gamma, \delta) = \prod_{i=1}^{n} [g_y^{-1}(\xi_0 + \beta X_i + \xi h(\gamma, C_i) + \delta C_i^+)]^{Y_i} [1 - g_y^{-1}(\xi_0 + \beta X_i + \xi h(\gamma, C_i) + \delta C_i^+)]^{1-Y_i}. \tag{1.2}$$

The primary objective is to estimate the causal effect of $X = 1$ vs. $X = 0$ on Y. Towards this end, the conditional parameter β may be of primary interest as this quantity represents the conditional (on the PS) causal effect. If the marginal effect is of primary interest, it can be obtained by marginalizing with respect to the empirical distribution of the covariates. Issues such as non collapsibility may prevent estimation of the marginal causal effect regardless of method used, but any effort to obtain the marginal effect requires estimation of β as a precursor step. Therefore, what follows equates estimation of causal effects to estimation of β for ease of illustration.

1.3.2 Traditional sequential estimation

Traditional PS procedures conduct estimation in the PS and outcome models completely separately. Estimates of γ are obtained from (1.1) to construct the estimated PS. Then, the estimated PS are treated as known quantities in the outcome model. That is, with estimated $\hat{\gamma}$, estimation of the treatment effect follows from $L(\mathbf{Y}|\beta, \xi, \mathbf{X}, \mathbf{C}, \hat{\gamma}, \delta)$ specified in (1.2).

An important feature of this approach is that it makes no attempt to recover the entire covariate-outcome relationship. Rather than specify a model for the relationship between each covariate and the outcome, the outcome model conditions on a one-dimensional summary of multivariate covariate information (the PS), with the dimension reduction specifically determined by fitting the PS model in (1.1). Of key importance is that this dimension reduction reflects the treatment

assignment mechanism to ensure the balancing score property. Other dimension reductions of C, e.g. with different values of γ, may fail to reflect $p(X = 1|C)$, and are not guaranteed to possess the balancing score property at the heart of PS methods.

With sequential estimation, estimates of γ from (1.1) are obtained in a manner that completely ignores quantities in the outcome model such as β, ξ, and Y. As we elaborate in the following sections, the primary difference with joint Bayesian estimation is the presence of feedback, which means that specification of the outcome model affects estimates of γ. The sequential Bayesian estimation ignores quantities in the outcome model when estimating γ but rather than treating the estimate of the PS as fixed quants, it considers their entire posterior distribution.

1.4 Bayesian Estimation and Model Feedback

In this section we formalize Bayesian PS estimation and illuminate in detail the role of model feedback. In contrast to the sequential procedure described in Section 1.3.2, Bayesian PS estimation combines the models in (1.1) and (1.2) into a single joint likelihood:

$$L(\mathbf{Y}, \mathbf{X}|\mathbf{C}, \gamma, \beta, \xi, \delta) = \prod_{i=1}^{n} [g_x^{-1}(C_i\gamma)]^{X_i}[1 - g_x^{-1}(C_i\gamma)]^{1-X_i} \times \tag{1.3}$$

$$[g_y^{-1}(\xi_0 + \beta X_i + \xi h(\gamma, C_i) + \delta C_i^+)]^{Y_i}[1 - g_y^{-1}(\xi_0 + \beta X_i + \xi h(\gamma, C_i) + \delta C_i^+)]^{1-Y_i}. \tag{1.4}$$

The likelihood in (1.3)-(1.4), together with the prior distribution for $(\gamma, \beta, \xi, \delta)$ serves as the basis for posterior inference. Recall that $h(\gamma, C)$ is a deterministic function of γ, which means that the PS themselves are treated as unknown quantities that are updated with every posterior update of γ. Model feedback in this case arises because both terms of the likelihood contribute to the posterior distribution of γ.

Throughout, we use a Metropolis-Hastings MCMC algorithm to sample from posterior distributions. We conduct the MCMC using two sampling blocks: one updating γ from its conditional posterior distribution, which corresponds to an update of the PS as well, and another block updating all parameters in the outcome model. Note from the likelihood in (1.3)-(1.4) that although updating γ conditional on (β, ξ, δ) -which corresponds to an update of the PS -will involve both terms of the likelihood, only term (1.4) - the likelihood for the outcome model - contributes to updating (β, ξ, δ) conditional on γ.

To illustrate the fundamental features of feedback implied by joint estimation of (1.3)-(1.4), the remainder of this section considers the simplified setting where the PS is included in the outcome model as a linear predictor; that is, we assume that $h(\gamma, C) = C\gamma$ and that $\xi = \xi_1$.

1.4.1 Algebraic Illustration of Feedback

Purely for illustration, take $g_x^{-1}(\cdot)$ and $g_y^{-1}(\cdot)$ as the Normal CDF, $\Phi(\cdot)$, representing Probit regression in the PS and outcome stages, and take all prior distributions $\propto 1$. Following *Albert and Chib* (1993), the Probit link allows Bayesian estimation with a data-augmentation procedure that iteratively samples latent continuous data from a truncated normal distribution with unit variance such that the latent $X^*(Y^*)$ are > 0 when $X = 1(Y = 1)$, and < 0 otherwise. Conditional on $(\mathbf{X}^*, \mathbf{Y}^*)$,

$$p(\gamma, \beta, \xi, \delta | \mathbf{X}^*, \mathbf{Y}^*, \mathbf{X}, \mathbf{C}) \propto \exp\{-\frac{1}{2}[(\mathbf{X}^* - \mathbf{C}\gamma)^{\mathbf{T}}(\mathbf{X}^* - \mathbf{C}\gamma) + (\mathbf{Y}^* - \xi_{\mathbf{0}}\mathbf{1_n} - \beta\mathbf{X} - \xi_{\mathbf{1}}\mathbf{C}\gamma - \mathbf{C}^{+}\delta)^{\mathbf{T}}(\mathbf{Y}^* - \xi_{\mathbf{0}}\mathbf{1_n} - \beta\mathbf{X} - \xi_{\mathbf{1}}\mathbf{C}\gamma - \mathbf{C}^{+}\delta)]\},$$

$\mathbf{C}$ is the $n \times (p+1)$ design matrix, and $\mathbf{1_n}$ is a $n-$dimensional vector with every entry equal to one. Thus, the conditional posterior distribution of γ can be written as:

$$p(\gamma | \mathbf{X}^*, \mathbf{Y}^*, \mathbf{X}, \mathbf{C}, \beta, \xi, \delta) \propto \exp\{\gamma^T(\mathbf{C^T C}(1 + \xi_1^2))\gamma - 2\gamma^T[\mathbf{C^T}(\mathbf{X}^* + \xi_1(\mathbf{Y}^* - \xi_0\mathbf{1_n} - \beta\mathbf{X} - \mathbf{C}^{+}\delta))]\}$$

which corresponds to the kernel of a Normal distribution with covariance matrix $(\mathbf{C^T C}(1+\xi_1^2))^{-1}$ and mean $(\mathbf{C^T C}(1+\xi_1^2))^{-1}(\mathbf{C^T}(\mathbf{X}^* + \xi_1(\mathbf{Y}^* - \xi_0\mathbf{1_n} - \beta\mathbf{X} - \mathbf{C}^+\delta)))$. Immediately we see that when $\xi_1 \neq 0$, quantities from the outcome model contribute to the posterior distribution of γ and, by extension, the PS. This is the nature of model feedback.

1.4.2 Implied parameterization of the covariate-outcome response surface

Until otherwise noted, asssume an outcome model that only adjusts for the PS; that is, assume $\delta = 0$. Considering the joint likelihood in (1.3)-(1.4) implies a parameterization of the covariate-outcome response surface conditional on X. We re-express $\xi_0 + \beta X + \xi h(\gamma, C)$ from term (1.4) as:

$$\xi_0 + \beta X + \xi_1(\gamma_0 + \gamma_1 C_1 + \ldots + \gamma_p C_p) = (\xi_0 + \xi_1\gamma_0) + \beta X + \xi_1\gamma_1 C_1 + \ldots + \xi_1\gamma_p C_p. \tag{1.5}$$

This parameterization implies that the covariate-outcome relationship for the k^{th} covariate is described by $\xi_1\gamma_k$, that is, that every covariate-outcome relationship is a rescaled version of the covariate-treatment relationship, with the same re-scaling factor (ξ_1) for every covariate. The key feature of model feedback is that posterior estimates of γ are informed in part by this parameterization of the outcome model, which may imply information about γ that is not consistent with the treatment assignment mechanism. In particular, this will occur if the underlying covariate-outcome relationship cannot be expressed as a simple rescaling of the covariate-treatment relationship.

To further illustrate, consider a simple setting where the true underlying relationships between p covariates, treatment, and outcome can be described as follows:

$$g_x(P(X_i = 1|C_i)) = \gamma_0 + \gamma_1 C_{i1} + \ldots + \gamma_p C_{ip} \qquad \text{and} \tag{1.6}$$

$$g_y(P(Y_i = 1|X_i, C_i)) = \alpha_0 + \beta X_i + \alpha_1 C_{i1} + \ldots + \alpha_p C_{ip}. \tag{1.7}$$

With the above data-generating mechanism, the joint likelihood in (1.3)-(1.4) with $\delta = 0$ correctly models (1.6), but entails linear adjustment for $g_x(PS)$, rather than a model for the complete covariate-outcome relationship in (1.7). Combining the above data-generating mechanism with the systematic component of the outcome model paramaterized as in the right hand side of (1.5) corresponds to $\gamma_0 = \frac{\alpha_0 - \xi_0}{\xi_1}$ and $\gamma_1 = \frac{\alpha_1}{\xi_1}, \gamma_2 = \frac{\alpha_2}{\xi_1}, \ldots, \gamma_p = \frac{\alpha_p}{\xi_1}$, meaning that the only way that the PS and outcome modeling stages can imply the same values of γ is if $\alpha_k = \xi_1 \gamma_k$ for all k. If this relationship does not hold, then feedback from the outcome model will yield posterior estimates of γ that do not reflect the true treatment-assignment mechanism in (1.6), meaning that $h(\gamma, C)$ is not technically a function of the PS and may not be a balancing score. Thus, Bayesian estimation with (1.3)-(1.4) and $\delta = 0$ is not guaranteed to yield estimates of β that reflect the causal treatment effect. In contrast, the sequential strategy in Section 1.3.2 estimates γ without regard to the outcome model, thus ensuring that $h(\gamma, C)$ maintains the balancing score property. We illustrate this phenomenon in the simulation study of Section 1.5.

1.4.3 Augmenting PS adjustment with individual covariates

The above feature of joint Bayesian PS estimation is not a feature of model feedback in general, but rather a byproduct of the dimension reduction implied by using the PS as a univariate summary of covariate information. Consider instead a model with $\delta \neq 0$ that adjusts for covariates in addition to the PS. With $h(\gamma, C) = C\gamma$, C^+ can include at most $(p-1)$ covariates to prevent perfect linear dependence in the design matrix for the outcome model. In this case, setting $C^+ = (C_2, \ldots, C_p)$, the right hand side of expression (1.5) becomes:

$$(\xi_0 + \xi_1\gamma_0) + \beta X + \xi_1\gamma_1 C_1 + (\xi_1\gamma_2 + \delta_1)C_2 + \ldots + (\xi_1\gamma_p + \delta_{p-1})C_p.$$

While setting $\delta \neq 0$ still implies feedback, the feedback does not impose the same restriction on the relationship between the covariate-treatment and covariate-outcome relationships which allows estimation of γ in accordance with the treatment assignment mechanism, thus maintaining the balancing score property. In other words, setting $\delta \neq 0$ allows the additional flexibility of modeling the covariate-outcome relationship without assuming that this relationship is a scalar multiple of the covariate-treatment relationship. The simulation study in Section 1.5 illustrates this phenomenon, and examines its benefits it situations where either the PS model or the covariate adjustment in the outcome model is misspecified.

1.4.4 Cutting the feedback

McCandless et al. (2010) present the idea of an approximately Bayesian method that 'cuts the feedback' from the outcome model to the PS model as an alternative to the fully Bayesian approach. We still use a Metropolis-Hastings MCMC algorithm but we do not sample from the joint posterior distribution. We cut the feedback from the outcome model to the PS model by first updating γ from the distribution defined by (1.3) and the prior distribution of γ. This posterior distribution ignores the likelihood contribution from (1.4). We then update (β, ξ, δ) given $\hat{\gamma}$ from the posterior defined by defined by (1.4) and the prior distribution of (β, ξ, δ). Cutting the feedback from the outcome model to the PS model eliminates any restrictions between the covariate-treatment/covariate-outcome relationship as they are modeled separately. Of note, the sequential Bayesian method primarily differs from the traditional sequential approach in that it does not treat the estimated PS as a fixed quantity. Rather, it makes use of the full posterior distribution of the PS by updating the estimated PS in the outcome model at every iteration of the MCMC. Residual confounding adjustment by allowing $\delta \neq 0$ is still possible but is no longer necessary to ensure $h(\gamma, \boldsymbol{C})$ maintains the balancing score property.

1.5 Simulation Study

In this section we present a simulation study to illustrate that the features described in the simplified setting of Section 1.4 persist in settings with more flexible specification of $h(\gamma, C)$. All simulated datasets contain $n = 1000$ observations and $p = 6$ covariates, simulated from the following data-generating scheme. First, $C_1, \ldots, C_6$ are simulated from a multivariate normal distribution with mean $(0, 0, 0, 0, 0, 0)$ and the identity covariance matrix. For all i, X_i is simulated from a Bernoulli distribution with:

$$P(X_i = 1|C_i) = \frac{\exp(\gamma_0 + \gamma_1 C_{i1} + \ldots + \gamma_6 C_{i6})}{1 + \exp(\gamma_0 + \gamma_1 C_{i1} + \ldots + \gamma_6 C_{i6})}. \tag{1.8}$$

All Y_i are similarly generated from Bernoulli distributions with:

$$P(Y_i = 1|X_i, C_i) = \frac{\exp(\alpha_0 + \beta X_i + \alpha_1 C_{i1} + \ldots + \alpha_6 C_{i6})}{1 + \exp(\alpha_0 + \beta X_i + \alpha_1 C_{i1} + \ldots + \alpha_6 C_{i6})}. \tag{1.9}$$

The values of γ specify the true treatment assignment mechanism, those of α specify the true covariate-outcome relationship, and β is the conditional treatment effect. For all simulations, we set $\beta = 0.0$.

We simulated 1000 data sets under each scenario described below, and analyzed the simulated data with the joint Bayesian method described in Section 1.4, both with and without residual confounding adjustment, and with the sequential Bayesian method described in Section 1.4.4. For comparison, we obtain maximum likelihood estimates of β using the traditional sequential procedure of Section 1.3.2 and from fitting model (1.9) directly, referring to the latter as the 'Gold Standard' since we know that this is the true data-generating mechanism.

Throughout analysis of the simulated data, we specify both $g_x^{-1}(\cdot)$ and $g_y^{-1}(\cdot)$ as $\frac{\exp(\cdot)}{1+\exp(\cdot)}$, indicating logistic regression in both model stages. Unlike the simple illustrations provided in Section 1.4, we take a more flexible modeling approach that stratifies units on quintiles of the logit(PS) and estimates the same β across PS strata. Adjustment for PS subclass is augmented with additional covariate adjustment ($\delta \neq 0$) when noted. For the Bayesian analyses, every posterior update of γ implies an update of the PS, so the quintiles of logit(PS) are recalculated and the PS subclasses redefined at every MCMC iteration. We specify diffuse prior distributions for all parameters as Normal with mean 0 and variance 10^{10}. In addition to comparing estimates of β, we also compare methods on the basis of estimates of γ, which determine the estimated PS. For point estimation, we use posterior mean estimates for the Bayesian methods, obtained from three MCMC chains, each run for 10,000 iterations, with the first 5,000 discarded as burn in and every 10^{th} sample saved for posterior inference. Note here that application of PS methods in practice should involve an investigation of whether covariates are balanced within PS subclass, which we forego in the simulation study. Balance checks are addressed in detail for the data analysis in Section 1.7.

1.5.1 Scenario where the covariate-outcome relationship is a simple rescaling of the covariate-treatment relationship

Scenario 1 generates data with parameters in (1.8) and (1.9) set to $(\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6) = (0.0, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3)$ and $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6) = (0.0, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5)$. This scenario represents a unique special case where $\gamma = \frac{\alpha}{\xi_1}$ and where the joint bayesian method should be capable of recovering the treatment effect without augmenting the outcome model with additional covariate adjustment.

We analyze the data with $\delta = 0$. Figure 1.1 depicts boxplots of the resulting posterior estimates of γ and β for both the joint Bayesian sequential Bayesian methods, along with estimates from the traditional sequential approach. We see that, on average, all three methods produce point estimates of γ that are similar and agree with the true parameter values from (1.8). For $\gamma_1, \ldots, \gamma_6$,

point estimates are less variable with the joint Bayesian method, which is to be expected because posterior distributions of these quantities involve additional information via feedback from the outcome model. Estimates of β are also similar between the methods. Again, this simulation illustrates the special case where the PS and outcome models imply the same values of γ, so posterior estimates of $h(\gamma, C)$ reflect the treatment assignment mechanism and the joint Bayesian method estimates the causal effect.

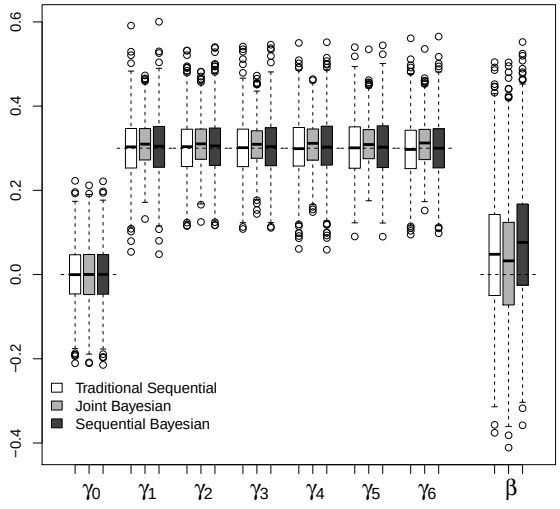


Figure 1.1: Scenario 1 with $(\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6) = (0.0, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3)$, and $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6) = (0.0, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5)$: boxplots of estimates of γ and β from the traditional sequential, joint Bayesian and sequential Bayesian analyses of 1000 replicated data sets. Horizontal dotted lines are at the true parameter values.

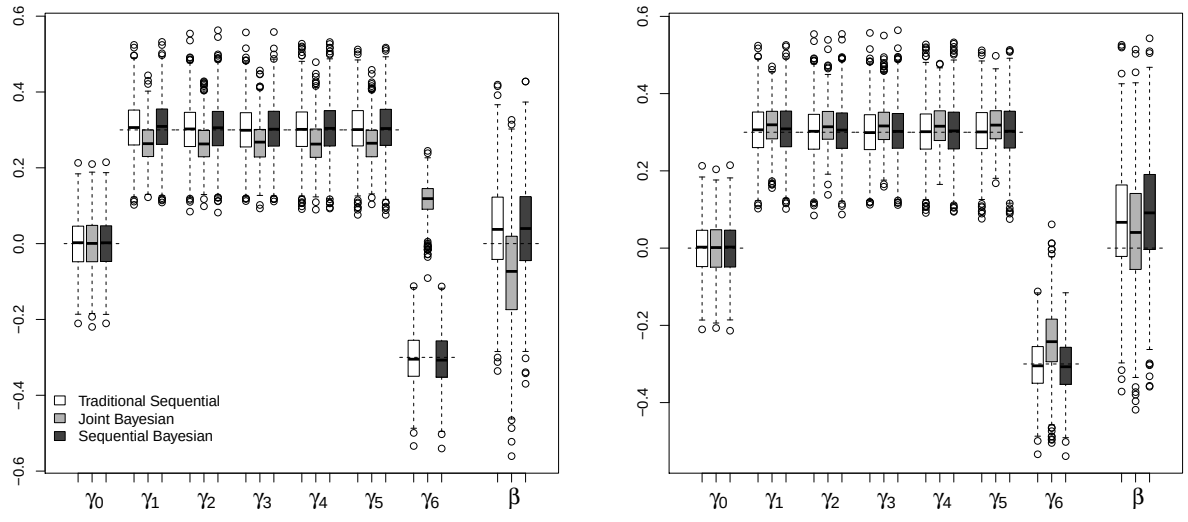
1.5.2 Scenario where one covariate exhibits a different covariate-treatment/ covariate-outcome relationship

Appealing to the discussion in Section 1.4, we simulate Scenario 2 with 5 covariates having the same covariate-treatment/covariate-outcome relationship, with the sixth covariate exhibiting a different relationship. This setting illustrates the effect that model feedback can have on joint Bayesian estimation when the covariate-outcome response surface cannot be expressed as a simple rescaling of the covariate-treatment surface. Towards this end, we simulate data as in

Scenario 1, except we change γ_6 to -0.3 so that $\gamma \neq \frac{\alpha}{\xi_1}$.

We first analyze the data with $\delta = 0$. Figure 1.2a depicts boxplots of estimates of γ and β from all three estimation methods. Unlike in Scenario 1, we see that, on average, the joint Bayesian method produces different estimates of $\gamma_1, \ldots, \gamma_6$ than either the sequential Bayesian method or traditional sequential method, whereas the latter two estimates agree. While both sequential approaches estimate γ in accordance with the treatment assignment mechanism in (1.8), the joint Bayesian method estimates different values of γ, with the most pronounced difference for γ_6. In the joint Bayesian method, the quantity $h(\gamma, C)$ does not reflect the treatment assignment mechanism, and is not guaranteed to serve as a balancing score. The result is posterior estimates of β with poor performance relative to estimates from the sequential procedures. This illustrates how feedback can distort the balancing score property of the PS and yield estimates of β that do not reflect a causal effect.

We argued in Section 1.4.3 that augmenting PS adjustment with individual covariates can prevent feedback from distorting estimates of γ in the joint Bayesian approach. Because we know in this simulated example that one covariate exhibits a different relationship with the treatment, we re-analyze these simulated data sets with an outcome model that adjusts for C_6 within PS subclass. That is, we let $\delta \neq 0$ and $C^+ = C_6$ in (1.2), referring to this analysis as Scenario 2^+. Point estimates from this analysis are compared in Figure 1.2b. We include the sequential Bayesian method for comparison purposes, although as noted in section 1.4.4 and shown in Figure 1.2a, this adjustment is not necessary for this method to maintain the balancing property. Compared to the analysis that adjusts only for the PS, the model that augments PS estimation with additional adjustment of C_6 produces estimates of $\gamma_1, \ldots, \gamma_6$ that are much more similar between the three estimation methods, implying that the joint Bayesian method with $\delta \neq 0$ comes closer to capturing the true treatment assignment mechanism. As a result, estimates of β are similar in the joint Bayesian and sequential estimation approaches, although the methods do not produce the exact same estimates.



(a) Scenario 2, $\delta = 0$

(b) Scenario 2^+, $\delta \neq 0$

Figure 1.2: Scenarios 2 and 2^+ with $(\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6) = (0.0, 0.3, 0.3, 0.3, 0.3, 0.3, -0.3)$, and $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6) = (0.0, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5)$: boxplots of estimates of γ and β from the sequential frequentist, joint Bayesian and sequential Bayesian analysis of 1000 replicated data sets. Horizontal dotted lines are at the true parameter values.

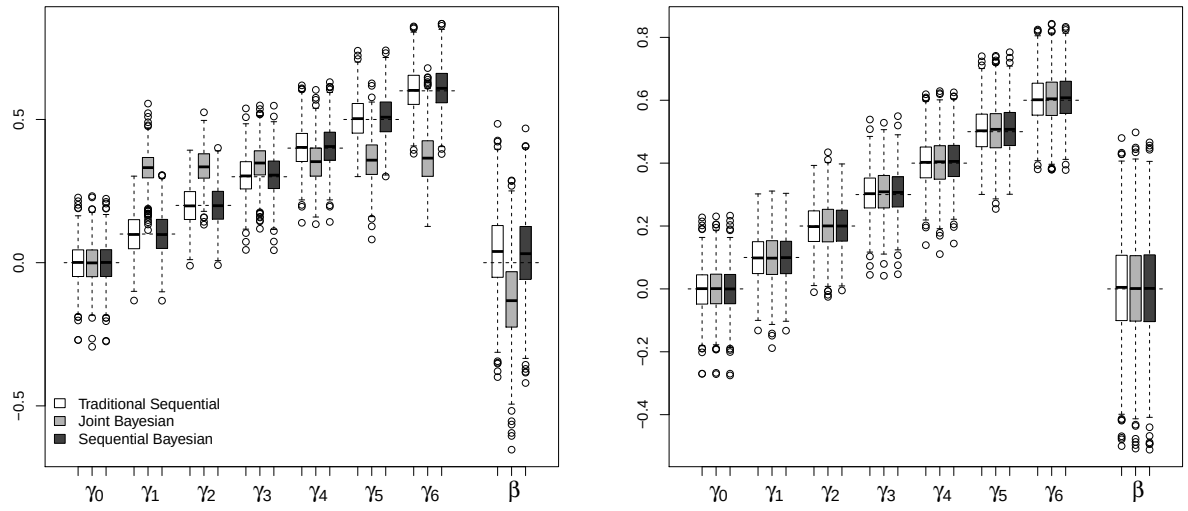
1.5.3 Scenario where every covariate exhibits different covariate-treatment/ covariate-outcome relationship

Finally, we simulate Scenario 3 so that the covariate-treatment/covariate-outcome relationship is different for every covariate. For the PS model (1.8) we set $(\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6) = (0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6)$. For the outcome model (1.9) we set $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6) = (0.0, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1)$.

We first analyze the data with $\delta = 0$. From Figure 1.3a, we see that the joint Bayesian method provides estimates of $\gamma_1, \ldots, \gamma_6$ that are all shrunken towards 0.35 (the average value of $\gamma_1, \ldots, \gamma_6$), which is a consequence of estimating these quantities with feedback from an outcome model that imposes restrictions on the covariate-treatment and covariate-outcome relationships. This is in stark contrast to the estimates from the sequential methods that are not informed by the outcome and accurately reflect a different γ_k for $k = 1, 2, \ldots, 6$. We also see that these vast discrepancies

between estimates of γ lead to estimates of β that are very different, with the joint Bayesian estimates performing very poorly. In a setting where the covariate-treatment/covariate-outcome relationship is different for every covariate, joint Bayesian estimation with $\delta = 0$ cannot adequately recover the treatment effect, even though sequential methods perform well.

We reanalyze the data simulated in Scenario 3 with $\delta \neq 0$ and $C^+ = (C_1, \ldots, C_6)$, referring to this analysis as Scenario 3^+. Results for these analyses are summarized in Figure 1.3b, which shows that the additional covariate adjustment in the outcome model prevents feedback from distorting estimates of γ, leading to estimates of γ from the joint Bayesian method that agree, on average, with those from the sequential procedures and the true treatment assignment mechanism. As a consequence, $h(\gamma, C)$ maintains the balancing score property, and Bayesian estimates of β agree very closely with estimates from the sequential procedure and with the true parameter value.



(a) Scenario 3, $\delta = 0$

(b) Scenario 3^+, $\delta \neq 0$

Figure 1.3: Scenarios 3 and 3^+ with $(\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6) = (0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6)$, and $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6) = (0.0, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1)$: boxplots of estimates of γ and β from the sequential frequentist, joint Bayesian and sequential Bayesian analyses of 1000 replicated data sets. Horizontal dotted lines are at the true parameter values.

Table 1.1 numerically summarizes the performance of each method in terms of estimates of β. This table also summarizes bias from an unadjusted analysis, which is high in all scenarios. The joint Bayesian analyses of scenarios containing different covariate-outcome and covariate-treatment relationships that do not augment PS adjustment (Scenarios 2 and 3) produce estimates of β with substantial bias, as compared to the sequential approaches and to the Gold Standard analysis. Joint Bayesian estimates for these scenarios also exhibit low coverage probabilities.

For Scenarios 1, 2^+, and 3^+ all methods performed comparably. It is interesting to note that, for the scenarios where one covariate exhibits a different covariate-treatment relationship (2 and 2^+), augmenting with the additional covariate actually resulted in slightly more bias for the sequential Bayesian method. The sequential Bayesian method also has conservative confidence interval coverage, with coverage at least 98% in all scenarios. The interval widths are nearly twice as wide as those from either the traditional sequential or joint Bayesian methods as the posterior standard deviation seems to overestimate the standard deviation of the posterior mean. (results not shown).

It is also important to note that the detrimental effects of feedback on causal estimates when $\delta = 0$ (as displayed in Scenarios 2 and 3) is a feature of the dimension-reduced feedback explicated in Section 1.4.2 and that this phenomenon cannot be remedied by increasingly flexible choices for $h(\gamma, C)$. To illustrate this point, Appendix A.1.1 conducts a simulation study paralleling that in Scenarios 3 and 3^+, but specifying a flexible spline basis for $h(\gamma, C)$. The results of this simulation are the same as those presented here; estimates of β are biased when $\delta = 0$, but not when $\delta \neq 0$, the latter case being analogous to the penalized spline of propensity prediction method of *Little* (2011).

1.6 Model Misspecification

We have shown that model feedback has the potential to distort effect estimates when doing joint Bayesian PS estimation. One recommendation to overcome this is when conducting joint Bayesian

Table 1.1: Numerical performance comparison of estimates of β from the unadjusted analysis (Unadj), traditional sequential (Seq), joint Bayesian method (JB) and sequential Bayesian (SB) methods and gold standard analysis (Gold). $\widehat{SE(\hat{\beta})}$ is the average standard error estimates or posterior standard deviations across replicated data sets. Coverage refers to the proportion of 95% confidence or posterior probability intervals that contain β.

	$Bias(\hat{\beta})$					$MSE(\hat{\beta})$				$\widehat{SE(\hat{\beta})}$				95 % Coverage			
	Unadj	Freq	JB	SB	Gold	Freq	JB	SB	Gold	Freq	JB	SB	Gold	Freq	JB	SB	Gold
Scenario 1	0.63	0.05	0.03	0.08	0	0.03	0.03	0.03	0.02	0.15	0.15	0.28	0.15	0.93	0.94	1	0.95
Scenario 2	0.42	0.04	-0.08	0.04	0	0.02	0.03	0.02	0.02	0.14	0.15	0.32	0.15	0.96	0.93	1	0.96
Scenario 2+	0.42	0.07	0.04	0.09	0	0.02	0.02	0.03	0.02	0.15	0.15	0.28	0.15	0.95	0.95	1	0.96
Scenario 3	0.4	0.04	-0.13	0.03	0	0.02	0.04	0.02	0.02	0.14	0.15	0.25	0.15	0.95	0.86	1	0.94
Scenario 3+	0.4	0	0	0	0	0.02	0.03	0.03	0.02	0.15	0.15	0.20	0.15	0.94	0.94	0.98	0.94

estimation with models for the PS and outcome stage augment the PS adjustment with adjustment for every covariate that appears in the PS model, a strategy akin to those previously developed to yield 'doubly robust' estimators that will estimate causal effects when either the PS model or the model for additional adjustment is correctly specified (*Bang and Robins*, 2005, *Little*, 2011). In this section, we use simulations to demonstrate that if we correctly specify the functional form of the covariate adjustment in the outcome model, we will get unbiased effect estimates, even if we misspecify the PS model. Likewise, if we misspecify the covariate adjustment in the outcome model but have the PS model correctly specified, we also get unbiased effect estimates. This is true, regardless of the method used (joint Bayesian, sequential frequentist or sequential Bayesian) but *only* if we include all C for residual confounding adjustment.

1.6.1 Misspecification of the Outcome Model

First, let's consider the case where the PS model is correctly specified but the residual covariate adjustment in the outcome model does not reflect the true data generating mechanism. We generate data from the situation where the covariate-treatment/covariate-outcome relationship are not simple rescalings - in other words, the residual confounding is necessary for the PS to maintain the balancing score property when using the joint Bayesian method. Specifically, after generating C as described in section 1.5, we generate X and Y from Bernoulli distributions with probabilities as follows

$$P(X_i = 1|C_i) = \frac{\exp(\gamma_0 + \gamma_1 C_{i1} + \ldots + \gamma_6 C_{i6})}{1 + \exp(\gamma_0 + \gamma_1 C_{i1} + \ldots + \gamma_6 C_{i6})}$$

$$P(Y_i = 1|X_i, C_i) = \frac{\exp(\alpha_0 + \beta X_i + \alpha_1 \log|C_{i1}| + \ldots + \alpha_6 \log|C_{i6}|)}{1 + \exp(\alpha_0 + \beta X_i + \alpha_1 \log|C_{i1}| + \ldots + \alpha_6 \log|C_{i6}|)}$$

where $(\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6) = (0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6)$, $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6) = (0.0, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1)$ and $\beta = 0$.

We analyzed the simulated data with $\delta \neq 0$ and $C^+ = (C_1, \ldots, C_6)$ using the joint Bayesian, sequential Bayesian and sequential frequentist approaches. Results for these analyses are summarized in Figure 1.4. We see that even though misspecified, the additional covariate adjustment in the outcome model still changes the nature of the feedback such that it does not distort estimates of γ, leading to joint Bayesian estimates of γ that agree, on average, with those from the sequential procedures and the true treatment assignment mechanism. As a consequence, $h(\gamma, C)$ maintains the balancing score property, and joint Bayesian estimates of β agree very closely with estimates from the sequential procedures and with the true parameter value. With a correctly specified PS model, we expect $X \perp\!\!\!\perp C_1, \ldots, C_p | PS$. It was the restriction on the covariate-treatment/covariate-outcome relationship imposed by fitting the joint likelihood that distorted estimates of γ, the PS and ultimately β. Adjusting for these covariates, even if the functional form of the adjustment does not reflect the true data generating mechanism, still allows estimation of γ in accordance with the treatment assignment mechanism, thus maintaining the balancing score property.

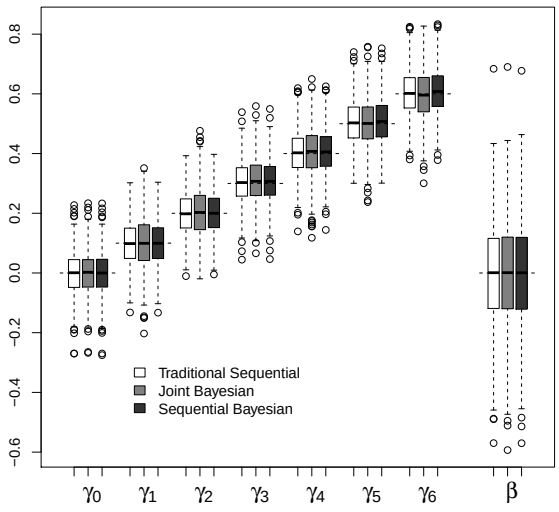


Figure 1.4: Misspecified outcome model: boxplots of estimates of γ and β from the sequential frequentist, joint Bayesian and sequential Bayesian analysis of 1000 replicated data sets. The dashed line represents the true parameter values.

1.6.2 Misspecification of the Propensity Score Model

Now, let's consider the case where the additional covariate adjustment in the outcome model reflects the true data generating mechanism but the PS model is incorrectly specified. In addition to the PS model misspecification, $\gamma \neq \frac{\alpha}{\xi_1}$; in other words, even if the functional form was correctly specified, we would still need additional covariate adjustment in the outcome model for $h(\gamma, C)$ to maintain the balancing score property in the joint Bayesian method. Specifically, after generating C as described in section 1.5, we generate X and Y from Bernoulli distributions with probabilities as follows

$$P(X_i = 1|C_i) = \frac{\exp(\gamma_0 + \gamma_1 \log|C_{i1}| + \ldots + \gamma_6 \log|C_{i6}|)}{1 + \exp(\gamma_0 + \gamma_1 \log|C_{i1}| + \ldots + \gamma_6 \log|C_{i6}|)}$$

$$P(Y_i = 1|X_i, C_i) = \frac{\exp(\alpha_0 + \beta X_i + \alpha_1 C_{i1} + \ldots + \alpha_6 C_{i6})}{1 + \exp(\alpha_0 + \beta X_i + \alpha_1 C_{i1} + \ldots + \alpha_6 C_{i6})}$$

where $(\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6) = (0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6)$, $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6) = (0.0, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1)$ and $\beta = 0$.

We analyzed the simulated data with $\delta \neq 0$ and $C^+ = (C_1, \ldots, C_6)$ using the joint Bayesian, sequential Bayesian and sequential frequentist approaches. Results for these analyses are summarized in Figure 1.5. We see that all three methods closely agree as to their estimates of β. In this case, the outcome model, minus $h(\gamma, C)$, is the true, 'gold standard' model. Adding in the misstated PS is essentially just adding in random noise and does not bias our estimates of β.

1.7 Comparing the Effectiveness of Cardiovascular Treatments

Carotid artery stenting (CAS) has recently emerged as a promising non-inferior alternative to carotid endarterectomy (CEA) for treatment of carotid artery disease, which is a primary cause

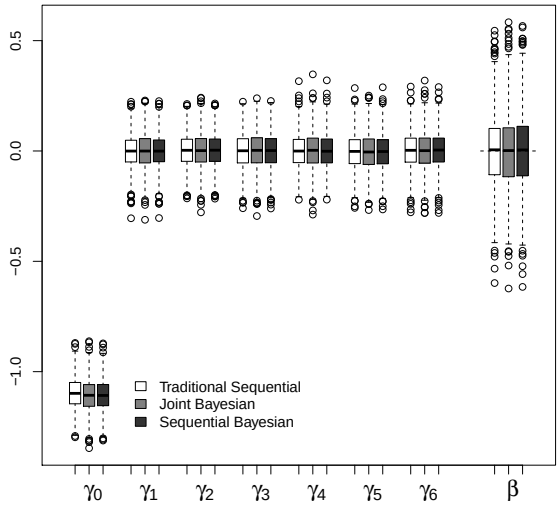


Figure 1.5: Misspecified PS model: boxplots of estimates of γ and β from the sequential frequentist, joint Bayesian and sequential Bayesian analysis of 1000 replicated data sets. The dashed line represents the true value of β. While the estimates of γ have no real meaning in this setting, we see that, on average, we estimate β without bias.

of stroke. To compare CEA ($X = 1$) vs. CAS ($X = 0$) for preventing death within one year of hospital admission ($Y = 1$ for death, 0 otherwise), we use hospital impatient data from 123,286 Medicare beneficiaries admitted to the hospital with a primary diagnosis of stroke during 2006 or 2007, as determined by the diagnosis codes found in *Lichtman et al.* (2009). An unadjusted comparison between 1-year mortality in CEA vs. CAS patients yields an odds ratio for death of 0.59 indicating worse outcomes with CAS, but this comparison is thought to be confounded by patient characteristics that help determine treatment choice. In particular, CAS patients generally have a higher baseline risk profile, as evident from Table 1.2, which summarizes patient characteristics in the CEA and CAS treatment groups. In pursuit of a causal effect estimate, we conduct a PS analysis that adjusts for the 25 variables in Table 1.2, including patient ethnicity, age, and gender, as well as baseline risk factors consisting of the Hierarchical Condition Categories (HCC) (*Pope et al.*, 2004) for current or previous presence of comorbidities.

We conduct the analysis using logistic regression in both the PS and the outcome stages, with $h(\gamma, C)$ specifying PS subclasses based on quintiles of the logit(PS). We checked that the entire range of PS values was represented in both treatment groups (i.e., that there was sufficient overlap) using maximum likelihood estimates of γ. For the Bayesian methods, the quintiles for defining PS subclasses were recalculated for every update of the PS. In light of the discussion in Sections 1.4 and 1.5, we consider an outcome model with $\delta \neq 0$ and $C^+ \equiv C$, implying residual adjustment for every covariate within PS subclass. We estimate the treatment effect using the joint Bayesian analysis of Section 1.4, the sequential Bayesian analysis of Section 1.4.4 as well as with a standard sequential analysis. Prior distributions for all parameters were considered Normal with mean 0 and variance 10^{10}. Three MCMC chains were run for 100,000 iterations, discarding the first 25,000 as burn in and saving every 20^{th} sample for posterior inference.

1.7.1 Results

From the joint Bayesian analysis, the posterior mean of the conditional causal odds ratio (OR), e^{β}, was 0.68, with a 95% posterior probability interval (0.61, 0.77), indicating a decreased odds of death within 1 year of hospital admission for CEA patients as compared to CAS patients. The analogous traditional sequential analysis produced the same point estimate and 95% confidence interval while the sequential Bayesian analysis produced the same point estimate but the wider interval (0.58, 0.80). Thus, our analysis fails to provide evidence that CAS is a non-inferior alternative to CEA for treating carotid artery disease in stroke patients, with increased conditional odds of death within 1-year of hospital admission among patients treated with CAS. As in our simulation study, all three analyses yield virtually identical results when $\delta \neq 0$.

We also note that for the Bayesian analyses MCMC performance was suspect for many parameters in the PS model, although performance was adequate for all parameters in the outcome model, including β. We revisit this point in the discussion.

Table 1.2: Baseline characteristics (% experiencing unless noted) and 1-year mortality rate for CAS and CEA patients.

	CAS (n=4038)	CEA (n=119248)
Age (mean)	75.3	75.1
White	92.3	93.8
Male	62.1	57.3
Prior Myocardial Infarction	5.1	2.1
Unstable Angina	5.2	2.5
Chronic Atherosclerosis	64.3	48.6
Respiratory Failure	3.3	1.9
Hypertension	75.3	78.8
Prior Stroke	7.5	6.7
Cerebrovascular Disease (non stroke)	26.7	17.1
Renal Failure	10.5	6
COPD	26.1	22.4
Pneumonia	5.4	3.6
Diabetes	35.3	32.3
Malnutrition	1.1	0.7
Dementia	3.6	3.1
Functional Disability	5.1	3.8
Peripheral Vascular Disease	15.2	9
Trauma in the Past Year	4	3.4
Major Psychiatric Disorder	1	1
Anemia	15.5	12.3
Depression	3.9	4.7
Parkinsons/Huntingtons	1.1	0.8
Asthma	1.7	2.6
Cancer	4.7	4.2
Death within 1 year of Admission	9.3	5.6

1.8 Discussion

Through a detailed assessment of model feedback, we have advanced existing research on Bayesian PS estimation. Using a simple example and simulated illustrations, we have shown that a joint likelihood for a PS model and an outcome model that adjusts for only the PS cannot uncover treatment effects in general settings. The key concept is that outcome models that adjust for the PS imply a characterization of the covariate-outcome response surface (conditional on X), and feedback from this outcome model can distort estimates from the PS model and compromise the desirable features of PS adjustment. This casts substantial doubt on the validity of using joint Bayesian PS estimation for an outcome model that adjusts for only the PS, and represents a vital feature that has been previously overlooked in the literature on Bayesian PS estimation.

One constructive approach that we explore here augments PS adjustment with additional covariate adjustment, which has been previously recommended in the PS literature (*Rubin*, 1985, *Stuart*, 2010). We have shown that joint Bayesian estimation using this strategy can accurately estimate the treatment effect in settings where adjustment for only the PS fails. Our recommendation is that, when conducting joint Bayesian estimation with models for the PS and outcome stage, PS adjustment should be augmented with adjustment for every covariate that appears in the PS model. Although this strategy could still provide substantial benefit over methods for direct covariate adjustment that do not use the PS (*Rubin*, 1985), adjusting for each individual covariate within PS subclass may be unappealing to researchers drawn to PS methods precisely because of their ability to provide reliable causal estimates without specifying every covariate in an outcome model. If, when specifying a model for the PS and a model for the outcome, researchers wish not to augment PS adjustment with adjustment for every covariate, then we recommend against using the type of joint Bayesian estimation presented here.

In that situation a researcher may use an approximately Bayesian approach that 'cuts the feedback' from the outcome model to the PS model. This method has an advantage in that it

treats the PS as the unknown quantity that it is and propagates uncertainty from its estimation into the outcome model. Additionally, this method fits into the broader literature on Bayesian methodology and may be incorporated into other methods that seek to use Bayesian approach to propensity scores, for example, methods to adjust for missing confounders using propensity scores (*McCandless et al.*, 2012).

In comparison with traditional sequential procedures, Bayesian PS estimation implies a significant computational burden. In the analysis of the Medicare data, achieving adequate MCMC performance and chain mixing was challenging for parameters in the PS model - which can be considered nuisance parameters in a PS analysis - particularly for the joint Bayesian method.

Our goal for this work is to shed light on the subtlety of model feedback when conducting joint Bayesian PS estimation when a model is used to conduct outcome comparisons adjusted for the PS. To achieve this goal, we made several simplifying assumptions. In particular, we specified an outcome model that stratified on PS quintiles, but assumed the same treatment effect across all PS subclasses. In analyzing the Medicare stroke data, we investigated the use of additional PS subclasses and the inclusion of PS-by-treatment interaction terms in (1.2) to estimate a different treatment effect in each subclass, but this did not qualitatively alter our results. Other interactions or more complicated modeling strategies could be implemented in either the PS stage or the outcome stage, but the salient features of model feedback would persist, as shown by *Zigler et al.* (2013). We also note that the entire joint Bayesian estimation paradigm relies on a likelihood based approach to both a PS model and an outcome model, and the issues addressed in this article have no clear analog to PS methods that exchange likelihood-based inference for matching or weighting in the outcome stage. Furthermore, the entirety of this article is predicated on the assumption of ignorable treatment assignment. While this assumption held by design in our simulation study, our results regarding the comparative effectiveness of CEA vs. CAS should be viewed in light of the prospect of unmeasured confounding, which may be present in our

example as the Medicare data lacks specific information on condition severity.

Better understanding of model feedback is essential to advance research on Bayesian methodology for use in problems involving the PS. For example, there has been recent interest in PS estimation when the set of necessary confounders is an unknown subset of those available for analysis (*Wang et al.*, 2012, *McCandless*, 2012, *Vansteelandt*, 2012). In principle, conducting Bayesian variable selection jointly on the PS and outcome models could ensure that important outcome predictors are included in the PS model, but our results here show that using model feedback to estimate coefficients in the PS model could prove detrimental. The sequential Bayesian method explored here would sacrifice the ability of the outcome to inform which variables to include in the PS. In another example of joint Bayesian PS estimation, *McCandless et al.* (2012) use PS ideas to adjust for confounding using external validation data within a joint Bayesian model, but do not directly address the role of feedback. Chapter 4 builds on this work by suggesting two approximately Bayesian approaches that do not allow feedback from the outcome model to the PS model. Investigation of feedback in these and other settings is an important avenue for future research, and provides sound motivation for further pursuit of Bayesian PS methods.

Bayesian Adjustment for Confounding in the Presence of Multiple Exposures

Abstract

Modern air pollution epidemiology demands a shift to considering the health effects of exposure to multiple pollutants but most epidemiological studies examine health effects associated with exposure to a single environmental contaminant at a time. For example, let's assume we are interested in estimating the effect of simultaneous exposure to ozone and $PM_{2.5}$ on cardiovascular (CVD) hospitalization in an observational study of 413 U.S. counties. We have data on over 50 measured confounders. There is limited literature that provides clear guidance on how to select confounders to include in the health effects model when the goal is multiple pollutant risk estimation. We propose a method to estimate the adverse health effect associated with a simultanous change in more than one exposure while addressing uncertainty in the selection of the confounders. We introduce Bayesian Adjustment for Confounding for Multiple Exposures (BAC-ME). For the situation with J exposure variables, our approach is based on specifying $J+1$ regression models, one for each of the exposure variables and one for the health effects model. The J regression models have each of the exposure variables as response variables and the set of measured confounders as predictor variables. We perform Bayesian variable selection on all models and link them through our specification of prior odds of including a predictor in the outcome model, given its inclusion in the exposure models. In simulation studies we show that our method estimates the multi pollutant adverse effect with smaller bias and mean squared error than traditional Bayesian Model Averaging (BMA) or adaptive LASSO and with improved coverage. We then apply BAC-ME, BMA and adaptive LASSO to an epidemiological study of over 14 million medicare enrollees for the study period 2008 to 2010. Using each approach, we estimate the change in emergency hospital admissions associated with a simultaneous change in long term exposure to both ozone and $PM_{2.5}$ adjusted for confounding.

2.1 Introduction

Most epidemiological studies examine health effects associated with exposure to a single environmental contaminant at a time. However, humans are exposed to many environmental agents

at once and therefore epidemiological studies need to change focus to this more realistic setting. For instance, suppose we are interested in estimating the adverse health effect associated with the simultaneous exposure to more than one pollutant, say ozone and $PM_{2.5}$. There may be any number of confounding factors we would like to account for. For example, in a retrospective epidemiological study of chronic health effects associated with long term exposure to both ozone and $PM_{2.5}$ potential confounders include weather variables, other pollutants (e.g. nitrogen dioxide, carbon monoxide and sulfur dioxide), geographic region and population characteristics. It may be impractical, impossible or undesirable to adjust for all possible confounders and yet we are not certain which are truly important. As the number of exposures included in the analysis for the estimation of a multi pollutant adverse health effect increases, so does the chance of excluding an important confounder from a large set of measured covariates.

One challenge with the transition from a single exposure to multiple exposures is the lack of a formal approach to select which measured confounders should be included in the health effects model. Standard approaches for selecting confounders in the context of a single exposure will not be adequate in this context; the set of confounders of an adverse health effect associated with simultaneous exposure to more than one pollutant cannot be fully characterized by the confounders of the effect of each individual pollutant separately. The key task is to identify confounders that are jointly associated with multiple exposures and the outcome.

Confounding adjustment in the epidemiological literature frequently relies on regression adjustment, and many air pollution studies have used a regression framework to identify the most toxic of a large set of exposures after adjustment for a pre-specified set of measured confounders (*Robins et al.*, 1992, *Greenland*, 1993, *Vedal and Kaufman*, 2011, *Dominici et al.*, 2010). In this paper we consider a different problem. Researchers are often confronted with choices regarding which of the available covariates should be included for confounding adjustment, especially when the number of variables is large relative to the sample size. In practice, they select a subset a priori based on some selection criteria: 'subject matter expert' knowledge, availability of data, etc.

Whatever method is used, there is always uncertainty surrounding that choice. Here we consider the question of which confounders to include in the health effects model (also called outcome model) when interest lies in multi pollutant risk estimation. In the single exposure setting, BAC (*Wang et al.*, 2012) has been recently introduced as a method to select confounders. BAC uses a Bayesian approach to model averaging to estimate the health effect associated with exposure to a single pollutant while acknowledging the uncertainty in the confounder selection. To our knowledge, the literature is lacking with respect to methods for confounding adjustment for the situation with multiple exposures.

In this paper, we will make two contributions. First, we will clarify the difference as to what constitutes a true confounder in multiple exposure settings versus single exposure settings. A true confounder in the multiple exposure (ME) setting is any covariate that confounds the relationship between simultaneous exposure to multiple pollutants and the outcome of interest. This could be a covariate that is marginally associated with one or more exposures - and, hence, would also be a confounder in the single exposure (SE) setting - or one that is jointly associated with multiple exposures (and might not be a confounder in the SE setting). Throughout this paper, when we refer to a true confounder we are referring to a true confounder in the ME setting unless otherwise specified. Second, we will develop a statistical framework to adjust for confounding in the presence of multiple exposures while accounting for uncertainty in the confounding adjustment. We introduce BAC for multiple exposures (BAC-ME) to extend this framework where selection of confounders is based on simultaneous exposure to multiple pollutants. Our method will allow us to select a subset of covariates to include to control for confounding in a linear regression model while protecting against the possibility of eliminating a true confounder. This will also help identify true confounders for future research efforts. We will show through simulation studies that it is of paramount importance to include all confounders in the outcome model and that excluding only one true confounder could lead to substantial bias in estimation of the multi pollutant adverse health effect. In section 2.2 we describe our method and present a simple illustrative example; in section 2.3 we present a simulation study that shows the advantage of

BAC-ME over traditional methods such as BMA and LASSO; in section 2.4 we apply our method to a retrospective epidemiological study aimed at estimating the multi pollutant adverse effect on cardiovascular (CVD) hospitalization associated with a simultaneous change in ozone and $PM_{2.5}$, controlling for weather data and population level characteristics; finally, in section 3.4 we conclude with a discussion.

2.2 Methods

2.2.1 Concept

Suppose we have multiple exposures, $X_1, \ldots, X_J$, a continuous outcome, Y, and M potential measured confounders (categorical and/or continuous), $\boldsymbol{C} = C_1, ...C_M$, and we want to estimate the effect of a simultaneous change in more than one exposure on the outcome. We will specifically examine the case with two exposures ($J = 2$) although the proposed framework can be easily generalized to more than two exposures. Our quantity of scientific interest is multi pollutant risk, defined here as the effect on cardiovascular outcome (Y) associated with the simultaneous change in exposure to two air pollutants (X_1 and X_2), adjusted for measured confounding. For J=2, we define the parameter of interest as:

$$\Delta_{\boldsymbol{x}}(\boldsymbol{\delta}) = \Delta_{(x_1,x_2)}(\delta_1, \delta_2) = E[Y|X_1 = x_1 + \delta_1, X_2 = x_2 + \delta_2] - E[Y|X_1 = x_1, X_2 = x_2] \qquad (2.1)$$

where δ_1 and δ_2 are simultaneous changes in exposure 1 and exposure 2, respectively and x_1 and x_2 are the current values of these exposures. For example, x_1 and x_2 could be the three year nationwide average level of $PM_{2.5}$ and ozone and δ_1 and δ_2 a 10 $\mu g/m^3$ increase in $PM_{2.5}$ and 10 ppm increase in ozone simultaneously.

2.2.2 Models

We specify three equations simultaneously: one for each exposure variable and one for the outcome.

$$X_{1i} = \eta_0^{X_1} + \sum_{m=1}^{M} \alpha_m^{X_1} \eta_m^{X_1} C_{mi} + \epsilon_i^{X_1} \tag{2.2}$$

$$X_{2i} = \eta_0^{X_2} + \gamma X_{1,i} + \sum_{m=1}^{M} \alpha_m^{X_2} \eta_m^{X_2} C_{mi} + \epsilon_i^{X_2} \tag{2.3}$$

$$Y_i = \eta_0^{Y} + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + \sum_{m=1}^{M} \alpha_m^{Y} \eta_m^{Y} C_{mi} + \epsilon_i^{Y} \tag{2.4}$$

where:

i = 1,...,N

$\epsilon^{X_1}, \epsilon^{X_2}, \epsilon^{Y} \overset{\text{iid}}{\sim} N(0, \sigma^2_{X_1}), N(0, \sigma^2_{X_2}), N(0, \sigma^2_{Y})$ respectively

The parameters $\boldsymbol{\alpha}^{X_1} \in \{0,1\}^M, \boldsymbol{\alpha}^{X_2} \in \{0,1\}^M, \boldsymbol{\alpha}^{Y} \in \{0,1\}^M$ are unknown vectors of indicator variables denoting whether or not a potential confounder is included in the regression model; $\alpha_m^{X_1} = 1$ if C_m is included in (2.2), $\alpha_m^{X_2} = 1$ if C_m is included in (2.3) and $\alpha_m^{Y} = 1$ if C_m is included in (2.4). In this setting, our scientific quantity of interest is $\Delta_{(x_1,x_2)}(\delta_1, \delta_2) = \delta_1\beta_1 + \delta_2\beta_2 + (\delta_1 x_2 + \delta_2 x_1 + \delta_1\delta_2)\beta_3$. Under the model formulation represented by (2.2) - (2.4), we assume that C_m is a true confounder if it is jointly associated with (X_1, X_2), and also associated with Y. This type of association can be manifested many different ways. The most obvious way is if C_m is associated with X_1 and Y and/or associated with X_2 and Y. However, these associations do not exhaust the possible confounding relationships in the multiple exposure setting. To conceptualize it is helpful to think of a binary covariate and exposures summarized in contingency tables, for example where the 2x2 contingency table for (X_1, C_m) is such that $p(C|X_1) = p(C)$ and likewise the contingency table for (X_2, C_m) is such that $p(C|X_2) = p(C)$ but the 4x2 contingency table for $(\{X_1, X_2\}, C_m)$ is such that $p(C|X_1, X_2) \neq p(C)$. For a specific example, see Appendix A.2.1. Such a situation is evident in the data analysis of Section 2.4 where

there exists covariates that are jointly associated with multiple exposures but not marginally associated with any single exposure. Throughout we assume that C_m is a pre-exposure variable; that is, we make the strong assumption that none of the C_m are intermediate variables that could be affected by any of the exposure variables.

We are interested in identifying the minimal outcome model; that is, the smallest outcome model that includes all true confounders and therefore will provide an unbiased estimate of $\Delta_{\boldsymbol{x}}(\boldsymbol{\delta})$. Adding C_m that are not true confounders into the health effects model will not bias estimation of $\Delta_{\boldsymbol{x}}(\boldsymbol{\delta})$. However, excluding even one of the true confounders yields a biased estimate. We will denote the minimal model as $\boldsymbol{\alpha}_0^Y$. Our goal is to estimate $\Delta_{\boldsymbol{x}}(\boldsymbol{\delta})$ when $\boldsymbol{\alpha}_0^Y$ is unknown. Any model, $\boldsymbol{\alpha}^Y$, that contains the minimal model, i.e. $\boldsymbol{\alpha}^Y \supseteq \boldsymbol{\alpha}_0^Y$, will yield a posterior distribution whose mean is an unbiased estimate of $\Delta_{\boldsymbol{x}}(\boldsymbol{\delta})$. The true model will always include the minimal model, $\boldsymbol{\alpha}_{TRUE}^Y \supseteq \boldsymbol{\alpha}_0^Y$, but may also include variables associated with only the outcome. Our method selects models that contain $\boldsymbol{\alpha}_0^Y$ by introducing prior dependence between $\boldsymbol{\alpha}^{X_1}, \boldsymbol{\alpha}^{X_2}$ and $\boldsymbol{\alpha}^Y$, ensuring that variables are selected based on joint associations with (X_1, X_2, Y).

2.2.3 Illustrative example

We will introduce our approach with an example to illustrate the danger of excluding even one of the true confounders when estimating $\Delta_{\boldsymbol{x}}(\boldsymbol{\delta})$. Consider the relationship in Figure 2.1. We have four true confounders (C_1, C_2, C_3, C_4) and one extraneous covariate C_5. The variable C_1 is strongly correlated with X_1, X_2 and Y; C_2 is strongly correlated with X_1 and Y; C_3 is strongly correlated with X_2 and weakly correlated with Y; C_4 is weakly correlated with X_2 and strongly correlated with Y and C_5 is uncorrelated with both exposures and the outcome. In addition, the two exposures are moderately correlated with each other. The minimal model guaranteed to provide an unbiased estimate of $\Delta_{\boldsymbol{x}}(\boldsymbol{\delta})$ is $\boldsymbol{\alpha}_0^Y = (1, 1, 1, 1, 0)$. This is also the true model. The full model, $\boldsymbol{\alpha}^Y = (1, 1, 1, 1, 1)$, includes $\boldsymbol{\alpha}_0^Y$, is fully adjusted for confounding and will yield an unbiased estimate as well. Any model that does not include $\boldsymbol{\alpha}_0^Y$ is not guaranteed to yield a

posterior mean that is an unbiased estimate $\Delta_{\boldsymbol{x}}(\boldsymbol{\delta})$.

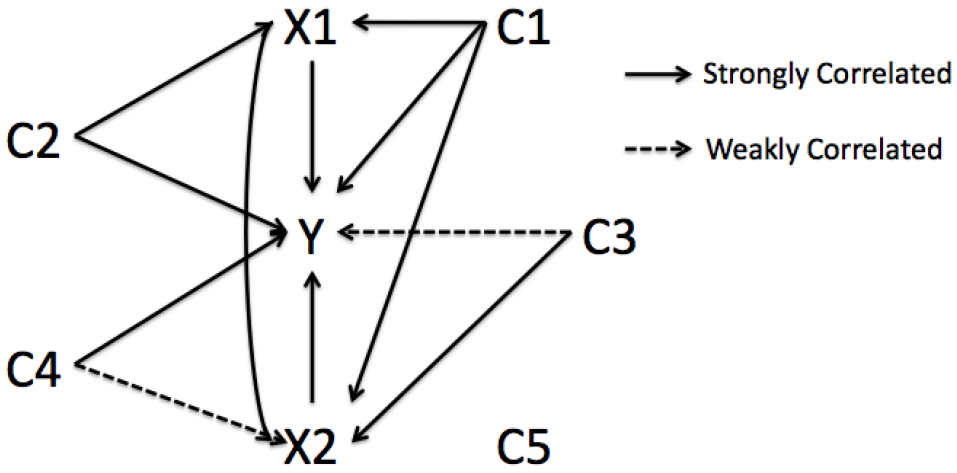


Figure 2.1: Illustrative Example

To illustrate we generated 1000 data sets from the following models which reflect the situation depicted in Figure 2.1:

$$X_{1i} = C_{1i} + C_{2i} + \epsilon_i^{X_1}$$
$$X_{2i} = 0.3X_{1i} + C_{1i} + C_{3i} + 0.1C_{4i} + \epsilon_i^{X_2}$$
$$Y_i = X_{1i} + X_{2i} - 0.5X_{1i}X_{2i} + C_{1i} + C_{2i} + 0.1C_{3i} + C_{4i} + \epsilon_i^{Y}$$

Throughout the remaining simulations, unless otherwise noted, i = 1,...,1000, $\epsilon^{X_1}, \epsilon^{X_2}, \epsilon^{Y} \overset{\text{iid}}{\sim} N(0,1)$, $C_{m,i} \overset{\text{iid}}{\sim} N(0,1)$

We estimated $\Delta_{0,0}(1,1) = \beta_1 + \beta_2 + \beta_3$ from the ordinary least squares estimates of $\boldsymbol{\beta}$ under different confounding adjustments. Table 2.1 column 2 shows the average bias in $\hat{\Delta}_{0,0}(1,1)$. We see that excluding C_1 which is strongly associated with both exposures and the outcome, introduces the most bias. Even the estimate from the model without C_3, which is only weakly correlated with the outcome, is biased. In fact, only $\boldsymbol{\alpha}_0^Y = \boldsymbol{\alpha}_{TRUE}^Y = (1,1,1,1,0)$ and $\boldsymbol{\alpha}^Y = (1,1,1,1,1)$ yield posterior mean estimates which are unbiased for $\Delta_{\boldsymbol{x}}(\boldsymbol{\delta})$.

Table 2.1: Bias, MSE and CI coverage in $\hat{\Delta}_{0,0}(1,1)$by model. The bias is the average bias arising from the ordinary least squares fit across simulations.

Model	Bias	MSE	CI Coverage
(1,1,1,1,0; True Model)	0.0003	0.0016	0.9520
(1,1,1,1,1)	0.0003	0.0016	0.9540
(0,1,1,1,0)	0.5661	0.3213	0.0000
(1,0,1,1,0)	0.4989	0.2505	0.0000
(1,1,0,1,0)	0.0352	0.0026	0.8440
(1,1,1,0,0)	0.0715	0.0082	0.7500

2.2.4 Prior specification for Bayesian model-averaged estimates.

When the goal is effect estimation accounting for model uncertainty, it is common to calculate the posterior distribution of the effect by taking a weighted average over models (*Hoeting et al.*, 1999, *Raftery*, 1995):

$$\sum_{\boldsymbol{\alpha}^Y} P(\Delta_{\boldsymbol{x}}^{\boldsymbol{\alpha}^Y}(\boldsymbol{\delta})|\boldsymbol{\alpha}^Y, D)P(\boldsymbol{\alpha}^Y|D) \tag{2.5}$$

Where $\Delta_{\boldsymbol{x}}^{\boldsymbol{\alpha}^Y}(\boldsymbol{\delta})$ is the model specific effect from the model $\boldsymbol{\alpha}^Y$, $P(\boldsymbol{\alpha}^Y|D)$ is the posterior probability of (or weight assigned to) $\boldsymbol{\alpha}^Y$, and D=($\boldsymbol{X}, Y, \boldsymbol{C}$). Equation (2.5) can be decomposed into two parts: the sum over models which include α_0^Y and the sum over the remaining models. i.e.,

$$\sum_{\boldsymbol{\alpha}^Y \supseteq \boldsymbol{\alpha}_0^Y} P(\Delta_{\boldsymbol{x}}^{\boldsymbol{\alpha}^Y}(\boldsymbol{\delta})|\boldsymbol{\alpha}^Y, D)P(\boldsymbol{\alpha}^Y|D) + \sum_{\boldsymbol{\alpha}^Y \not\supseteq \boldsymbol{\alpha}_0^Y} P(\Delta_{\boldsymbol{x}}^{\boldsymbol{\alpha}^Y}(\boldsymbol{\delta})|\boldsymbol{\alpha}^Y, D)P(\boldsymbol{\alpha}^Y|D) \tag{2.6}$$

When $\boldsymbol{\alpha}^Y$ contains $\boldsymbol{\alpha}_0^Y$, the posterior mean of $P(\Delta_{\boldsymbol{x}}^{\boldsymbol{\alpha}^Y}(\boldsymbol{\delta})|\boldsymbol{\alpha}^Y, D)$ is an unbiased estimate of $\Delta_{\boldsymbol{x}}(\boldsymbol{\delta})$. If we have a method that assigns posterior weights only to the models in the first term of (2.6), we are averaging across models that yield unbiased estimates of $\Delta_{\boldsymbol{x}}(\boldsymbol{\delta})$. By contrast, any method that assigns high weights to the models in the second term of (2.6) is averaging across models that are unlikely to yield an unbiased estimate of $\Delta_{\boldsymbol{x}}(\boldsymbol{\delta})$. Note that this quantity is defined based solely on quantities in (2.4) but (2.2) and (2.3) inform estimation of $\boldsymbol{\alpha}^Y$ in (2.4). Our goal is to specify a prior distribution on $\boldsymbol{\alpha}^Y|\boldsymbol{\alpha}^{X_1}, \boldsymbol{\alpha}^{X_2}$ that assigns the posterior mass mostly to models that contain

$\boldsymbol{\alpha}_0^Y$. This will ensure averaging across unbiased estimates of $\Delta_{\boldsymbol{x}}(\boldsymbol{\delta})$.

BAC-ME jointly considers both exposure models, (2.2) and (2.3), and the outcome model, (2.4). To assign more posterior mass to the health effects models that contain $\boldsymbol{\alpha}_0^Y$, we assign prior probabilities that ensure variables related to either exposure variable are included in the outcome model. Specifically, first we specify a prior distribution on $\boldsymbol{\alpha}^Y|\boldsymbol{\alpha}^{X_1}, \boldsymbol{\alpha}^{X_2}$ by defining a dependence parameter, ω, that represents the prior odds of including a covariate in the outcome model when it is in either (or both) exposure models. These priors can be extremely general – for instance different dependence parameters for each exposure and even each confounder (see Appendix A.2.2 for a more general formulation) – but for simplicity we will investigate the case where:

$$\frac{P(\alpha_m^Y = 1|\alpha_m^{X_1} = 1, \alpha_m^{X_2} = 1)}{P(\alpha_m^Y = 0|\alpha_m^{X_1} = 1, \alpha_m^{X_2} = 1)} = \frac{P(\alpha_m^Y = 1|\alpha_m^{X_1} = 1, \alpha_m^{X_2} = 0)}{P(\alpha_m^Y = 0|\alpha_m^{X_1} = 1, \alpha_m^{X_2} = 0)} = \frac{P(\alpha_m^Y = 1|\alpha_m^{X_1} = 0, \alpha_m^{X_2} = 1)}{P(\alpha_m^Y = 0|\alpha_m^{X_1} = 0, \alpha_m^{X_2} = 1)} = \omega \tag{2.7}$$

$$\frac{P(\alpha_m^Y = 1|\alpha_m^{X_1} = 0, \alpha_m^{X_2} = 0)}{P(\alpha_m^Y = 0|\alpha_m^{X_1} = 0, \alpha_m^{X_2} = 0)} = 1 \tag{2.8}$$

In the case of a single exposure, *Wang et al.* (2012) note that $\omega = \infty$ is usually conservative and provides unbiased results. Setting $\omega = \infty$ makes the default formulation that if C_m is associated with (X_1, X_2) then C_m is forced into the outcome model. If C_m is also associated with Y, then C_m is a confounder and must be in the outcome model to guarantee an unbiased estimate of $\Delta_{\boldsymbol{x}}(\boldsymbol{\delta})$. If C_m is not associated with Y, we may loose efficiency in our estimation but we still obtain an unbiased estimate of $\Delta_{\boldsymbol{x}}(\boldsymbol{\delta})$. While this method is flexible enough to alter ω to any value, setting $\omega = \infty$ maximizes our chance of selecting any variable associated with exposure(s) into the outcome model. For all results presented here, we set $\omega = \infty$. See Appendix A.2.2 for the prior distribution of $\boldsymbol{\alpha}^{X_1}|\boldsymbol{\alpha}^Y$, $\boldsymbol{\alpha}^{X_2}|\boldsymbol{\alpha}^Y$ and the joint, marginal and conditional probabilities implied by the odds ratios in (2.7) and (2.8).

Our goal is to estimate the posterior distribution of $(\boldsymbol{\alpha}^Y, \boldsymbol{\alpha}^{X_1}, \boldsymbol{\alpha}^{X_2}, \Delta_{(x_1,x_2)}(\delta_1, \delta_2))$. We assume the following priors for other model parameters:

$$(\gamma, \boldsymbol{\eta}^{X_j})|(\boldsymbol{\alpha}^{X_j}, \sigma^2_{X_j}) \sim N(\boldsymbol{\mu}_{0\boldsymbol{\alpha}^{X_j}}, \sigma^2_{X_j}\phi^2\boldsymbol{\Sigma}_{0\boldsymbol{\alpha}^{X_j}}), j = 1, 2$$
$$(\beta_1, \beta_2, \beta_3, \boldsymbol{\eta}^Y)|(\boldsymbol{\alpha}^Y, \sigma^2_Y) \sim N(\boldsymbol{\mu}_{0\alpha^Y}, \sigma^2_Y\phi^2\Sigma_{0\alpha^Y})$$
$$\frac{1}{\sigma^2_{X_1}}, \frac{1}{\sigma^2_{X_2}}, \frac{1}{\sigma^2_Y} \sim Gamma(\nu/2, \nu\lambda/2)$$

where $(\nu\lambda/2)$ is the rate parameter of the Gamma distribution (i.e. $E[\frac{1}{\sigma^2_{X_1}}] = 1/\lambda$) and $\nu, \lambda, \phi, \boldsymbol{\mu}_{0\boldsymbol{\alpha}^{X_j}}, \boldsymbol{\mu}_{0\boldsymbol{\alpha}^Y}, \boldsymbol{\Sigma}_{0\boldsymbol{\alpha}^{X_j}}$, and $\boldsymbol{\Sigma}_{0\boldsymbol{\alpha}^Y}$ are hyperparameters that are specified as recommended by *Raftery et al.* (1997)

We used an MCMC algorithm to draw posterior samples of $(\boldsymbol{\alpha}^{X_1}, \boldsymbol{\alpha}^{X_2}, \boldsymbol{\alpha}^Y, \gamma, \boldsymbol{\beta})$. We used the MC^3 method (*Madigan et al.*, 1995) to sample from the first three full conditionals. Derivation of the posterior distributions for all parameters may be found in Appendix A.2.3.

2.3 Simulations Studies

In section 2.3.1, in the simple setting of three true confounders, ten variables associated with outcome only and 30 extraneous covariates, we will show the reduction in bias in estimation of $\Delta_{\boldsymbol{x}}(\boldsymbol{\delta})$ that BAC-ME provides over methods that select variables based solely on their ability to predict Y; in section 2.3.2 we will simulate data sets from a more complex scenario: 20 confounders, 10 variables associated with outcome only and 30 extraneous variables.

Table 2.2: Comparative Methods

Method	Description
BAC-ME	Bayesian Adjustment for Confounding - Multiple Exposures
FBMA	Forced Bayesian Model Averaging - Exposures are forced into the model
NLASSO	Not-forced Adaptive Least Absolute Shrinkage and Selection Operator - Exposures are allowed to enter or leave the model just as confounders

With a lack of methods designed specifically for confounding adjustment in the presence of multiple exposures, much less that account for uncertainty in that confounding adjustment, we chose methods that have traditionally been used for model selection. LASSO and adaptive LASSO are common methods for simultaneous estimation and variable selection (*Zou*, 2006). BMA is a standard Bayesian method for model selection. BMA also goes a step further and has been proposed as a method to account for uncertainty in confounding adjustment (*Hoeting et al.*, 1999, *Raftery*, 1995). The specific methods we explored are detailed in Table 2.2. Both BMA and adaptive LASSO perform variable selection on only the health effects model (equation (2.4)). These methods all choose a model based on its ability to predict Y and not on its ability to estimate the multi pollutant adverse health effect of a change in X_1 and X_2 on Y properly adjusted for confounding. Note that the last two columns of Table 2.3 list model weights for FBMA and NLASSO for the datasets in section 2.2.3. Both FBMA and NLASSO assign the most weight to model (1,1,0,1,0), which does not include $\boldsymbol{\alpha}_0^Y$. FBMA only selected a model which contained $\boldsymbol{\alpha}_0^Y$ 30.3% of the time and NLASSO only 0.1% of the time. These methods tended to select outcome models that do not include C_3 because this variable is only weakly associated with Y but strongly associated with an exposure. In contrast, BAC-ME, which simultaneously fits both exposure models and the outcome model, always selected a model containing $\boldsymbol{\alpha}_0^Y$.

Table 2.3: Bias by model and proportion of time that model is selected by method. The bias is the average bias arising from the ordinary least squares fit across simulations.The weight for BAC-ME and FBMA is the posterior probability of $\boldsymbol{\alpha}^Y$. The weight for NLASSO is the proportion of time that that model was selected. FBMA used a uniform prior on $\boldsymbol{\alpha}^Y$. BAC-ME used the priors defined by (2.7) - (2.8) with $\omega = \infty$.

Model	Bias($\hat{\Delta}_{0,0}(1,1)$)	BAC-ME weight	FBMA weight	NLASSO weight
(1,1,1,1,0; True Model)	0.0003	0.975	0.296	0.001
(1,1,1,1,1)	0.0003	0.025	0.007	0.000
(0,1,1,1,0)	0.5661	0.000	0.000	0.000
(1,0,1,1,0)	0.4989	0.000	0.000	0.000
(1,1,0,1,0)	0.0352	0.000	0.681	0.999
(1,1,1,0,0)	0.0715	0.000	0.000	0.000
Includes $\boldsymbol{\alpha}_0^Y$		1	0.303	0.001

2.3.1 Bias by degree of confounding

This set of simulations will explore bias as a function of the degree of confounding across methods. This will demonstrate the settings in which BAC-ME has the largest advantage over BMA and adaptive LASSO. For these simulations we have one confounder associated with both exposures (C_1), one confounder of X_1 only (C_2), one confounder of X_2 only (C_3), ten variables associated with outcome (Y) only and 30 extraneous covariates – variables not associated with either of the exposures nor the outcome – as noted below.

$$
\begin{aligned}
E[X_1] &= C_1 + C_2 \\
E[X_2] &= 0.3X_1 + C_1 + \eta_3^{X_2} C_3 \\
E[Y] &= 0.2X_1 + 0.2X_2 + 0.1X_1X_2 + C_1 + C_2 + \eta_3^{Y} C_3 + \sum_{i=4}^{13} C_i
\end{aligned}
$$

We vary the true coefficients, $\eta_3^{X_2}$ and η_3^{Y}, on a grid from 0.1 to 1 to assess the bias we see in different methods under different strengths of confounding for the relationship between X_2 and Y. For each combination of $\eta_3^{X_2}$ and η_3^{Y} we generated 100 data sets with a sample size of 500 each. The heat maps below demonstrate that BAC-ME provides unbiased estimates for a much larger range of true coefficients. Figure 2.2 shows the bias of BAC-ME and FBMA. Darker colors indicate more bias. If both coefficients are relatively small, little bias results from excluding them in the outcome model. If η_3^{Y} is large, both methods will tend to select C_3 into the outcome model and will yield an unbiased estimate. However, when $\eta_3^{X_2}$ is large and η_3^{Y} is small, FBMA yields biased estimates.

A figure of the comparative bias of BAC-ME and NLASSO would show that regardless of the values of $\eta_3^{X_2}$ and η_3^{Y}, NLASSO is more biased than BAC-ME. Comparisons as a function of CI coverage rather than bias are nearly identical.

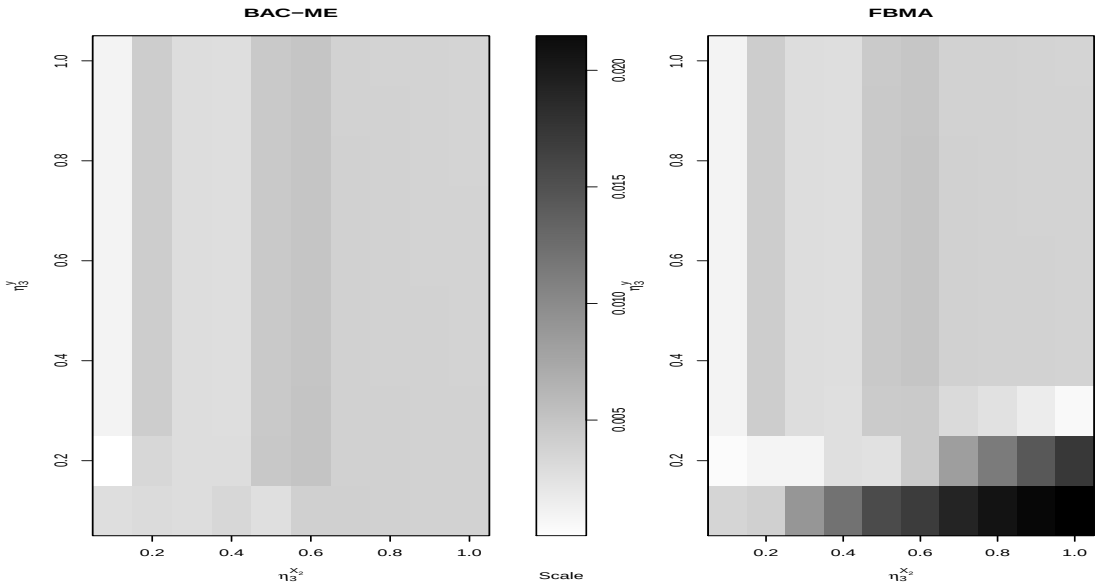


Figure 2.2: Heat maps of bias by strength of confounding. Darker shades indicate more bias.

2.3.2 More complex simulations

For the final set of simulations we will present a more complex scenario with 20 confounders – $(C_1, ..., C_5)$ associated with both X_1 and X_2, $(C_6, ..., C_{10})$ associated with X_1 only and $(C_{11}, ..., C_{20})$ associated with X_2 only – 10 variables associated with outcome only and 30 extraneous variables. We will show that BAC-ME results in significantly less bias than FBMA or NLASSO. We generated data from the following models:

$$E[X_1] = \sum_{i=1}^{5} C_i + \sum_{i=6}^{10} C_i$$

$$E[X_2] = 0.3X_1 + \sum_{i=1}^{5} C_i + \sum_{i=11}^{15} C_i + \sum_{i=16}^{20} 0.1C_i$$

$$E[Y] = 0.2X_1 + 0.2X_2 + 0.1X_1X_2 + \sum_{i=1}^{5} C_i + \sum_{i=6}^{10} C_i + \sum_{i=11}^{15} 0.1C_i + \sum_{i=16}^{20} C_i + \sum_{i=21}^{30} C_i$$

Table 2.4 summarizes results from these simulations with respect to $\Delta_{0,0}(1, 1)$. We see that BAC-ME outperforms FBMA and NLASSO with respect to bias, MSE and CI coverage with FBMA and

NLASSO having more than ten times the bias of BAC-ME. We can see the reason from columns two and three. The second column is the percent of time the method chose a model that included the minimal model; the third column shows the percent of time the method chose the true model. Only our method picks models which include the minimal model 100% of the time. FBMA never selects models which include the minimal model and NLASSO does so only 2% of the time. This is not surprising. Exposure 2 (X_2) had confounders that were only weakly associated with the outcome but strongly associated with the exposure. FBMA and NLASSO have no mechanism by which to identify these confounders, they are typically not selected into the model, and bias results.

Table 2.4: Results for $\hat{\Delta}_{0,0}(1,1)$ for the simulation in Section 2.3.2. *Incl. Min* is the proportion of time the method selected a model that contained the minimal model. *True* is the proportion of time the method selected the true model. *Bias* is the difference in the true value of $\Delta_{0,0}(1,1)$ and $\hat{\Delta}_{0,0}(1,1)$ where $\hat{\Delta}_{0,0}(1,1)$ is the average posterior mean for BAC-ME and FBMA and the average estimate for NLASSO.

Method	Incl. Min	True	Bias	SE	MSE	CI Coverage
BAC-ME	1.00	0.43	0.0034	0.0351	0.0012	0.9800
FBMA	0.00	0.00	0.0574	0.0306	0.0042	0.6400
NLASSO	0.02	0.02	-0.0615	0.0566	0.0070	0.9100

2.4 Data Analysis

In this section, we apply BAC-ME to a retrospective epidemiological study of over 14 million medicare enrollees, weather, pollution and demographic data. The data includes county level characteristics for 413 counties throughout the US for the period 2008-2010. These include rate of cardiovascular (CVD) hospital admissions, county level traits, and temperature and dew point averages and standard deviations, for a total of 57 potential confounders. Specific details on the data set may be found in Appendix A.2.4. The goal is to estimate the change in the rate of emergency hospitalizations for CVD associated with a simultaneous increase of one interquartile range in both ozone and $PM_{2.5}$ while accounting for weather, demographics (age, race and gender) and population level characteristics (e.g. proportion who are overweight) from the U.S. Census and the CDC's Behavioral Risk Factor Surveillance System. The hospitalization rate is

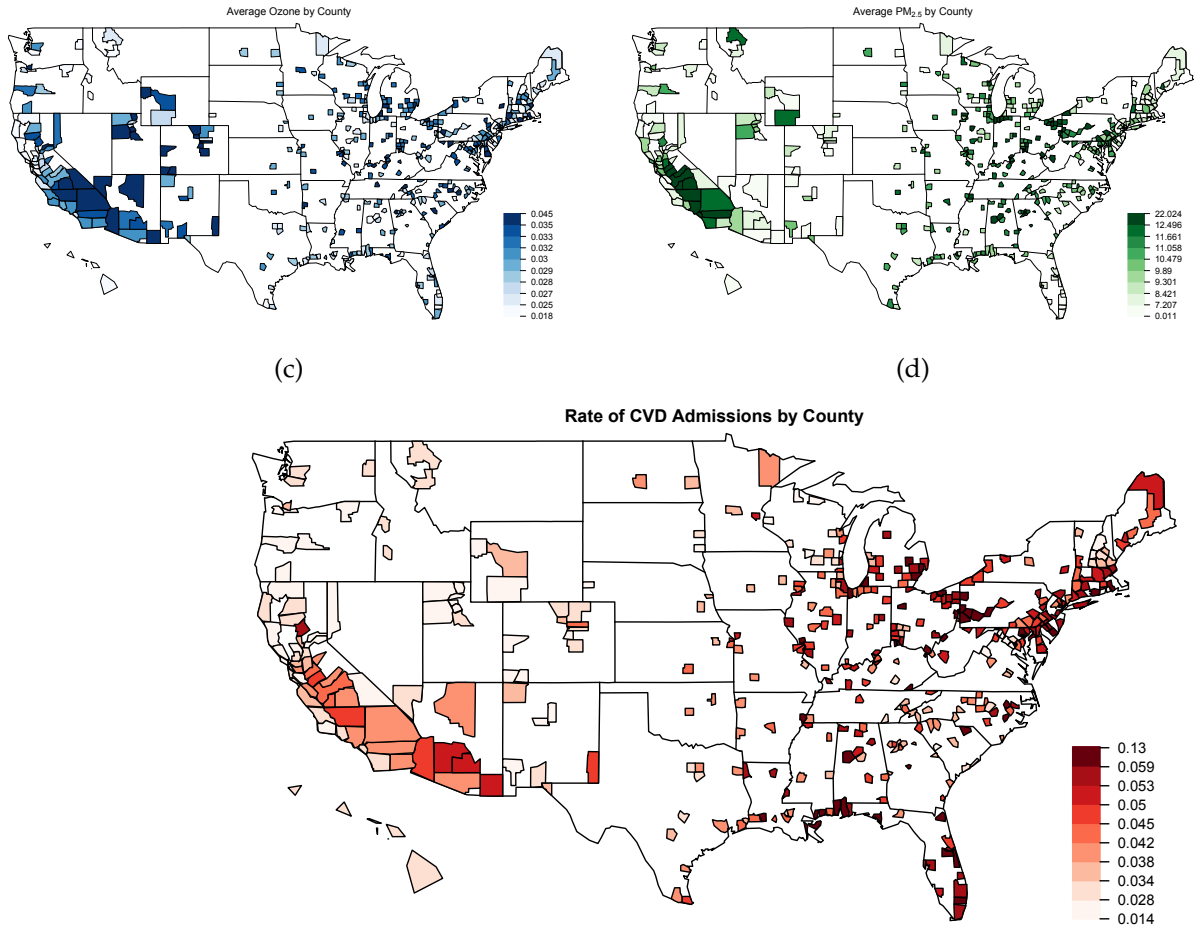
recorded separately for each group (gender and race).

We conducted all analyses described both with and without an interaction term between ozone and $PM_{2.5}$ in the outcome model. No analysis provided evidence of a nonzero interaction between exposures, so we present results for the models without the interaction. To start we considered a full model with all available covariates.

$$
\begin{aligned}
O_{3i} &= \eta_0^{X_1} + \sum_{m=1}^{M} \alpha_m^{X_1} \eta_m^{X_1} C_{mi} + \epsilon_i^{X_1} \\
PM_{2.5i} &= \eta_0^{X_2} + \gamma O_{3i} + + \sum_{m=1}^{M} \alpha_m^{X_2} \eta_m^{X_2} C_{mi} + \epsilon_i^{X_2} \\
Y_i &= \eta_0^{Y} + \beta_1 O_{3i} + \beta_2 PM_{2.5i} + \sum_{m=1}^{M} \alpha_m^{Y} \eta_m^{Y} C_{mi} + \epsilon_i^{Y}
\end{aligned} \tag{2.9}
$$

$Y_i = \dfrac{CVD}{N_i}$ where CVD is total number of cardiovascular admissions, defined as Heart Failure, Heart Rhythm Disturbances, Ischemic Heart Disease or Peripheral Vascular Disease and N_i is the total person-years at risk in county i from 2008-2010. Based on preliminary analysis, the use of a linear model for the outcome Y_i is reasonable. Figures 2.3c and 2.3d show the average levels of ozone and $PM_{2.5}$ by county for the 413 counties in our data; Figure 2.3e shows the rate of CVD admissions for the same 413 counties. The vector $\boldsymbol{C}_i$ denotes the set of potential confounders described above and given in detail in Appendix A.2.4. We eliminated 10 potential confounders due to missing data or high correlation (> 0.8) with other confounders; the eliminated confounders are listed in Appendix A.2.4, Table A.2.

To control for weather, we include seventh degree polynomial terms for temperature and dew point. For these polynomial terms, a term could not exit the model unless there were no higher order terms currently in the model. Likewise, a term could not enter the model unless all lower order terms were already in the model. We assume the residuals are independent and identically distributed $N(0, \sigma^2)$ random variables. We considered five approaches, BAC-ME, BMA forcing



(c) (d)

(e)

Figure 2.3: (a) shows the average ozone levels by county (ppm) (b) the average $PM_{2.5}$ levels ($\mu g/m^3$) (c) the rate of CVD admissions (admissions per person-year). Levels shown for Hawaii are for Honolulu county.

exposures into the model, adaptive LASSO not forcing exposures into the model, ordinary least squares (OLS) and BAC (single exposure). OLS is the least squares estimate from equation (2.9) with $\alpha_m^Y = 1$ for $m = 1, \ldots, M$. For BAC (single exposure) we fit two models to the data. The first is a model with ozone as the exposure and treating $PM_{2.5}$ as one of the potential measured confounders. This model is defined by (2.10) - (2.11) and with prior odds ratios given in (2.12) with $\omega = \infty$.

$$O_{3i} = \sum_{m=1}^{M} \alpha_m^X \eta_m^X C_{mi} + \alpha_{M+1}^X \eta_{M+1}^X PM_{2.5i} + \epsilon_i^X \tag{2.10}$$

$$Y_i = \beta O_{3i} + \sum_{m=1}^{M} \alpha_m^Y \eta_m^Y C_{mi} + \alpha_{M+1}^Y \eta_{M+1}^Y PM_{2.5i} + \epsilon_i^Y \tag{2.11}$$

$$\frac{P(\alpha_m^Y = 1|\alpha_m^X = 1)}{P(\alpha_m^Y = 0|\alpha_m^X = 1)} = \omega, \qquad \frac{P(\alpha_m^Y = 1|\alpha_m^X = 0)}{P(\alpha_m^Y = 0|\alpha_m^X = 0)} = 1 \tag{2.12}$$

Similarly, the second BAC (single exposure) model treats $PM_{2.5}$ as the exposure and ozone as a potential measured confounder. As in the simulation studies, results from NLASSO were vastly different than from the other approaches and we will only show results for the other four methods here.

Table 2.5 shows the estimated regression coefficients for one inter-quartile range (IQR) increase in both ozone and $PM_{2.5}$ per 10,000 person-years at risk. With BAC-ME we estimate coefficients for each pollution variable that are not statistically significant individually. More specifically, we found that a simultaneous change in ozone from its 25^{th} to 75^{th} percentile and $PM_{2.5}$ from its 25^{th} to 75^{th} percentile is associated with an increase in CVD hospital emergency admissions of 25.6 per 10,000 person years at risk. Notably, the 95% credible interval for $\hat{\Delta}_{(x_1,x_2)}(\delta_1, \delta_2)$ does not contain zero. FBMA provides point estimates for β_1 and β_2 that are substantially different than those

provided by BAC-ME and an estimate for β_1 that is statistically significant. For this particular change, $\hat{\Delta}_{(x_1,x_2)}(\delta_1, \delta_2)$ is very similar for both methods. However, this would not necessarily be the case if we investigated a different change (i.e. other than one IQR for both pollutants). OLS provides point estimates that are smaller than the estimates provided by BAC-ME and an estimate of the multi pollutant effect that is not statistically significant.

Table 2.5: Effect estimates for one IQR increase in ozone (β_1) and $PM_{2.5}$ (β_2) per 10,000 person-years at risk. Δ is the expected change in the rate of CVD admissions per 10,000 person years at risk for a change in both pollutants from their 25^{th} percentiles to their 75^{th} percentiles.

Method	Parameter	Estimate	SE	95% Interval
BAC-ME	β_1 (Ozone)	13.2	7.3	(-1.0, 27.4)
	β_2 ($PM_{2.5}$)	12.5	8.5	(-4.2, 29.0)
	Δ	25.6	11.2	(3.5, 47.4)
BAC (Ozone)	β	10.4	7.4	(-3.9, 24.8)
BAC ($PM_{2.5}$)	β	14.5	8.4	(-1.8, 30.7)
FBMA	β_1	17.0	7.0	(2.9, 30.7)
	β_2	8.6	7.9	(-7.0, 23.9)
	Δ	25.6	10.7	(4.8, 46.2)
OLS	β_1	9.4	7.7	(-5.7, 24.4)
	β_2	8.5	8.5	(-8.1, 25.2)
	Δ	17.9	11.4	(-4.4, 40.3)

This data analysis highlights the importance of conducting multi pollutant analysis differently than single pollutant analysis. Consider the BAC (single exposure) models. The estimates for the individual effects of a change in ozone or $PM_{2.5}$ vary somewhat between BAC (single exposure) models and that defined by (2.9). Figure 2.4 shows the posterior inclusion probability ($P(\boldsymbol{\alpha}^Y|D)$) of each of the 47 potential confounders from the multiple exposure model compared to the the two single exposure models. Notice that the posterior probabilities of $\boldsymbol{\alpha}^Y$ differ greatly across models for some covariates and inclusion in model (2.9) is not simply the union of those included in the two single exposure models. Under the single exposure models we are estimating different parameters than under the multiple exposure model. In the single exposure models, e.g. (2.11), we are searching for the covariates that are associated with Y and ozone (or $PM_{2.5}$) but never for covariates that are associated with Y and with (ozone, $PM_{2.5}$) jointly. For instance, consider that the posterior probability of including mean age in outcome model (2.9) is 0.46 whereas the posterior probabilities of including it in either of the single exposure models are 0. This difference

is because mean age is not marginally associated with either ozone or $PM_{2.5}$ but it is associated with them jointly. To better illustrate the differences it is helpful to think of the exposures as binary variables. The variable mean age is not associated with ozone marginally. That is, counties with high ozone have a similar age distribution to that of counties with low ozone. Similarly, mean age is not associated with $PM_{2.5}$ marginally. However, counties with both high ozone and low $PM_{2.5}$ tend to have a younger medicare population than the rest of the country. That is, mean age is jointly associated with (ozone, $PM_{2.5}$) even though it is marginally associated with neither. The example in Appendix A.2.1 illustrates a similar situation. But just as important, with the BAC (single exposure) models, there is no clear way to obtain an estimate of the posterior distribution of the multi pollutant effect, and hence no way to capture the uncertainty surrounding any estimate of this effect.

2.5 Discussion

A formal method to estimate a multi pollutant adverse health effect fully adjusted for confounding is currently lacking. BAC-ME gives a means to identify true confounders in a multiple exposure setting while guarding against the possibility of ignoring variables only weakly associated with outcome but strongly associated with one or more exposure variables. Importantly, BAC-ME is designed to detect true confounders on the basis of joint association with multiple exposures, rather than restrict attention to the subset of confounders that are marginally associated with at least one exposure. Further, BAC-ME is designed to acknowledge the uncertainty in the confounder selection, an issue that is exacerbated when there are multiple exposures and when the vector of available covariates in high-dimensional. Our simulation studies show that in a variety of settings our method outperforms methods that include potential confounders into the health effects model based solely on their ability to predict the outcome and ignoring completely their association with exposures.

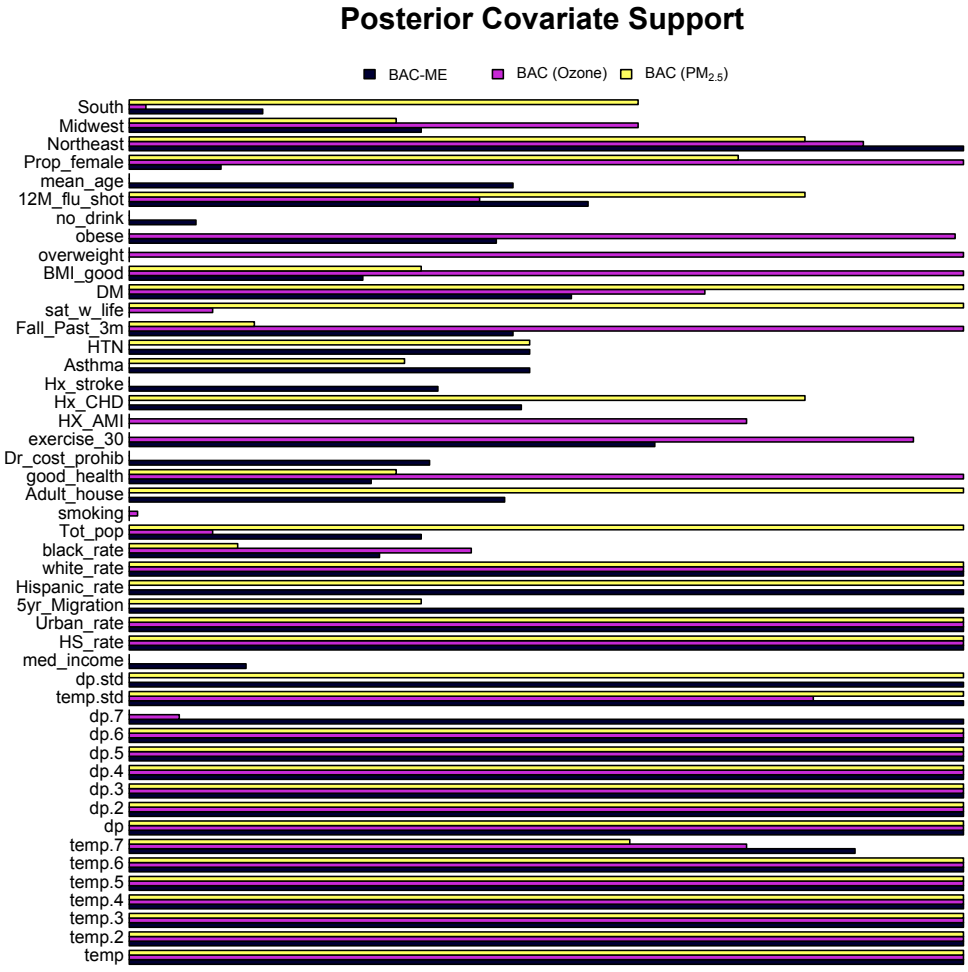


Figure 2.4: Posterior probabilities of including each of the potential confounders in the multiple exposure health effects model (Equation (2.9)) and in the single exposure models (e.g. Equation (2.11))

.

One might be tempted to use existing methods to estimate effects of ozone and $PM_{2.5}$ separately but this can only approximate the targeted effect of a simultaneous exposure. An approach that treats the exposures separately would ignore the potential interaction between the two components. Only in the most simple case is the multi pollutant adverse health effect simply the sum of the individual effects. Additionally, this approach does not control for confounding properly. Even absent an interaction term, the coefficients from the models $E[Y] = \beta_1 X_1 + \boldsymbol{\eta_1 C}$ and $E[Y] = \beta_2 X_2 + \boldsymbol{\eta_2 C}$ are not generally the same as those from the model $E[Y] = \beta_1 X_1 + \beta_2 X_2 + \boldsymbol{\eta C}$ and cannot be interpreted as the adverse health effect of a simultaneous change in both exposures.

For all simulations presented, we also compared the performance of BMA when exposures were not forced into the model, adaptive LASSO when they were forced into the model and a two stage LASSO / OLS procedure. We found that BAC-ME outperformed all other methods examined. We ran additional simulations varying the covariance structure and sample and effect sizes; results were consistent with those presented here.

Identifying the set of true confounders may be a goal in and of itself. In large studies, collecting data on unnecessary confounders wastes time and resources at best and could even be medically invasive. By identifying the set of true confounders, future studies may be designed more efficiently: data need not be collected on unnecessary covariates and extra care can be taken to collect information for a known confounder.

This method may be easily extended to accommodate any order interaction between any terms, including exposure – confounder interactions. In practice, one must carefully consider which terms should be included as 'exposures' (and, hence, forced into the model) and which should be treated as 'confounders' (and be allowed to enter and leave the model.) This method may also be easily extended to include any number of exposures, although the number it is wise to include may be limited by the situation and available data. Additionally, the method may be extended

to include more complex model formulations such as GLMs, though this would require more advanced computational techniques.

Propensity Score Methods for Combining Data Sources

Abstract

Comparative effectiveness research increasingly relies on observational studies based on large administrative databases, analysis of which raises several methodological challenges. Additionally, it is often desirable to combine heterogenous sources of information to estimate effects in an overall population while making using of confounders available only for a small subset of the population. When these additional, partially measured confounders are high dimensional, correlated or contain both continuous and categorical variables, traditional approaches such as Bayesian data augmentation are very challenging. We propose two methods that build on work by *McCandless et al.* (2012). Our methods use 'conditional propensity scores' to reduce the partially measured covariates to a scalar quantity, which may then be imputed in the main data. We conduct a simulation study which shows that in a variety of settings our methods reduce bias over the more common approaches of adjusting for only the fully measured covariates or complete case analysis.

3.1 Introduction

Comparative effectiveness research (CER) increasingly relies on observational studies based on large administrative databases. Compared to clinical trials, the analyses of these databases allow us to study a much larger population and investigate additional questions of interest. However, analysis of these large and complex administrative databases raises several methodological challenges and requires the development of new statistical methods. Comparing the effectiveness of treatment strategies in observational data is challenging both because patients are not randomly assigned to treatment strategies and because medical providers are not randomly assigned to quality of care interventions, which introduces the likely possibility that outcome comparisons are confounded by factors that simultaneously relate to treatment choices, providers, and health outcomes.

In addition, it is often desirable to combine heterogeneous sources of information, specifically information coming from a primary data source (e.g. Medicare) that provides information for treatments, outcomes, and a limited set of measured confounders on a large number of people and smaller supplementary data sources (e.g. SEER-Medicare) containing a much richer set of covariates. These additional covariates may be high dimensional, frequently exceeding the number of covariates available in the primary data. Additionally, these covariates often contain important confounders not measured in the primary data.

For example, glioblastoma multiforme (GBM), the most prevalent of the primary brain tumors, is a devastating disease with high mortality and high medical costs. Our goal is to estimate the average causal effect in the elderly of a treatment, e.g. major craniotomy, on an outcome of interest, e.g. 1-year mortality, adjusting for both fully and partially observed covariates. Elderly GBM patients (65 or older) are characterized by a high rate of associated comorbidities and are often excluded from clinical trials. GBM is a relatively rare disease and many questions regarding patient level outcomes can only be addressed with the linkage and analysis of very large administrative databases. Because a diagnosis of GBM is most commonly made during a hospitalization, Medicare Part A inpatient claims data captures almost entirely the population of elderly GBM patients. Medicare Part A contains patient demographic data (age, sex, etc.) and comorbidity information. The SEER-Medicare (Surveillance, Epidemiology and End Results) linked database can be used to identify the Medicare enrollees with GBM that are in SEER. SEER-Medicare provides a wealth of information about important confounders such as cancer site, stage, and histology for approximately 9.5% of the Part A study population. However, SEER-Medicare is not a representative sample of the Medicare population; it is taken from 20 regional registries representing only a small part of the Medicare population. Combining the two data sources would allow us estimate effects in the entire Medicare population with GBM while making use of the important confounders available only in SEER.

There are currently a number of challenges to combining such data sources for analysis. Traditional Bayesian data augmentation methods require specifying the joint distribution of outcome, treatment and covariates (*Little and Rubin*, 2002). Often the missing covariates are high dimensional, correlated, or contain both continuous and dichotomous or categorical variables. For instance, SEER-Medicare data has nearly 100 potential confounders, including dichotomous, categorical and continuous variables, many of which are correlated. Correctly specifying the joint distribution in this setting is nearly impossible. Recently *McCandless et al.* (2012) suggest a method to use 'conditional propensity scores' to adjust for confounders available only in a supplementary dataset; this reduces the q dimensional partially measured covariates to a scalar quantity. We propose two methods that build on their work.

In section 3.2 we will present two approximately Bayesian methods to adjust for missing confounders using supplemental data. We assume that the supplemental data is drawn from the same underlying population but may not be a random sample of the entire population. In section 3.3 we present simulation results that compare our methods to complete case analysis using only the supplemental data and a 'naive' analysis that uses only the fully measured covariates. Finally, in section 3.4 we conclude with a discussion.

3.2 Methods

Suppose we have X, a dichotomous treatment, Y, a dichotomous outcome, $\boldsymbol{C}$, a set of fully measured covariates and $\boldsymbol{U}$, a set of partially measured covariates. We will develop our methods in the context of Medicare part A (which we will simply called Medicare) and SEER-Medicare for illustrative purposes although they can clearly be applied to any sources of data meeting our assumptions. We will denote the primary data as $\{Y_i, X_i, \boldsymbol{C}_i, \boldsymbol{U}_i\}$ for $i = 1, \ldots, n$ and the supplemental data as $\{Y_i, X_i, \boldsymbol{C}_i, \boldsymbol{U}_i\}$ for $j = n+1, \ldots, n+m$. The quantity $\boldsymbol{U}_i$ is completely unobserved in the primary data. Our goal is to estimate the marginal average causal effect (ACE) of a binary treatment, X, on outcome, Y, in the Medicare population. For comparison purposes, we will de-

fine our effect of interest as the risk difference, although we could easily define it as the risk ratio or odds ratio. More generally, we define the parameter of interest as

$$\Delta = P(Y = 1|X = 1) - P(Y = 1|X = 0)$$

Throughout, we let the subscript sup indicate observations from the supplemental data set and and the subscript $prim$ denote observations from the primary data set. Variables without either subscript include the full data. In addition to assuming that the supplemental data is from the same underlying population as the primary data (the population of interest), we further assume that the missingness depends only on observed data. Specifically, $f(\boldsymbol{U}_{prim}|X_{prim}, Y_{prim}, \boldsymbol{C}_{prim}) = f(\boldsymbol{U}_{sup}|X_{sup}, Y_{sup}, \boldsymbol{C}_{sup})$. For instance, our primary data might be Medicare while our supplemental data could be linked SEER-Medicare, which is taken from the Medicare population but is not a nationwide sample (*NCI*, 2013). But, we assume that, conditional on observed characteristics (i.e. $X, Y, \boldsymbol{C}$), the SEER-Medicare data is a random sample of the Medicare population.

We will build on the concept of conditional propensity scores presented by *McCandless et al.* (2012). We present two approximately Bayesian methods to adjust for missing confounders using supplemental data.

3.2.1 Models

Define treatment and outcome models as follows:

$$g(P(X_i = 1|\boldsymbol{C}_i, \boldsymbol{U}_i)) = \boldsymbol{C}_i\boldsymbol{\gamma} + \boldsymbol{U}_i\tilde{\boldsymbol{\gamma}} \qquad i = 1, \ldots, n+m \tag{3.1}$$

$$g(P(Y_i = 1|X_i, \boldsymbol{C}_i, Z_i)) = \beta X_i + \boldsymbol{C}_i\boldsymbol{\xi} + \boldsymbol{h}\{Z_i\}\tilde{\boldsymbol{\xi}} \qquad i = 1, \ldots, n+m \tag{3.2}$$

The parameter β represents the effect of X on Y conditional on $\boldsymbol{C}$ and $\boldsymbol{U}$, $g(\cdot)$ is a link function and $Z_i = \boldsymbol{U}_i\tilde{\boldsymbol{\gamma}}$ is the propensity score conditional on $\boldsymbol{C}$ (*McCandless et al.*, 2012). The deterministic function $\boldsymbol{h}\{Z_i\}$ specifies how Z_i enters the outcome model, for instance as a linear predictor or natural cubic spline basis. While not a true propensity score, Z is a balancing quantity in that it balances the distribution of the *missing* covariates ($\boldsymbol{U}$), conditional on the observed covariates ($\boldsymbol{C}$), between treated ($X = 1$) and untreated ($X = 0$). *McCandless et al.* (2012) prove that if there is no unmeasured confounding conditional on $(\boldsymbol{C}, \boldsymbol{U})$ then there is no unmeasured confounding conditional on $(\boldsymbol{C}, Z)$. Therefore, we can estimate the treatment effect by modeling the conditional distribution of Y given $(X, \boldsymbol{C}, Z)$. Note that we will always adjust for $\boldsymbol{C}$ by including $\boldsymbol{C}$ as a linear term in the regression model although further extensions are possible.

The corresponding complete data likelihood for (3.1) - (3.2) is then:

$$P(Y, X|\boldsymbol{C}, \boldsymbol{U}, \boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}, \beta, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}}) = \prod_{i=1}^{n+m} P(Y_i, X_i|\boldsymbol{C}_i, \boldsymbol{U}_i, \boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}, \beta, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}}) \tag{3.3}$$

Complicating matters is the concept of 'feedback'. In a fully Bayesian model, we would sample from the joint posterior distribution which can be obtained by combining the complete data likelihood in (3.3) with the prior distributions on our model parameters. But note that this likelihood can be broken into the components corresponding to treatment assignment (e.g. the propensity score model) and outcome model as follows

$$P(Y, X|\boldsymbol{C}, \boldsymbol{U}, \boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}, \beta, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}}) = \underbrace{P(X|\boldsymbol{C}, \boldsymbol{U}, \boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}})}_{\text{PS Model}} \underbrace{P(Y|X, \boldsymbol{C}, \boldsymbol{U}, \tilde{\boldsymbol{\gamma}}, \beta, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}})}_{\text{Outcome Model}}$$

Notice that the parameter $\tilde{\gamma}$ appears in both terms in the likelihood. That is, quantities in the outcome model, specifically Y, influence estimation of $\tilde{\gamma}$ and hence, Z. This is the nature of feedback. *Zigler et al.* (2013) examine this concept in detail. They found that, in general, this feedback distorts estimates of the parameters in the treatment (or propensity score) model and that this distortion adversely impacts the balancing score property of the PS. As a result, Bayesian propensity score models that use a joint likelihood for a PS model and an outcome model are not guaranteed to uncover treatment effects. *Zigler et al.* (2013) show that an outcome model that adjusts for the PS and also for every covariate included in the PS model can accurately estimate the treatment effect. However, this additional covariate adjustment is not available in our setting due to the nature of the missing data. We will instead take the sequential approach examined in Chapter 1 and develop an approximately Bayesian procedure.

The likelihood in (3.3) can be decomposed into the portion corresponding to the primary data and the portion corresponding to the supplemental data as follows:

$$\begin{aligned} P(Y, X|\boldsymbol{C}, \boldsymbol{U}, \boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}, \beta, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}}) &= P(Y_{prim}, X_{prim}|\boldsymbol{C}_{prim}, Z_{prim}, \boldsymbol{\gamma}, \beta, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}}) \\ &\quad \times P(Y_{sup}, X_{sup}|\boldsymbol{C}_{sup}, \boldsymbol{U}_{sup}, \boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}, \beta, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}}) \\ &= \prod_{i=1}^{n} P(Y_i, X_i|\boldsymbol{C}_i, Z_i, \boldsymbol{\gamma}, \beta, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}}) \prod_{i=n+1}^{n+m} P(Y_i, X_i|\boldsymbol{C}_i, \boldsymbol{U}_i, \boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}, \beta, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}}) \end{aligned}$$

Similarly, we can decompose (3.1) into (3.4) and (3.6) and (3.2) into (3.5) and (3.7):

$$g(P(X_i = 1|\boldsymbol{C}_i, Z_i)) = \boldsymbol{C}_i\boldsymbol{\gamma} + Z_i \qquad i = 1, \ldots, n \qquad (3.4)$$

$$g(P(Y_i = 1|X_i, \boldsymbol{C}_i, Z_i)) = \beta X_i + \boldsymbol{C}_i\boldsymbol{\xi} + \boldsymbol{h}\{Z_i\}\tilde{\boldsymbol{\xi}} \qquad i = 1, \ldots, n \qquad (3.5)$$

$$g(P(X_i = 1|\boldsymbol{C}_i, \boldsymbol{U}_i)) = \boldsymbol{C}_i\boldsymbol{\gamma} + \boldsymbol{U}_i\tilde{\boldsymbol{\gamma}} \qquad i = n+1, \ldots, n+m \qquad (3.6)$$

$$g(P(Y_i = 1|X_i, \boldsymbol{C}_i, Z_i)) = \beta X_i + \boldsymbol{C}_i\boldsymbol{\xi} + \boldsymbol{h}\{\boldsymbol{U}_i\tilde{\boldsymbol{\gamma}}\}\tilde{\boldsymbol{\xi}} \qquad i = n+1, \ldots, n+m \qquad (3.7)$$

where the Z_i in (3.4) and (3.5) is a missing variable. Notice that we could define $\boldsymbol{U}_i\tilde{\gamma}$ in (3.6) and (3.7) as Z_i; we choose to leave it as $\boldsymbol{U}_i\tilde{\gamma}$ to highlight the fact that $\tilde{\gamma}$ appears in these models. We then take a Bayesian data augmentation approach to calculate the posterior distribution of $P(Z_{prim}, \boldsymbol{\gamma}, \tilde{\gamma}, \beta, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}}|Y, X, \boldsymbol{C}, \boldsymbol{U}_{sup})$. We must first specify a model for the distribution of Z. We assume that Z and $\boldsymbol{C}$ are not independent - an assumption we feel is reasonable in most settings - and specify $f(Z|\boldsymbol{C}, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a parameter vector; i.e. $E[Z_i|\boldsymbol{C}_i] = s(\boldsymbol{C}_i, \boldsymbol{\theta})$.

Both of our proposed methods use Bayesian data augmentation to impute Z in the primary data but they differ in how they make use of the available data. First, in section 3.2.2 we propose a sequential Bayesian model (SB) that cuts the feedback between the PSand outcome models. This method simultaneously imputes the missing Z_i in the primary data and estimates the coefficients in the PS model (3.1). Then, given (Z_i) we estimate the coefficients in the outcome model (3.2). Next, in section 3.2.3 we propose a two-stage approach (TSB). In stage one, Z_i is estimated from (3.1) for all subjects in the supplemental data. In stage two, Z_i is imputed for subjects in the primary data while simultaneously estimating the regression coefficients from (3.2).

The quantity Δ in the Medicare part A population is fully defined by the regression parameters in (3.2) and the distributions of $\boldsymbol{C}$ and Z. Specifically,

$$\begin{aligned}\Delta = P(Y=1|X=1) - P(Y=1|X=0) = &\int\int P(Y=1|X=1, \boldsymbol{C}, Z)P(\boldsymbol{C}, Z)\mathrm{d}\boldsymbol{C}\mathrm{d}Z \\ &- \int\int P(Y=1|X=0, \boldsymbol{C}, Z)P(\boldsymbol{C}, Z)\mathrm{d}\boldsymbol{C}\mathrm{d}Z\end{aligned}$$

It is important to note that this quantity is not even estimable using the SEER-Medicare data alone because it is not a random sample from the Medicare population.

3.2.2 Sequential Bayesian

Here we propose an approximately Bayesian method (SB) that 'cuts the feedback' (*McCandless et al.*, 2010) from the outcome model to the PS model.

1. First we estimate the posterior distribution of $(Z_{prim}, \boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}})$ given the both the primary and supplemental data, $P(Z_{prim}, \boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}} | X, \boldsymbol{C}, \boldsymbol{U}_{sup})$.
2. Next we estimate the posterior distribution of $(\beta, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}})$ again given both the primary and supplemental data and $(\tilde{\boldsymbol{\gamma}}, Z_{prim})$. That is, for every $(Z_{prim}, \boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}})$ sampled from the posterior distribution in step 1, we sample $(\beta, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}})$ from $P(\beta, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}} | Y, X, \boldsymbol{C}, \boldsymbol{U}_{sup}, Z_{prim}, \tilde{\boldsymbol{\gamma}})$.

This 'cuts the feedback' from the outcome model to the treatment model in the sense that information from the outcome model is not used to estimate quantities in the treatment model.

Posterior simulation is accomplished using MCMC. Details may be found in appendix A.3.1. Given posterior samples of Z_{prim} and all unknown parameters, it is straightforward to estimate the posterior distribution of Δ in the Medicare population from the empirical distribution of $\boldsymbol{C}$.

3.2.3 Two-Stage Approach

Next we take a two-stage approach (TSB).

1. In stage one, we estimate the posterior distribution of the parameters in the PS model (3.1) using only the supplemental data, $P(\boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}} | X_{sup}, \boldsymbol{C}_{sup}, \boldsymbol{U}_{sup})$. Given a posterior distribution of $\tilde{\boldsymbol{\gamma}}$ and $\boldsymbol{U}_{sup}$, we also have a posterior distribution of Z_{sup}.

2. In stage two, we estimate the posterior distribution of $(Z_{prim}, \beta, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}})$ given $\hat{Z}_{sup}$, which is defined as the mean of the posterior distribution of Z_{sup}, that is $P(Z_{prim}, \beta, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}} | Y, X, \boldsymbol{C}, \hat{Z}_{sup})$.

This method is similar to the SB method but differs primarily in two ways. In TSB all parameters from (3.1) are estimated using the supplemental data only, whereas in SB, the primary data contributed to estimation of $\boldsymbol{\gamma}$. Additionally, in TSB the parameters in the outcome model are estimated using only a point estimate for Z_{sup}, rather than the entire posterior distribution. TSB is similar to a traditional PS approach where, once calculated, the estimated PS are treated as a fixed quantity. TSB does not use the data as fully as SB but retains desirable properties when extended to other settings, a point we will come back to in section 3.4.

Posterior simulation is again accomplished using MCMC. Details may be found in appendix A.3.2. Given $P(Z_{prim}, \beta, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}} | Y, X, \boldsymbol{C}, \hat{Z}_{sup})$ and $\hat{Z}_{sup}$, it is again straightforward to estimate Δ from the empirical distribution of $\boldsymbol{C}$.

3.3 Simulation Study

We conducted a simulation study to compare our methods to two commonly used methods: complete case analysis - that is, analysis in the supplemental data only - and analysis using only the fully measured covariates, $\boldsymbol{C}$, which we will call 'naive'. As previously noted, the regression coefficient β is the effect of treatment, X (e.g. major craniotomy), on outcome, Y (e.g. 1-year risk of death), conditional on $\boldsymbol{C}$ and Z. While complete case analysis analysis allows us to estimate this effect, we cannot consistently estimate Δ in the population of interest (e.g. Medicare enrollees). The ACE in this analysis is instead the effect of X on Y in the population from which the supplemental data was drawn (e.g. the SEER-Medicare population), which is not our population of interest. Not surprisingly, if we attempt to estimate the ACE, Δ, in the entire population using only the supplemental data we found large bias in most scenarios. Here we will present results for our methods and the naive analysis.

3.3.1 Design

We generated 500 datasets from 18 scenarios: where the supplemental data = 10, 50 and 70% of the total data, where $\boldsymbol{C}$ and $\boldsymbol{U}$ are weakly and moderately correlated, and where $\boldsymbol{C}$ and $\boldsymbol{U}$ are approximately equally important in terms of confounding, where $\boldsymbol{C}$ contains 'stronger' confounders than $\boldsymbol{U}$ and where $\boldsymbol{U}$ contains stronger confounders than $\boldsymbol{C}$. All scenarios had six fully measured confounders and 10 partially measured confounders. For simplicity, the first column of $\boldsymbol{C}$ is the intercept. Total sample size for all scenarios was 2000. These scenarios are laid out in Table 3.1. We assumed a probit link function for both (3.1) and (3.2). The specific data generating mechanism is described in Appendix A.3.3. We fix $\beta = 0.5$ for all scenarios and adjust for Z in the outcome as a linear covariate. i.e. $\boldsymbol{h}\{Z\} = Z$.

Table 3.1: Simulation Scenarios. Breakdown of 18 simulation scenarios. m is the supplemental data sample size, *Confounding* represents the relative importance of confounders in $\boldsymbol{C}$ and $\boldsymbol{U}$ and $\rho_{\boldsymbol{C},\boldsymbol{U}}$ is the correlation between $\boldsymbol{C}$ and $\boldsymbol{U}$.

	m	Confounding	$\rho_{\boldsymbol{C},\boldsymbol{U}}$
1	200	=	low
2	200	=	mod
3	200	$\boldsymbol{U}$ stronger	low
4	200	$\boldsymbol{U}$ stronger	mod
5	200	$\boldsymbol{C}$ stronger	low
6	200	$\boldsymbol{C}$ stronger	mod
7	1000	=	low
8	1000	=	mod
9	1000	$\boldsymbol{U}$ stronger	low
10	1000	$\boldsymbol{U}$ stronger	mod
11	1000	$\boldsymbol{C}$ stronger	low
12	1000	$\boldsymbol{C}$ stronger	mod
13	1400	=	low
14	1400	=	mod
15	1400	$\boldsymbol{U}$ stronger	low
16	1400	$\boldsymbol{U}$ stronger	mod
17	1400	$\boldsymbol{C}$ stronger	low
18	1400	$\boldsymbol{C}$ stronger	mod

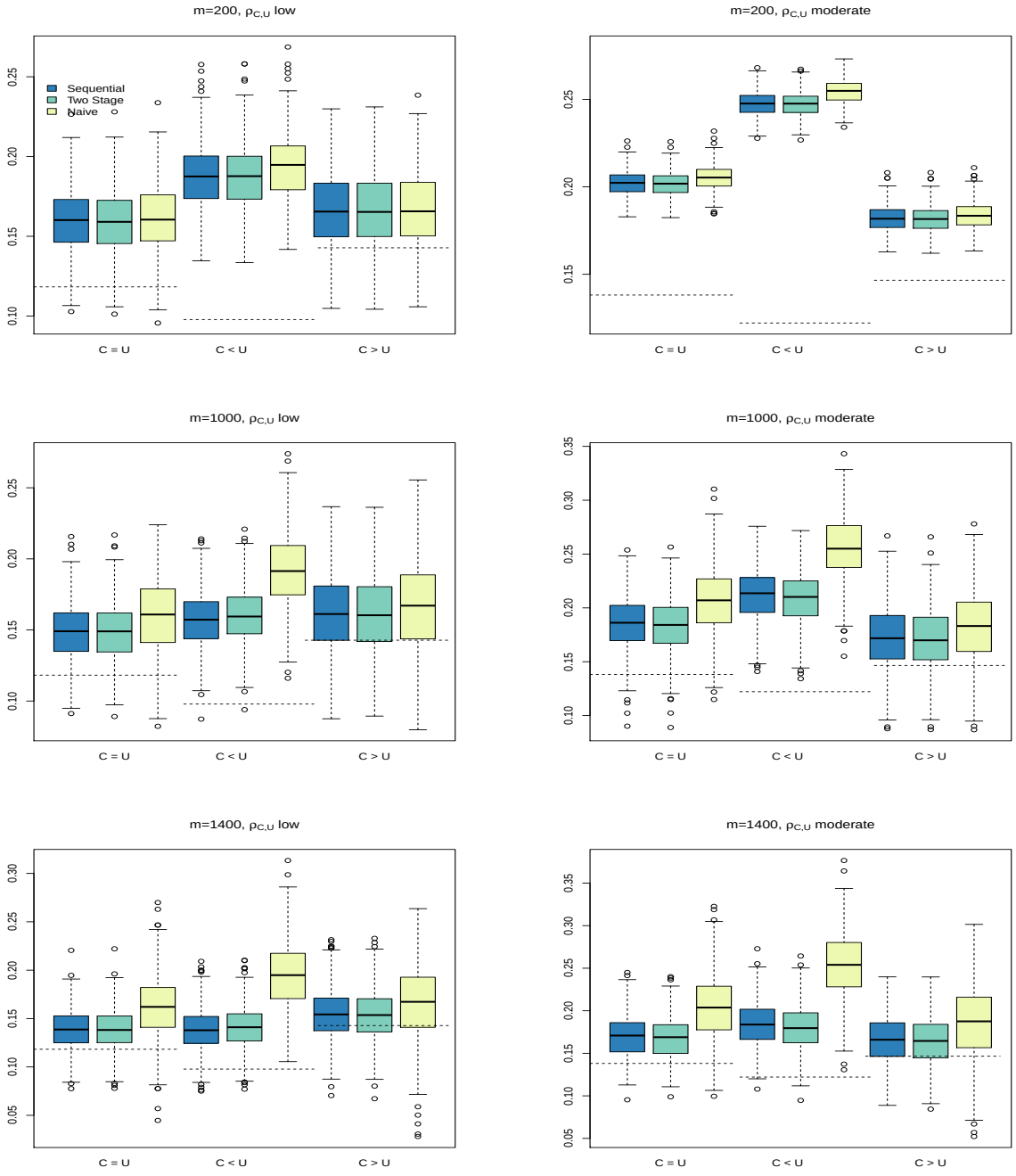


Figure 3.1: Boxplots of $\hat{\Delta}$ from the sequential Bayesian, two-stage Bayesian and naive analysis of 500 data sets for all 18 scenarios. The darkest boxes are from the SB analysis, the medium boxes from the TSB analysis and the light boxes from the naive analysis. Dashed lines indicate the true value of Δ for each scenario. The first row of plots are from the scenarios with m=200, the second row m=1000 and the third row m=1400. The left column of plots are from scenarios where $\boldsymbol{C}$ and $\boldsymbol{U}$ are weakly correlated, the right columns where they are moderately correlated. Within each plot, the leftmost three boxes are scenarios where $\boldsymbol{C}$ and $\boldsymbol{U}$ contain approximately equally important confounders, the center three boxes are scenarios where $\boldsymbol{U}$ contains the most important confounders and the right three boxes are scenarios where $\boldsymbol{C}$ contains the most important confounders.

3.3.2 Results

We found that across scenarios our methods never do worse than the naive analysis and in many cases perform much better. Figure 3.1 shows boxplots of the results of our simulations. Each box represents the distribution of effect estimates across data sets. Our estimate, $\hat{\Delta}$, is the mean of the posterior distribution of Δ.

When m=200, or 10% of the sample size (the first row of figure 3.1), we found that there is not much gain in bias or mean squared error (MSE) by using our methods over simply fitting the fully measured covariates. All methods performed comparably so there is little to motivate the extra effort required to use our methods. When m=1000, or 50% of the sample size (the second row of figure 3.1), we see a significant reduction in both bias and MSE with both of our methods over adjusting for only the fully measured covariates except when $\boldsymbol{C}$ contains the important confounders. When m=1400, or 70% of the overall sample (the third row of figure 3.1), we again see a significant reduction in both bias and MSE with both of our methods over adjusting for only the fully measured covariates in most scenarios. Table 3.2 shows the % bias and MSE reduction by scenario for both the SB and TSB methods over the naive analysis.

It is worth noting that, even if our interest were in estimating the conditional effect, β, complete case analysis fails to converge in many data sets when m=200 and even when m=1000 or 1400, the MSE is significantly greater than either of our methods in most scenarios. (results not shown)

Ultimately, both the SB analysis and the TSB analysis performed comparably across scenarios. They never performed worse than the naive analysis and in situations with a reasonably sized supplemental data set, they tended to perform significantly better, in some cases eliminating bias altogether.

Table 3.2: Bias and MSE Reduction. Same 18 simulation scenarios as depicted in Table 3.1. % Reduction is the reduction using SB and TSB, respectively, compared to the naive analysis.

	m	Confounding	$\rho_{\boldsymbol{C},\boldsymbol{U}}$	% Bias (SB) Reduction	%Bias (TSB) Reduction	% MSE (SB) Reduction	%MSE (TSB) Reduction
1	200	=	low	5.7	4.7	10.4	8.9
2	200	=	mod	6.1	4.9	11.2	9.2
3	200	$\boldsymbol{U}$ stronger	low	6.6	6.6	12.4	12.5
4	200	$\boldsymbol{U}$ stronger	mod	6.4	5.8	12.1	11.1
5	200	$\boldsymbol{C}$ stronger	low	5.5	4.4	8.9	7.7
6	200	$\boldsymbol{C}$ stronger	mod	5.0	4.1	8.4	7.2
7	1000	=	low	28.6	28.7	44.2	44.3
8	1000	=	mod	33.2	30.0	51.6	48.0
9	1000	$\boldsymbol{U}$ stronger	low	34.2	37.0	55.5	58.9
10	1000	$\boldsymbol{U}$ stronger	mod	35.0	32.7	56.5	53.6
11	1000	$\boldsymbol{C}$ stronger	low	25.6	23.7	36.9	35.9
12	1000	$\boldsymbol{C}$ stronger	mod	32.4	28.1	44.4	41.3
13	1400	=	low	51.3	51.9	69.0	69.3
14	1400	=	mod	54.6	50.9	74.4	71.8
15	1400	$\boldsymbol{U}$ stronger	low	54.7	57.7	77.0	79.3
16	1400	$\boldsymbol{U}$ stronger	mod	55.6	52.3	78.2	75.5
17	1400	$\boldsymbol{C}$ stronger	low	54.2	52.5	59.9	59.4
18	1400	$\boldsymbol{C}$ stronger	mod	55.9	51.9	67.1	65.4

3.3.3 Sensitivity Analysis

More flexible adjustment for Z in the outcome model could, in theory, yield better results for both of our methods. For instance, for the SB method, we also let $\boldsymbol{h}\{\cdot\}$ be a natural cubic spline basis function. Our results for this adjustment were nearly identical to our results where $\boldsymbol{h}\{\cdot\}$ is the identity function and we present the results for the more simple adjustment here.

Throughout our simulations we assume a linear regression model for $f(Z|\boldsymbol{C}, \boldsymbol{\theta})$. In our data sets this was a reasonable assumption, but in other data sets, any model should be evaluated for feasibility.

3.4 Discussion

Combining heterogeneous sources of information has the potential to allow us to make use of very large data sets that are perhaps missing key confounders and smaller supplemental data with rich covariate information on only a subset of the population in order to estimate treatment effects in the larger population. Existing methods are difficult to implement in the situation where the partially measured covariates are high dimensional or contain both continuous and categorical covariates.

Sturmer et al. (2005) previously proposed propensity score calibration as a method to combine heterogenous data sources. Their method treats a propensity score calculated using only the fully measured covariates as measured with error. They then use validation data and a measurement error model to account for covariates present only in the validation data. However, their method relies on the assumption that the propensity score measured with error is a surrogate for the 'gold standard' propensity score available in the validation data - a condition that would be violated any time the direction of confounding from the partially observed covariates differs from that of the fully observed covariates (*Sturmer et al.*, 2007). Further, they show that violations of this assumption can actually lead to an increase in bias. *McCandless et al.* (2012) suggest another propensity score method that does not rely on the surrogacy assumption. Their method uses 'conditional propensity scores' to adjust for confounders available only in a supplementary dataset and reduces the q dimensional partially measured covariates to a scalar quantity. Our two proposed methods build on their work but address an important limitation with their methods. We do not fit the PS and outcome models using the joint likelihood, as this is shown by *Zigler et al.* (2013) to give biased estimates of the desired causal effect in most settings. We instead use approximately Bayesian methods that cut the feedback from the outcome model to the PS model.

Our second method, TSB, has the distinguishing feature that the conditional propensity score is estimated using only the supplemental data. Although this method doesn't use the data as

fully as the SB method, it provides the building block for more complex extensions. *Zigler and Dominici* (2013) recently propose methods of Bayesian variable selection in PS models. The TSB method can easily be extended to accommodate variable selection of the potential confounders based on their joint association with exposure and outcome. While variable selection can also be accomplished in the SB method, we would sacrifice the ability of the outcome to inform which variables to include in the PS model.

Throughout we assume a linear regression model for $f(Z|\boldsymbol{C}, \boldsymbol{\theta})$ – a reasonable assumption in our simulated data. However, more complex models might be necessary in other settings. Strategies to marginalize over the distribution of Z, rather than imputing the missing Z, could also be implemented. We expect that they would perform similarly to the methods presented but do not investigate this here. Additionally, if we were uncertain of our choice of model for $f(Z|\boldsymbol{C}, \boldsymbol{\theta})$, a multiple imputation approach might be advisable over a full Bayesian data augmentation. We evaluated this approach in our simulations using the TSB method and found the results were nearly identical to the full Bayesian data augmentation in our scenarios. Throughout we assume no interactions between the fully measured covariates $\boldsymbol{C}$ and the partially measured covariates $\boldsymbol{U}$. While presented in the context of a dichotomous outcome, extending these methods to a continuous or categorical outcome is straightforward.

Although in many of the settings we investigated, our estimates are still biased, the bias is significantly reduced over fitting only the fully measured covariates, a common approach. Further extensions, such as a model averaging approach, could improve the performance of our methods in many settings.

Appendices

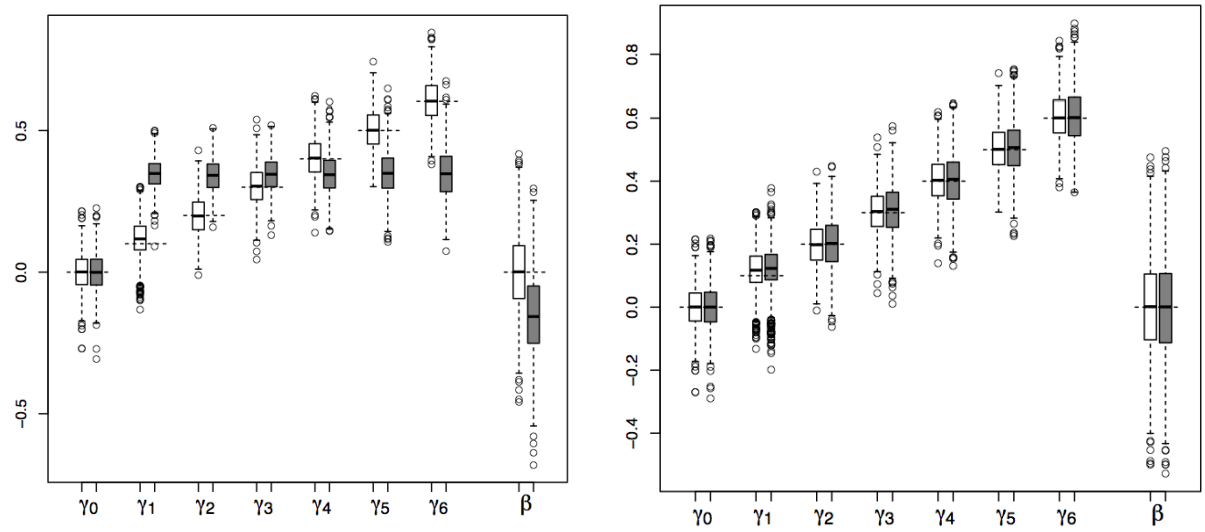
A.1 Model Feedback in Bayesian Propensity Score Estimation - Appendix

A.1.1 Simulation study with very flexible specification of $h(\gamma, C)$

To further demonstrate the behavior of feedback in the joint Bayesian method, we conduct a simulation study paralleling that of Scenarios 3 and 3^+ of the main text where every covariate exits a unique treatment-covariate/outcome-covariate relationship, but we analyze the data with a more flexibly specified outcome model. We refer to this simulation as Scenario 4. We simulate data from expressions (1.8) and (1.9) of the main text, with $(\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6) = (0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6)$ and $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6) = (0.0, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1)$. Rather than adjusting for subclasses based on PS quintiles, we specify $h(\gamma, C)$ as a natural cubic spline basis with 10 knots, placed at the deciles of logit (PS). Figure A.1a depicts boxplots of posterior mean estimates of γ and β for an analysis with $\delta = 0$ for the sequential frequentist and joint Bayesian and methods. Figure A.1b depicts the same for an analysis with $\delta \neq 0$ and $\mathrm{C}^+ \equiv C$, labeled Scenario 4^+. Note that the latter analysis with $\delta \neq 0$ is analogous to the penalized spline of propensity prediction method of *Little* (2011). The results of this simulation closely parallel those in the main text; even when the PS enters the outcome in a highly flexible manner, failure to adjust for additional covariates leads feedback to distort the estimates of γ and, ultimately the causal effect.

A.1.2 Acknowledgements

This work was funded by NCI P01CA134294, USEPA RD83479801, and HEI 4909. The contents of this work are solely the responsibility of the grantee and do not necessarily represent the official views of the USEPA. Further, USEPA does not endorse the purchase of any commercial products or services mentioned in the publication. The authors thank Giovanni Parmigiani, Sebastien Haneuse, and Matt Cefalu for helpful discussion.



(a) Scenario 2, $\delta = 0$

(b) Scenario 2^+, $\delta \neq 0$

Figure A.1: Scenarios 4 and 4^+ with $(\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6) = (0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6)$, and $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6) = (0.0, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1)$: boxplots of estimates of γ and β from the sequential frequentist and joint Bayesian analysis of 1000 replicated data sets. Horizontal dotted lines are at the true parameter values.

A.2 Bayesian Adjustment for Confounding in the Presence of Multiple Exposures - Appendices

A.2.1 Example with Marginal but not Joint Independence

The following is an example of where we have marginal independence between a covariate and two exposures separately but not between the covariate and the exposures jointly. To illustrate we use dichotomous exposures and covariates, which can easily be summarized in contingency tables. For instance, C could be an indicator variable such that $C = 1$ if a county has a medicare population that is younger than average and 0 otherwise. Also, imagine that $X_1 = 1$ if a county had "high" ozone and 0 otherwise and $X_2 = 1$ if a county had "high" $PM_{2.5}$ and 0 otherwise. Further, suppose that our data can be summarized as follows:

Then, $P(C = 1|X_1 = 1) = \frac{4}{44} = .0909, P(C = 1|X_1 = 0) = \frac{8}{88} = .0909$ and $P(C = 1) = \frac{12}{132} = .0909 \Rightarrow C \perp\!\!\!\perp X_1$

		C	
		1	0
X_1	1	4	40
	0	8	80

		C	
		1	0
X_2	1	10	100
	0	2	20

	C	
	1	0
1,1	4	36
1,0	0	4
0,1	6	64
0,0	2	16

(X_1, X_2)

Also, $P(C = 1|X_2 = 1) = \frac{10}{110} = .0909, P(C = 1|X_2 = 0) = \frac{2}{22} = .0909$ and $P(C = 1) = \frac{12}{132} = .0909 \Rightarrow C \perp\!\!\!\perp X_2$

BUT, $P(C = 1|X_1 = 1, X_2 = 1) = \frac{4}{40} = .100 \neq P(C = 1)$ and $P(C = 1|X_1 = 1, X_2 = 0) = \frac{0}{4} = 0, P(C = 1|X_1 = 0, X_2 = 1) = \frac{6}{70} = .085, P(C = 1|X_1 = 1, X_2 = 1) = \frac{2}{18} = .111 \Rightarrow C \not\perp\!\!\!\perp \{X_1, X_2\}$

A.2.2 Prior Distributions

A.2.2.1 Complete distributions for prior odds ratios given in section 2.2.4

Consider the most simple formulation of prior odds ratios given in section 2.2.4

$$\frac{P(\alpha_m^Y = 1|\alpha_m^{X_1} = 1, \alpha_m^{X_2} = 1)}{P(\alpha_m^Y = 0|\alpha_m^{X_1} = 1, \alpha_m^{X_2} = 1)} = \frac{P(\alpha_m^Y = 1|\alpha_m^{X_1} = 1, \alpha_m^{X_2} = 0)}{P(\alpha_m^Y = 0|\alpha_m^{X_1} = 1, \alpha_m^{X_2} = 0)} = \frac{P(\alpha_m^Y = 1|\alpha_m^{X_1} = 0, \alpha_m^{X_2} = 1)}{P(\alpha_m^Y = 0|\alpha_m^{X_1} = 0, \alpha_m^{X_2} = 1)} = \omega$$

$$\frac{P(\alpha_m^Y = 1|\alpha_m^{X_1} = 0, \alpha_m^{X_2} = 0)}{P(\alpha_m^Y = 0|\alpha_m^{X_1} = 0, \alpha_m^{X_2} = 0)} = 1$$

Additionally, in order to ensure a marginal probability $P(\alpha_m^{X_1} = 1) = P(\alpha_m^{X_2} = 1) = \frac{1}{2}$, we specify the following prior odds ratios:

$$\frac{P(\alpha_m^{X_1} = 1|\alpha_m^Y = 1)}{P(\alpha_m^{X_1} = 0|\alpha_m^Y = 1)} = \frac{P(\alpha_m^{X_2} = 1|\alpha_m^Y = 1)}{P(\alpha_m^{X_2} = 0|\alpha_m^Y = 1)} = \frac{4\omega}{3\omega + 1} \tag{1}$$

$$\frac{P(\alpha_m^{X_1}=1|\alpha_m^Y=0)}{P(\alpha_m^{X_1}=0|\alpha_m^Y=0)} = \frac{P(\alpha_m^{X_2}=1|\alpha_m^Y=0)}{P(\alpha_m^{X_2}=0|\alpha_m^Y=0)} = \frac{4}{\omega+3} \overset{\omega\to\infty}{\longrightarrow} 0 \tag{2}$$

The prior odds in (2) assign a very low probability (0 for $\omega = \infty$) of being included in either of the two exposure models if that potential confounder is not included into the outcome model. A marginal probability of $\frac{1}{2}$ implies a priori naiveté as to which covariates are associated with the exposures; (1) and (2) could be adjusted for a different prior belief. These odds ratios also assume a priori that the two exposure models are independent.

The above prior odds ratios imply the following joint, conditional and marginal probabilities:
Joint Probabilities:

$$\begin{aligned}
&P(\alpha_m^Y=1,\alpha_m^{X_1}=1,\alpha_m^{X_2}=1) = P(\alpha_m^Y=1,\alpha_m^{X_1}=1,\alpha_m^{X_2}=0)\\
&= P(\alpha_m^Y=1,\alpha_m^{X_1}=0,\alpha_m^{X_2}=1) = \frac{1}{4}\frac{\omega}{\omega+1}\\
&P(\alpha_m^Y=0,\alpha_m^{X_1}=1,\alpha_m^{X_2}=1) = P(\alpha_m^Y=0,\alpha_m^{X_1}=1,\alpha_m^{X_2}=0)\\
&= P(\alpha_m^Y=0,\alpha_m^{X_1}=0,\alpha_m^{X_2}=1) = \frac{1}{4}\frac{1}{\omega+1}\\
&P(\alpha_m^Y=1,\alpha_m^{X_1}=0,\alpha_m^{X_2}=0) = P(\alpha_m^Y=0,\alpha_m^{X_1}=0,\alpha_m^{X_2}=0) = \frac{1}{8}
\end{aligned}$$

Conditional Probabilities:

$$\begin{aligned}
&P(\alpha_m^Y=1|\alpha_m^{X_1}=1,\alpha_m^{X_2}=1) = P(\alpha_m^Y=1|\alpha_m^{X_1}=1,\alpha_m^{X_2}=0)\\
&= P(\alpha_m^Y=1|\alpha_m^{X_1}=0,\alpha_m^{X_2}=1) = \frac{\omega}{\omega+1}
\end{aligned}$$

$$P(\alpha_m^Y=1|\alpha_m^{X_1}=0,\alpha_m^{X_2}=0) = \frac{1}{2}$$

$$P(\alpha_m^{X_1} = 1|\alpha_m^Y = 1) = P(\alpha_m^{X_2} = 1|\alpha_m^Y = 1) = \frac{4\omega}{7\omega + 1}$$
$$P(\alpha_m^{X_1} = 1|\alpha_m^Y = 0) = P(\alpha_m^{X_2} = 1|\alpha_m^Y = 0) = \frac{4}{\omega + 7}$$

Marginal Probabilities:

$$P(\alpha_m^{X_1} = 1) = P(\alpha_m^{X_2} = 1) = 1/2$$
$$P(\alpha_m^Y = 1) = \frac{3}{4}\frac{\omega}{\omega + 1} + \frac{1}{8}$$

As previously mentioned, the above priors are unnecessarily restrictive. A more general formulation follows. This formulations allows us to treat the two exposure models differently or give higher probability to those covariates associated with both treatments.

Conditional Odds Ratios

$$\frac{P(\alpha_m^Y = 1|\alpha_m^{X_1} = 1, \alpha_m^{X_2} = 1)}{P(\alpha_m^Y = 0|\alpha_m^{X_1} = 1, \alpha_m^{X_2} = 1)} = \omega_3$$
$$\frac{P(\alpha_m^Y = 1|\alpha_m^{X_1} = 1, \alpha^{X_2} = 0)}{P(\alpha_m^Y = 0|\alpha_m^{X_1} = 1, \alpha^{X_2} = 0)} = \omega_1$$
$$\frac{P(\alpha_m^Y = 1|\alpha^{X_1} = 0, \alpha_m^{X_2} = 1)}{P(\alpha_m^Y = 0|\alpha^{X_1} = 0, \alpha_m^{X_2} = 1)} = \omega_2$$
$$\frac{P(\alpha_m^Y = 1|\alpha_m^{X_1} = 0, \alpha_m^{X_2} = 0)}{P(\alpha_m^Y = 0|\alpha_m^{X_1} = 0, \alpha_m^{X_2} = 0)} = 1$$

$$\frac{P(\alpha_m^{X_1} = 1|\alpha_m^Y = 1)}{P(\alpha_m^{X_1} = 0|\alpha_m^Y = 1)} = \frac{2(\omega_2 + 1)(\omega_1 + \omega_3 + 2\omega_1\omega_3)}{(1 + \omega_1)(1 + 3\omega_2)(1 + \omega_3)}$$
$$\frac{P(\alpha_m^{X_2} = 1|\alpha_m^Y = 1)}{P(\alpha_m^{X_2} = 0|\alpha_m^Y = 1)} = \frac{2(\omega_1 + 1)(\omega_2 + \omega_3 + 2\omega_2\omega_3)}{(1 + \omega_2)(1 + 3\omega_1)(1 + \omega_3)}$$
$$\frac{P(\alpha_m^{X_1} = 1|\alpha_m^Y = 0)}{P(\alpha_m^{X_1} = 0|\alpha_m^Y = 0)} = \frac{2(1 + \omega_2)(2 + \omega_1 + \omega_3)}{(1 + \omega_1)(2 + \omega_2)(1 + \omega_3)}$$
$$\frac{P(\alpha_m^{X_2} = 1|\alpha_m^Y = 0)}{P(\alpha_m^{X_2} = 0|\alpha_m^Y = 0)} = \frac{2(1 + \omega_1)(2 + \omega_2 + \omega_3)}{(1 + \omega_2)(2 + \omega_1)(1 + \omega_3)}$$

Joint Probabilities:

$$
\begin{aligned}
P(\alpha_m^Y = 1, \alpha_m^{X_1} = 1, \alpha_m^{X_2} = 1) &= \frac{1}{4}\frac{\omega_3}{\omega_3 + 1} \\
P(\alpha_m^Y = 1, \alpha_m^{X_1} = 1, \alpha_m^{X_2} = 0) &= \frac{1}{4}\frac{\omega_1}{\omega_1 + 1} \\
P(\alpha_m^Y = 1, \alpha_m^{X_1} = 0, \alpha_m^{X_2} = 1) &= \frac{1}{4}\frac{\omega_2}{\omega_2 + 1} \\
P(\alpha_m^Y = 0, \alpha_m^{X_1} = 1, \alpha_m^{X_2} = 1) &= \frac{1}{4}\frac{1}{\omega_3 + 1} \\
P(\alpha_m^Y = 0, \alpha_m^{X_1} = 1, \alpha_m^{X_2} = 0) &= \frac{1}{4}\frac{1}{\omega_1 + 1} \\
P(\alpha_m^Y = 0, \alpha_m^{X_1} = 0, \alpha_m^{X_2} = 1) &= \frac{1}{4}\frac{1}{\omega_2 + 1}
\end{aligned}
$$

$$
P(\alpha_m^Y = 1, \alpha_m^{X_1} = 0, \alpha_m^{X_2} = 0) = P(\alpha_m^Y = 0, \alpha_m^{X_1} = 0, \alpha_m^{X_2} = 0) = \frac{1}{8}
$$

Conditional Probabilities:

$$
\begin{aligned}
P(\alpha_m^Y = 1 | \alpha_m^{X_1} = 1, \alpha_m^{X_2} = 1) &= \frac{\omega_3}{\omega_3 + 1} \\
P(\alpha_m^Y = 1 | \alpha_m^{X_1} = 1, \alpha_m^{X_2} = 0) &= \frac{\omega_1}{\omega_1 + 1} \\
P(\alpha_m^Y = 1 | \alpha_m^{X_1} = 0, \alpha_m^{X_2} = 1) &= \frac{\omega_2}{\omega_2 + 1} \\
P(\alpha_m^Y = 1 | \alpha_m^{X_1} = 0, \alpha_m^{X_2} = 0) &= \frac{1}{2}
\end{aligned}
$$

$$
\begin{aligned}
P(\alpha_m^{X_1} = 1 | \alpha_m^Y = 1) &= \frac{\frac{\omega_3}{\omega_3+1} + \frac{\omega_1}{\omega_1+1}}{\frac{\omega_3}{\omega_3+1} + \frac{\omega_2}{\omega_2+1} + \frac{\omega_1}{\omega_1+1} + \frac{1}{2}} \\
P(\alpha_m^{X_2} = 1 | \alpha_m^Y = 1) &= \frac{\frac{\omega_3}{\omega_3+1} + \frac{\omega_2}{\omega_2+1}}{\frac{\omega_3}{\omega_3+1} + \frac{\omega_2}{\omega_2+1} + \frac{\omega_1}{\omega_1+1} + \frac{1}{2}} \\
P(\alpha_m^{X_1} = 1 | \alpha_m^Y = 0) &= \frac{\frac{1}{\omega_3+1} + \frac{1}{\omega_1+1}}{\frac{1}{\omega_3+1} + \frac{1}{\omega_2+1} + \frac{1}{\omega_1+1} + \frac{1}{2}} \\
P(\alpha_m^{X_2} = 1 | \alpha_m^Y = 0) &= \frac{\frac{1}{\omega_3+1} + \frac{1}{\omega_2+1}}{\frac{1}{\omega_3+1} + \frac{1}{\omega_2+1} + \frac{1}{\omega_1+1} + \frac{1}{2}}
\end{aligned}
$$

Marginal Probabilities:

$$
P(\alpha_m^{X_1} = 1) = P(\alpha_m^{X_2} = 1) = 1/2
$$

$$P(\alpha_m^Y = 1) = \frac{1}{8}(1 + \frac{2\omega_1}{\omega_1 + 1} + \frac{2\omega_2}{\omega_2 + 1} + \frac{2\omega_3}{\omega_3 + 1})$$

Priors on all other parameters are as recommended by *Raftery et al.* (1997) and are given below:

Priors:

- $\eta^{\alpha^{X_1}}|(\alpha^{X_1}, \sigma^2_{X_1}) \sim N(\mu_{0\alpha^{X_1}}, \sigma^2_{X_1}\phi^2\Sigma_{0\alpha^{X_1}})$
- $\eta^{\alpha^{X_2}}|(\alpha^{X_2}, \sigma^2_{X_2}) \sim N(\mu_{0\alpha^{X_2}}, \sigma^2_{X_2}\phi^2\Sigma_{0\alpha^{X_2}})$
- $(\beta_1, \beta_2, \eta^{\alpha^Y})|(\alpha^Y, \sigma^2_Y) \sim N(\mu_{0\alpha^Y}, \sigma^2_Y\phi^2\Sigma_{0\alpha^Y})$
- $\sigma^2_{X_1}, \sigma^2_{X_2}, \sigma^2_Y \sim Gamma(\nu/2, \nu\lambda/2)$, where $(\nu\lambda/2)$ is the inverse-scale parameter of the Gamma distribution (i.e. E[$\sigma^2_{X_1}$] = $1/\lambda$)

Hyperparameters:

- $\mu_{0\alpha^{X_1}} = \mu_{0\alpha^{X_2}} = \mu_{0\alpha^Y} = \mathbf{0}$
- $\Sigma_{0\alpha^{X_1}}, \Sigma_{0\alpha^{X_2}}, \Sigma_{0\alpha^Y}$ are diagonal matrices with elements equal to s^2_m
- $\phi = 2.85$
- $\nu = 2.58$
- $\lambda = 0.28$

A.2.3 Posterior Distributions

A.2.3.1 Assumptions

In order to derive the full conditionals and simplify the MCMC, several reasonable assumptions are necessary. Roughly speaking, we can think of these in terms of five basic assumptions. First,

given a fixed outcome model and the exposure, selecting that exposure model does not depend on the other exposure model, model coefficients, or data not in that exposure model. Second, given an exposure model, selection of the outcome model is independent of that exposure. Third, given fixed exposure models and the outcome, selection of the outcome model does not depend on the exposures or model coefficients. Fourth, given an outcome model, selection of the exposure models does not depend on the outcome itself. Finally, given a model and data, estimation of the coefficient(s) from that model does not depend on the other models or their coefficients. These assumptions are given explicitly below. We believe that these are reasonable assumptions.

(A1) $\boldsymbol{\alpha}^{X_1} \perp (X_2, Y, \boldsymbol{\beta}, \gamma, \boldsymbol{\alpha}^{X_2}) | \, (\boldsymbol{\alpha}^{Y}, X_1, \boldsymbol{C})$

(A2) $X_1 \perp \boldsymbol{\alpha}^{Y} | \, (\boldsymbol{\alpha}^{X_1}, \boldsymbol{C})$

(A3) $\boldsymbol{\alpha}^{X_2} \perp (X_1, Y, \boldsymbol{\beta}, \gamma, \boldsymbol{\alpha}^{X_1}) | \, (\boldsymbol{\alpha}^{Y}, \tilde{X}_2, \boldsymbol{C})$

(A4) $\tilde{X}_2 \perp \boldsymbol{\alpha}^{Y} | \, (\boldsymbol{\alpha}^{X_2}, \boldsymbol{C})$

(A5) $\alpha^{Y} \perp (X_1, X_2, \boldsymbol{\beta}, \gamma) | \, (\alpha^{X_1}, \alpha^{X_2}, \tilde{Y}, \boldsymbol{C})$

(A6) $\tilde{Y} \perp (\alpha^{X_1}, \alpha^{X_2}) | \, (\alpha^{Y}, \boldsymbol{C})$

(A7) $\gamma \perp (\boldsymbol{\alpha}^{X_1}, \boldsymbol{\alpha}^{Y}, \boldsymbol{\beta}) | \, (\boldsymbol{\alpha}^{X_2}, D)$

(A8) $\boldsymbol{\beta} \perp (\alpha^{X_1}, \alpha^{X_2}, \gamma) | \, (\boldsymbol{\alpha}^{Y}, D)$

where $\tilde{X}_2 = X_2 - \gamma X_1$ and $\tilde{Y} = Y - (\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2)$.

A.2.3.2 Full Conditionals.

Our goal is to estimate the posterior distribution of $(\boldsymbol{\alpha}^{Y}, \boldsymbol{\alpha}^{X_1}, \boldsymbol{\alpha}^{X_2}, \Delta_{(x_1,x_2)}(\delta_1, \delta_2))$. We accomplish this by iteratively sampling from $P(\boldsymbol{\alpha}^{X_1} | \boldsymbol{\alpha}^{X_2}, \boldsymbol{\alpha}^{Y}, \boldsymbol{\beta}, \gamma, D))$, $P(\boldsymbol{\alpha}^{X_2} | \boldsymbol{\alpha}^{X_1}, \boldsymbol{\alpha}^{Y}, \boldsymbol{\beta}, \gamma, D))$, $P(\boldsymbol{\alpha}^{Y} | \boldsymbol{\alpha}^{\boldsymbol{X_1}}, \boldsymbol{\alpha}^{X_2}, \boldsymbol{\beta}, \gamma, D))$, $P(\gamma | \boldsymbol{\alpha}^{X_1}, \boldsymbol{\alpha}^{X_2}, \boldsymbol{\alpha}^{Y}, \boldsymbol{\beta}, D))$, and $P(\boldsymbol{\beta} | \boldsymbol{\alpha}^{X_1}, \boldsymbol{\alpha}^{X_2}, \boldsymbol{\alpha}^{Y}, \gamma, D))$. Recall that

D=($\boldsymbol{X}, Y, \boldsymbol{C}$). Derivations of these follow:

1. $P(\boldsymbol{\alpha}^{X_1}|\boldsymbol{\alpha}^{X_2}, \boldsymbol{\alpha}^{Y}, \boldsymbol{\beta}, \gamma, D) \overset{A1}{=} P(\boldsymbol{\alpha}^{X_1}|\boldsymbol{\alpha}^{Y}, X_1, \boldsymbol{C}) \overset{BayesThm}{=}$

   $\frac{P(X_1|\boldsymbol{\alpha}^{X_1}, \boldsymbol{\alpha}^{Y}, \boldsymbol{C})P(\boldsymbol{C})P(\boldsymbol{\alpha}^{X_1}|\boldsymbol{\alpha}^{Y})}{P(X_1, \boldsymbol{C}|\boldsymbol{\alpha}^{Y})} \overset{A2}{=} \frac{P(X_1|\boldsymbol{\alpha}^{X_1}, \boldsymbol{C})P(\boldsymbol{C})P(\boldsymbol{\alpha}^{X_1}|\boldsymbol{\alpha}^{Y})}{P(X_1, \boldsymbol{C}|\boldsymbol{\alpha}^{Y})} \propto$

   $P(X_1|\boldsymbol{\alpha}^{X_1}, \boldsymbol{C})P(\boldsymbol{\alpha}^{X_1}|\boldsymbol{\alpha}^{Y})$,

   where $P(X_1|\boldsymbol{\alpha}^{X_1}, \boldsymbol{C}) =$

   $$\frac{\Gamma(\frac{\nu+n}{2})(\nu\lambda)^{\nu/2}}{\pi^{n/2}\Gamma(\nu/2)|I_n + \phi^2 W_{\alpha^{X_1}}\Sigma_{0\alpha^{X_1}}W'_{\alpha^{X_1}}|^{1/2}}$$

   $$\times\{\lambda\nu + (X_1 - W_{\alpha^{X_1}}\mu_{0\alpha^{X_1}})'(I_n + \phi^2 W_{\alpha^{X_1}}\Sigma_{0\alpha^{X_1}}W'_{\alpha^{X_1}})^{-1}(X_1 - W_{\alpha^{X_1}}\mu_{0\alpha^{X_1}})\}^{-\frac{\nu+n}{2}}$$

2. $P(\boldsymbol{\alpha}^{X_2}|\boldsymbol{\alpha}^{X_1}, \boldsymbol{\alpha}^{Y}, \boldsymbol{\beta}, \gamma, D) \overset{A3}{=} P(\boldsymbol{\alpha}^{X_2}|\boldsymbol{\alpha}^{Y}, \tilde{X}_2, \boldsymbol{C}) \overset{BayesThm}{=}$

   $\frac{P(\tilde{X}_2|\boldsymbol{\alpha}^{X_2}, \boldsymbol{\alpha}^{Y}, \boldsymbol{C})P(\boldsymbol{C})P(\boldsymbol{\alpha}^{X_2}|\boldsymbol{\alpha}^{Y})}{P(\tilde{X}_2, \boldsymbol{C}|\boldsymbol{\alpha}^{Y})} \overset{A4}{=} \frac{P(\tilde{X}_2|\boldsymbol{\alpha}^{X_2}, \boldsymbol{C})P(\boldsymbol{C})P(\boldsymbol{\alpha}^{X_2}|\boldsymbol{\alpha}^{Y})}{P(\tilde{X}_2, \boldsymbol{C}|\boldsymbol{\alpha}^{Y})} \propto$

   $P(\tilde{X}_2|\boldsymbol{\alpha}^{X_2}, \boldsymbol{C})P(\boldsymbol{\alpha}^{X_2}|\boldsymbol{\alpha}^{Y})$,

   where $P(\tilde{X}_2|\boldsymbol{\alpha}^{X_2}) =$

   $$\frac{\Gamma(\frac{\nu+n}{2})(\nu\lambda)^{\nu/2}}{\pi^{n/2}\Gamma(\nu/2)|I_n + \phi^2 W_{\alpha^{X_2}}\Sigma_{0\alpha^{X_2}}W'_{\alpha^{X_2}}|^{1/2}}$$

   $$\times\{\lambda\nu + (\tilde{X}_2 - W_{\alpha^{X_2}}\mu_{0\alpha^{X_2}})'(I_n + \phi^2 W_{\alpha^{X_2}}\Sigma_{0\alpha^{X_2}}W'_{\alpha^{X_2}})^{-1}(\tilde{X}_2 - W_{\alpha^{X_2}}\mu_{0\alpha^{X_2}})\}^{-\frac{\nu+n}{2}}$$

   Where $\boldsymbol{W}_{\alpha^{X_j}}$ is the design matrix for exposure regression j

3. $P(\boldsymbol{\alpha}^{Y}|\boldsymbol{\alpha}^{X_1}, \boldsymbol{\alpha}^{X_2}, \boldsymbol{\beta}, D) \overset{A5}{=} P(\boldsymbol{\alpha}^{Y}|\boldsymbol{\alpha}^{X_1}, \boldsymbol{\alpha}^{X_2}, \tilde{Y}, \boldsymbol{C}) \overset{BayesThm}{=}$

   $\frac{P(\tilde{Y}|\boldsymbol{\alpha}^{X_1}, \boldsymbol{\alpha}^{X_2}, \boldsymbol{\alpha}^{Y}, \boldsymbol{C})P(\boldsymbol{C})P(\boldsymbol{\alpha}^{Y}|\boldsymbol{\alpha}^{X_1}, \boldsymbol{\alpha}^{X_2})}{P(\tilde{Y}, \boldsymbol{C}|\boldsymbol{\alpha}^{X_1}, \boldsymbol{\alpha}^{X_2})} \overset{A6}{=} \frac{P(\tilde{Y}|\boldsymbol{\alpha}^{Y}, \boldsymbol{C})P(\boldsymbol{C})P(\boldsymbol{\alpha}^{Y}|\boldsymbol{\alpha}^{X_1}, \boldsymbol{\alpha}^{X_2})}{P(\tilde{Y}, \boldsymbol{C}|\boldsymbol{\alpha}^{X_1}, \boldsymbol{\alpha}^{X_2})} \propto$

   $P(\tilde{Y}|\boldsymbol{\alpha}^{Y}, \boldsymbol{C})P(\boldsymbol{\alpha}^{Y}|\boldsymbol{\alpha}^{X_1}, \boldsymbol{\alpha}^{X_2})$,

where $P(\tilde{Y}|\boldsymbol{\alpha}^Y) =$

$$\frac{\Gamma(\frac{\nu+n}{2})(\nu\lambda)^{\nu/2}}{\pi^{n/2}\Gamma(\nu/2)|I_n + \phi^2 W_{\alpha^Y}\Sigma_{0\alpha^Y}W'_{\alpha^Y}|^{1/2}}$$

$$\times \{\lambda\nu + (\tilde{Y} - W_{\alpha^Y}\mu_{0\alpha^Y})'(I_n + \phi^2 W_{\alpha^Y}\Sigma_{0\alpha^Y}W'_{\alpha^Y})^{-1}(\tilde{Y} - W_{\alpha^Y}\mu_{0\alpha^Y})\}^{-\frac{\nu+n}{2}}$$

Where $\boldsymbol{W}_{\alpha^Y}$ is the design matrix for the outcome regression

4. Finally, $P(\boldsymbol{\beta}|\boldsymbol{\alpha}^{X_1}, \boldsymbol{\alpha}^{X_2}, \boldsymbol{\alpha}^Y, \gamma, D) \overset{A8}{=} P(\boldsymbol{\beta}|\boldsymbol{\alpha}^Y, D)$

$$\beta_j|\boldsymbol{\alpha}^Y, D \sim t_{n+v}(\beta_{j,n\boldsymbol{\alpha}^Y}, \sigma^2_{j,n\boldsymbol{\alpha}^Y}) \text{ for j = 1, 2, 3}$$

Where $\beta_{j,n\alpha^Y}$ is the j^{th} element of $\theta_{n\alpha^Y}$, $\sigma^2_{j,n\alpha^Y}$ is the (j, j) element of $S_{n\alpha^Y}$:

$$\theta_{n\alpha^Y} = (W'_{\alpha^Y}W_{\alpha^Y} + \Sigma^{-1}_{0\alpha^Y}/\phi^2)^{-1}(\Sigma^{-1}_{0\alpha^Y}\mu_{0,\alpha^Y}/\phi^2 + W'_{\alpha^Y}Y)$$

$$S_{n\alpha^Y} = (n+\nu)^{-1}\{\nu\lambda + (Y - W_{\alpha^Y}\theta_{n\alpha^Y})'Y + (\mu_{0\alpha^Y} - \theta_{n\alpha^Y})'\Sigma^{-1}_{0\alpha^Y}\mu_{0\alpha^Y}/\phi^2\}\{(W'_{\alpha^Y}W_{\alpha^Y} + \Sigma^{-1}_{0\alpha^Y}/\phi^2)^{-1}\}$$

Similarly, $P(\gamma|\boldsymbol{\alpha}^{X_1}, \boldsymbol{\alpha}^{X_2}, \boldsymbol{\alpha}^Y, \boldsymbol{\beta}, D) \overset{A7}{=} P(\gamma|\boldsymbol{\alpha}^{X_2}, D)$

$$\gamma|\boldsymbol{\alpha}^{X_2}, D \sim t_{n+v}(\gamma_{j,n\boldsymbol{\alpha}^{X_2}}, \sigma^2_{j,n\boldsymbol{\alpha}^{X_2}})$$

A.2.4 Data Analysis

In this appendix we describe the variables and data sources used in section 2.4. Table A.1 lists all available covariates (C) and their data sources. All variables were averaged over the period 2008-2010. Table A.2 shows the variables that were dropped before beginning our analysis and the reason (missing data or highly correlated (> 0.8) with other covariates). Figure A.2 shows the distribution of covariates included in the analysis. Note that the plots are scaled by covariate. Finally, Table A.3 shows the posterior probability of inclusion in (2.9) for each covariate for BAC-ME and FBMA and whether or not it was included for NLASSO.

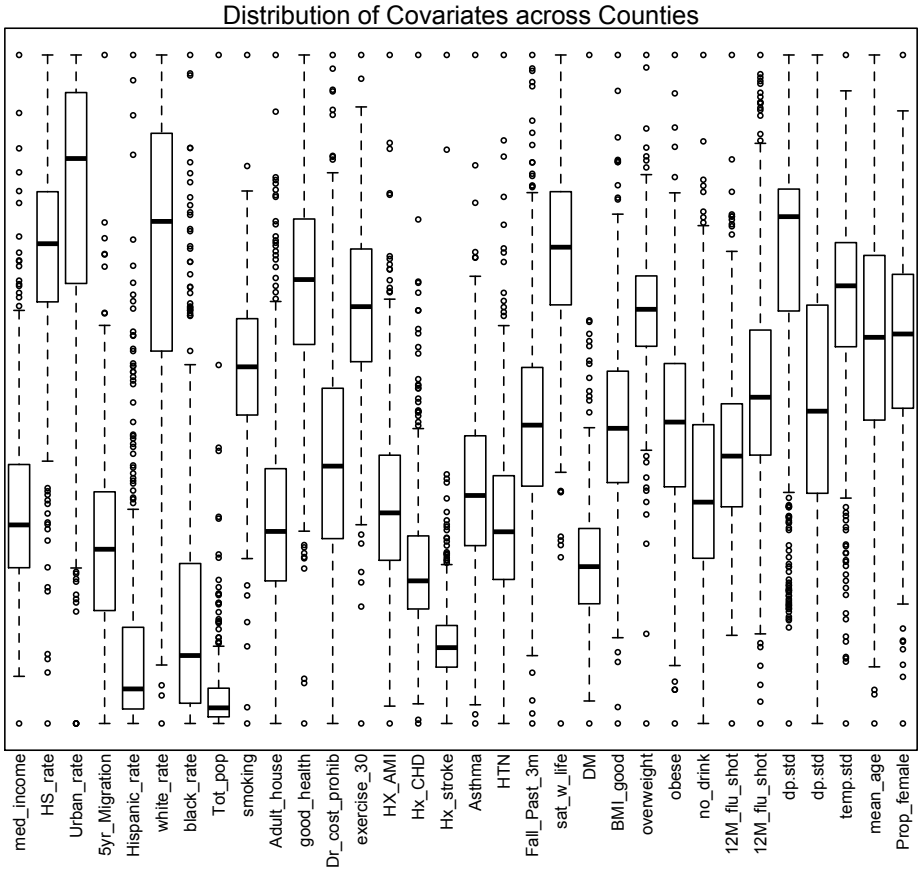


Figure A.2: Distribution of each covariate included in the analysis, by county. Each box plot is on its own scale.

Table A.1: All variables were averaged over the period 2008-2010.

Potential Confounders

Variable	**Source**
Median Income HS grad rate Urban rate 5 year migration rate White rate Black rate Hispanic rate Total Population	County Level Census Data
Smoking rate Mean # of adults in house Proportion in general good health Prop. who did not see dr due to cost Prop exercise in past 30 days Prop with AMI Prop with CHD Prop stroke Prop with asthma Prop alcohol past 30 days Prop with Hypertension Prop who had a fall past 3 mo Prop satisfied with life Prop prior diabetes Prop with good BMI Prop with overweight BMI Prop with obese BMI Prop who don't drink daily Prop with Flu shot 12 mo Prop with Pneumonia shot 12 mo Prop Pre DM Prop Own Home	County level data from the CDC's Behavioral Risk Factor Surveillance System
Mean age Female rate White rate Black rate	Medicare Beneficiary Enrollment Data (Medicare recipients 65 and older)
temp,. . ., temp^7 temp stand dev dew point,. . .,dew point^7 dew point stand dev	County level weather data (*www.ncdc.noaa.gov*, 2012)
Mean NO_2 Mean SO_2 Mean CO Mean Lead	Pollution Data from the EPA's Air Quality System Database (*US EPA*, 2012)
South Midwest Northeast	Geographic Regions as defined by the Census Bureau (*www.census.gov*, 2012)

Table A.2: These variables were eliminated from the data before beginning our analysis.

Variable	Reason
NO_2, CO and SO_2 and Lead	Not enough data points
Black rate and White rate from the medicare data	Highly correlated with race proportions from Census Data
Prop with pneumonia shot 12 mo	Highly correlated with flu shot
Prop alcohol past 30 days	Highly correlated with Prop who don't drink daily
Pre_DM & Own_Home	Too many missing values

Table A.3: Posterior Support by Method. Posterior inclusion probabilities ($P(\boldsymbol{\alpha}^Y|D)$) of each of the 47 potential confounders where the α_m^Y are defined in (2.9) for BAC-ME and FBMA and whether or not a variable was included in (2.9) for NLASSO.

	Variable	BAC-ME	FBMA	NLASSO
1	temp	1	1	1
2	temp.2	1	1	0
3	temp.3	1	1	0
4	temp.4	1	1	0
5	temp.5	1	1	0
6	temp.6	1	1	0
7	temp.7	0.15	0	0
8	dp	1	1	1
9	dp.2	1	1	0
10	dp.3	1	1	0
11	dp.4	1	1	0
12	dp.5	1	1	0
13	dp.6	1	1	0
14	dp.7	0.24	0.74	0
15	temperature_annual_STD	1	1	0
16	Dew_point_annual_STD	1	0	0
17	Median_income	0	0.75	0
18	HS_rate	1	1	0
19	Urban_rate	1	1	0
20	Migration_5_year_rate	0.13	0	0
21	Hispanic_rate	0.75	1	0
22	white_rate	1	1	0
23	black_rate	0.26	0	0
24	Tot_pop	1	1	0
25	smoking	0	0.93	0
26	Mean_adult_in_house	0.06	0.22	0
27	General_Health_good	0	1	0
28	Not_See_Dr_BC_Cost	0	0	0
29	Exercise_Past_30D	0.28	1	0
30	HX_AMI	0.82	0	0
31	Hx_CHD	1	0	1
32	Hx_stroke	0	0.09	0
33	Asthma	0.23	1	0
34	HTN	1	0.21	0
35	Fall_Past_3m	0.11	1	0
36	Satisfaction_with_life_Yes	0.38	0.76	0
37	DM	0.89	0.24	1
38	BMI_good	1	0.94	0
39	BMI_overweight	0	0.94	0
40	BMI_OB	0.19	0.86	0
41	No_Drink_daily	1	0.38	0
42	flu_shot_past_12m	1	1	0
43	mean_age_D	1	0.59	0
44	Female_rate_D	1	1	1
45	Northeast	0.29	0	0
46	Midwest	0.78	0.04	0
47	South	1	0.97	0

A.2.5 Acknowledgements

Support for the research was provided by EPA grants R834894 and RD83479801, NIH grants R21 ES020152, R01 ES019955, R01 ES019560 and R01 ES012054, NCI grant P01 CA134294-02 and HEI grant 4909. The contents of this work are solely the responsibility of the grantee and do not necessarily represent the official views of the USEPA. Further, USEPA does not endorse the purchase of any commercial products or services mentioned in the publication. Many thanks to Chi Wang for his guidance on the technical aspects of his implementation of BAC and to Matt Cefalu for his thoughtful discussion and insight.

A.3 Propensity Score Methods for Combining Data Sources - Appendices

A.3.1 MCMC Details for Sequential Bayesian Approach

We develop the MCMC algorithm for binary treatment and outcomes, assuming a probit link function and using a latent variable approach (*Albert and Chib*, 1993). Further, we adjust for Z in the outcome model as a linear covariate. That is $\boldsymbol{h}\{Z\} = Z$. We specify a linear regression model for $E[Z|C]$. In other words, $Z_i|C_i \sim N(\boldsymbol{C}_i\boldsymbol{\eta}, \sigma^2)$ for $i = 1, \ldots, n$.

A.3.1.1 Models

$$
\begin{aligned}
g(P(X_i = 1|\boldsymbol{C}_i, Z_i)) &= \boldsymbol{C}_j\boldsymbol{\gamma} + Z_i && i = 1, \ldots, n \\
g(P(X_i = 1|\boldsymbol{C}_i, \boldsymbol{U}_i)) &= \boldsymbol{C}_i\boldsymbol{\gamma} + \boldsymbol{U}_i\tilde{\gamma} && i = n+1, \ldots, n+m \\
E[Z_i|\boldsymbol{C}_i] &= \boldsymbol{C_i\eta} && i = 1, \ldots, n
\end{aligned}
$$

$$
\begin{aligned}
g(P(Y_i = 1|X_i, \boldsymbol{C}_i, Z_i)) &= \beta X_i + \boldsymbol{C}_i\boldsymbol{\xi} + \boldsymbol{h}\{Z_i\}\tilde{\boldsymbol{\xi}} && i = 1, \ldots, n \\
g(P(Y_i = 1|X_i, \boldsymbol{C}_i, Z_i)) &= \beta X_i + \boldsymbol{C}_i\boldsymbol{\xi} + \boldsymbol{h}\{U_i\tilde{\gamma}\}\tilde{\boldsymbol{\xi}} && i = n+1, \ldots, n+m
\end{aligned}
$$

A.3.1.2 Prior Distributions

Let $\boldsymbol{\theta} = (\beta, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}})$, q be the number of fully measured confounders plus 1 and p be the number of partially measured confounders (recall that the first column of $\boldsymbol{C}$ is the intercept). We factorize the prior distribution $P(\boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}, \boldsymbol{\theta}, \boldsymbol{\eta}, \sigma^2)$ as $P(\boldsymbol{\gamma})P(\tilde{\boldsymbol{\gamma}})P(\boldsymbol{\theta})P(\boldsymbol{\eta}|\sigma^2)P(\sigma^2)$.

We assume the following prior distributions:

- $\tilde{\boldsymbol{\gamma}} \sim N_p(\mathbf{0}, \lambda_{\tilde{\boldsymbol{\gamma}}}\boldsymbol{I}_)$
- $\boldsymbol{\gamma} \sim N_q(\mathbf{0}, \lambda_{\boldsymbol{\gamma}}\boldsymbol{I}_q)$
- $\boldsymbol{\theta} \sim N_{q+2}(\mathbf{0}, \lambda_{\boldsymbol{\theta}}\boldsymbol{I}_{q+2})$
- $\sigma^2 \sim IG(a_0, b_0)$
- $\boldsymbol{\eta}|\sigma^2 \sim N_q(\mathbf{0}, \sigma^2 k\boldsymbol{I}_q)$

and let $\lambda_{\boldsymbol{\gamma}} = \lambda_{\tilde{\boldsymbol{\gamma}}} = \lambda_{\boldsymbol{\theta}} = 1000$, $k = 10,000$, $a_0 = 20.1$ and $b_0 = 2$.

A.3.1.3 Posterior Simulation

We iteratively sample from $P(Z_{prim}, \boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}|X^*, \boldsymbol{C}, \boldsymbol{U}_{sup})$ then
$P(\boldsymbol{\theta}|Y^*, X, \boldsymbol{C}, \boldsymbol{U}_{sup}, \tilde{\boldsymbol{\gamma}}, Z_{prim})$. Note that Y^*, X^* are the underlying latent variables from the probit regression models. From the priors specified above and (3.1) we have

$$
\begin{aligned}
&P(Z_{prim}, \boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}, \boldsymbol{\eta}, \sigma^2 | X^*, \boldsymbol{C}, \boldsymbol{U}_{sup}) \\
&\quad \propto P(X^*_{sup} | \boldsymbol{C}_{sup}, \boldsymbol{U}_{sup}, \boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}) P(X^*_{prim} | \boldsymbol{C}_{prim}, Z_{prim}, \boldsymbol{\gamma}) P(Z_{prim} | \boldsymbol{C}_{prim}, \boldsymbol{\eta}, \sigma^2) \\
&\quad \times P(\boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}, \boldsymbol{\eta}, \sigma^2) \\
&\quad \propto exp\{-\frac{1}{2}(X^*_{sup} - (\boldsymbol{C}_{sup}\boldsymbol{\gamma} + \boldsymbol{U}_{sup}\tilde{\boldsymbol{\gamma}}))'(X^*_{sup} - (\boldsymbol{C}_{sup}\boldsymbol{\gamma} + \boldsymbol{U}_{sup}\tilde{\boldsymbol{\gamma}}))\} \\
&\quad \times exp\{-\frac{1}{2}(X^*_{prim} - (\boldsymbol{C}_{prim}\boldsymbol{\gamma} + Z_{prim}))'(X^*_{prim} - (\boldsymbol{C}_{prim}\boldsymbol{\gamma} + Z_{prim}))\} \\
&\quad \times exp\{-\frac{1}{2}(Z_{prim} - \boldsymbol{C}_{prim}\boldsymbol{\eta})'(\sigma^2 \boldsymbol{I}_n)^{-1}(Z_{prim} - \boldsymbol{C}_{prim}\boldsymbol{\eta}))\} exp\{-\frac{1}{2}(\boldsymbol{\gamma}'(\lambda_{\boldsymbol{\gamma}} \boldsymbol{I}_q)^{-1}\boldsymbol{\gamma}\} \\
&\quad \times exp\{-\frac{1}{2}(\tilde{\boldsymbol{\gamma}}'(\lambda_{\tilde{\boldsymbol{\gamma}}} \boldsymbol{I}_q)^{-1}\tilde{\boldsymbol{\gamma}}\}(\sigma^2)^{-q/2} exp\{-\frac{1}{2}\boldsymbol{\eta}'(\sigma^2 k \boldsymbol{I}_q)^{-1}\boldsymbol{\eta}\}(\sigma^2)^{-(a_0+1)} \\
&\quad \times exp\{\frac{-b_0}{\sigma^2}\}
\end{aligned}
$$

and

$$
\begin{aligned}
&P(\boldsymbol{\theta} | Y^*, X, \boldsymbol{C}, \boldsymbol{U}_{sup}, Z_{prim}, \tilde{\boldsymbol{\gamma}}) \\
&\quad \propto P(Y^*_{sup} | X_{sup}, \boldsymbol{C}_{sup}, \boldsymbol{U}_{sup}, \tilde{\boldsymbol{\gamma}}, \boldsymbol{\theta}) P(Y^*_{prim} | X_{prim}, \boldsymbol{C}_{prim}, Z_{prim}, \boldsymbol{\theta}) P(\boldsymbol{\theta}) \\
&\quad \propto exp\{-\frac{1}{2}(Y^*_{sup} - (\beta X_{sup} + \boldsymbol{C}_{sup}\boldsymbol{\xi} + (\boldsymbol{U}_{sup}\tilde{\boldsymbol{\gamma}})\tilde{\boldsymbol{\xi}}))'(Y^*_{sup} - (\beta X_{sup} + \boldsymbol{C}_{sup}\boldsymbol{\xi} + (\boldsymbol{U}_{sup}\tilde{\boldsymbol{\gamma}})\tilde{\boldsymbol{\xi}}))\} \\
&\quad \times exp\{-\frac{1}{2}(Y^*_{prim} - (\beta X_{prim} + \boldsymbol{C}_{prim}\boldsymbol{\xi} + Z_{prim}\tilde{\boldsymbol{\xi}}))'(Y^*_{prim} - (\beta X_{prim} + \boldsymbol{C}_{prim}\boldsymbol{\xi} + Z_{prim}\tilde{\boldsymbol{\xi}})))\} \\
&\quad \times exp\{-\frac{1}{2}\boldsymbol{\theta}'(\lambda_{\boldsymbol{\theta}} \boldsymbol{I}_{q+2})^{-1}\boldsymbol{\theta}\}
\end{aligned}
$$

A.3.1.4 MCMC Algorithm

Let $Z^{(t)} = \begin{pmatrix} \boldsymbol{U}_{sup}\tilde{\boldsymbol{\gamma}}^{(t)} \\ Z^{(t)}_{prim} \end{pmatrix}$ and $\boldsymbol{W}^{(t)} = (X, \boldsymbol{C}, Z^{(t)})$. Define $V_{\tilde{\boldsymbol{\gamma}}} = \boldsymbol{U}_{sup}(\boldsymbol{U}'_{sup}\boldsymbol{U}_{sup} + \frac{1}{\lambda_{\tilde{\boldsymbol{\gamma}}}}\boldsymbol{I}_p)^{-1}\boldsymbol{U}'_{sup}$ and $V_{\boldsymbol{\gamma} sup} = \boldsymbol{C}_{sup}(\boldsymbol{C}'_{sup}\boldsymbol{C}_{sup} + \frac{1}{\lambda_{\boldsymbol{\gamma}}}\boldsymbol{I}_q)^{-1}\boldsymbol{C}'_{sup}$.

From the posterior distributions in A.3.1.3, it is relatively straightforward to calculate the marginal and conditional distributions that follow. The MCMC algorithm for iteration (t+1):

1. Draw $X_{sup}^{*(t+1)}$ from a truncated normal distribution with mean $\boldsymbol{C}_{sup}\boldsymbol{\gamma}^{(t)} + \boldsymbol{U}_{sup}\tilde{\boldsymbol{\gamma}}^{(t)}$ and variance 1 as described in *Albert and Chib* (1993).

2. Draw $X_{prim}^{*(t+1)}$ from a truncated normal distribution with mean $\boldsymbol{C}_{prim}\boldsymbol{\gamma}^{(t)} + Z_{prim}^{(t)}$ and variance 1.

3. Draw $Z_{prim}^{(t+1)}$ from $N\left(\frac{\sigma^{2(t)}}{1+\sigma^{2(t)}}(X_{prim}^{*(t+1)} - \boldsymbol{C}_{prim}\boldsymbol{\gamma}^{(t)} + \frac{1}{\sigma^{2(t)}}\boldsymbol{C}_{prim}\boldsymbol{\eta}^{(t)}), \frac{\sigma^{2(t)}}{1+\sigma^{2(t)}}\right)$

4. Draw $\sigma^{2(t+1)}$ from $IG(a_0 + \frac{m+n}{2}, b_0 + 1/2(Z^{(t+1)\prime}Z^{(t+1)} - Z^{(t+1)\prime}\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{C} + \frac{1}{k}\boldsymbol{I}_q)^{-1}\boldsymbol{C}'Z^{(t+1)}))$

5. Draw $\boldsymbol{\eta}^{(t+1)}$ from $N_q((\boldsymbol{C}'\boldsymbol{C} + \frac{1}{k}\boldsymbol{I}_q)^{-1}\boldsymbol{C}'Z^{(t+1)}, \sigma^{2(t+1)}(\boldsymbol{C}'\boldsymbol{C} + \frac{1}{k}\boldsymbol{I}_q)^{-1})$

6. Draw $\boldsymbol{\gamma}^{(t+1)}$ from $N_q((\boldsymbol{C}'_{prim}\boldsymbol{C}_{prim} + \boldsymbol{C}'_{sup}(\boldsymbol{I}_m - V_{\tilde{\boldsymbol{\gamma}}})\boldsymbol{C}_{sup} + \frac{1}{\lambda_{\boldsymbol{\gamma}}}\boldsymbol{I}_q)^{-1}(\boldsymbol{C}'_{prim}(X_{prim}^{*(t+1)} - Z_{prim}^{(t+1)}) + \boldsymbol{C}'_{sup}(\boldsymbol{I}_m - V_{\tilde{\boldsymbol{\gamma}}})X_{sup}^{*(t+1)}), (\boldsymbol{C}'_{prim}\boldsymbol{C}_{prim} + \boldsymbol{C}'_{sup}(\boldsymbol{I}_m - V_{\tilde{\boldsymbol{\gamma}}})\boldsymbol{C}_{sup} + \frac{1}{\lambda_{\boldsymbol{\gamma}}}\boldsymbol{I}_q)^{-1})$

7. Draw $\tilde{\boldsymbol{\gamma}}^{(t+1)}$ from $N_p((\boldsymbol{U}'_{sup}(\boldsymbol{I}_m - V_{\boldsymbol{\gamma}J})\boldsymbol{U}_{sup} + \frac{1}{\lambda_{\tilde{\boldsymbol{\gamma}}}}\boldsymbol{I}_p)^{-1}(\boldsymbol{U}'_{sup}(\boldsymbol{I}_m - V_{\boldsymbol{\gamma}J})X_{sup}^{*(t+1)}), (\boldsymbol{U}'_{sup}(\boldsymbol{I}_m - V_{\boldsymbol{\gamma}J})\boldsymbol{U}_{sup} + \frac{1}{\lambda_{\tilde{\boldsymbol{\gamma}}}}\boldsymbol{I}_p)^{-1})$

8. Draw $Y^{*(t+1)}$ from a truncated normal distribution with mean $\boldsymbol{W}^{(t)}\boldsymbol{\theta}^{(t+1)}$ and variance 1.

9. Draw $\boldsymbol{\theta}^{(t+1)}$ from $N_{q+2}((\boldsymbol{W}^{(t+1)\prime}\boldsymbol{W}^{(t+1)} + \frac{1}{\lambda_\theta}\boldsymbol{I}_{q+2})^{-1}\boldsymbol{W}^{(t+1)\prime}Y^{*(t+1)}, (\boldsymbol{W}^{(t+1)\prime}\boldsymbol{W}^{(t+1)} + \frac{1}{\lambda_\theta}\boldsymbol{I}_{q+2})^{-1})$

A.3.2 MCMC Details for Two-Stage Approach

We again develop the MCMC algorithm for binary treatment and outcomes, assuming a probit link function and using a latent variable approach (*Albert and Chib*, 1993). Further, we adjust for Z in the outcome model as a linear covariate. That is $\boldsymbol{h}\{Z\} = Z$. We specify a linear regression model for $E[Z|C]$. In other words, $Z_i|\boldsymbol{C}_i \sim N(\boldsymbol{C}_i\boldsymbol{\eta}, \sigma^2)$ for $i = 1, \ldots, n$. Let $\boldsymbol{\theta}_X = (\boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}})$, $\boldsymbol{\theta}_Y = (\beta, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}})$, $\boldsymbol{W}_X = (\boldsymbol{C}, \boldsymbol{U})$, $\boldsymbol{W}_Y = (\boldsymbol{C}, Z)$ and $Z = \begin{pmatrix} \hat{Z}_{sup} \\ Z_{prim} \end{pmatrix}$.

A.3.2.1 Models

1. *Stage 1*

$$g(P(X_i = 1|\boldsymbol{C}_i, \boldsymbol{U}_i)) = \boldsymbol{C}_i\boldsymbol{\gamma} + U_i\tilde{\boldsymbol{\gamma}} \qquad i = n+1, \ldots, n+m$$

2. *Stage 2*

$$E[Z_i] = \boldsymbol{C_i\eta} \qquad i = 1, \ldots, n$$
$$g(P(Y_i = 1|X_i, \boldsymbol{C}_i, Z_i)) = \beta X_i + \boldsymbol{C}_i\boldsymbol{\xi} + \boldsymbol{h\{}Z_i\boldsymbol{\}}\tilde{\boldsymbol{\xi}} \qquad i = 1, \ldots, n+m$$

A.3.2.2 Prior Distributions

We factorize the prior distribution $P(\boldsymbol{\theta}_X, \boldsymbol{\theta}_Y, \boldsymbol{\eta}, \sigma^2)$ as $P(\boldsymbol{\theta}_X)P(\boldsymbol{\theta}_Y)P(\boldsymbol{\eta}|\sigma^2)\ P(\sigma^2)$.

We assume the following prior distributions:

- $\boldsymbol{\theta}_X \sim N_p(\mathbf{0}, \lambda_X \boldsymbol{W}_X'\boldsymbol{W}_X)$
- $\boldsymbol{\theta}_Y \sim N_{q+2}(\mathbf{0}, \lambda_Y \boldsymbol{I}_{q+2})$
- $\sigma^2 \sim IG(a_0, b_0)$
- $\boldsymbol{\eta}|\sigma^2 \sim N_q(\mathbf{0}, \sigma^2 k \boldsymbol{I}_q)$

and let $\lambda_X = \lambda_Y = 1000$, $k = 10,000$, $a_0 = 20.1$ and $b_0 = 2$.

A.3.2.3 Posterior Simulation

A.3.2.3.1 Stage 1 In stage 1, we sample from $P(\boldsymbol{\theta}_X|X^*_{sup}, \boldsymbol{W}_{X,sup})$. From the priors specified above and (3.1) we have

$$\begin{aligned} P(\boldsymbol{\theta}_X|X^*_{sup}, \boldsymbol{W}_{X,sup}) \propto & P(X^*_{sup}|\boldsymbol{W}_{X,sup}, \boldsymbol{\theta}_X)P(\boldsymbol{\theta}_X) \\ \propto & exp\{-\frac{1}{2}(X^*_{sup} - \boldsymbol{W}_{X,sup}\boldsymbol{\theta}_X)'(X^*_{sup} - \boldsymbol{W}_{X,sup}\boldsymbol{\theta}_X)\} \\ & exp\{-\frac{1}{2}(\boldsymbol{\theta}'_X \frac{1}{\lambda_X}\boldsymbol{W}'_{X,sup}\boldsymbol{W}_{X,sup}\boldsymbol{\theta}_X)\} \end{aligned}$$

A.3.2.3.2 Stage 2 In stage 2, we sample from $P(Z_{prim}, \boldsymbol{\theta}_Y, \boldsymbol{\eta}, \sigma^2 | Y^*, X, \boldsymbol{W}_Y)$. From the priors specified above and (3.1) we have

$$\begin{aligned} P(Z_{prim}, \boldsymbol{\theta}_Y, \boldsymbol{\eta}, \sigma^2 | Y^*, X, \boldsymbol{C}, \hat{Z}_{sup}) &\propto P(Y^* | X, \boldsymbol{C}, Z, \boldsymbol{\theta}_Y) P(Z_{prim} | \boldsymbol{C}_{prim}, \boldsymbol{\eta}, \sigma^2) P(\boldsymbol{\theta}_Y) P(\boldsymbol{\eta} | \sigma^2) P(\sigma^2) \\ &\propto exp\{-\frac{1}{2}(Y^* - \boldsymbol{W}_Y \boldsymbol{\theta}_Y)'(Y^* - \boldsymbol{W}_Y \boldsymbol{\theta}_Y)\} \\ &\times exp\{-\frac{1}{2}(Z_{prim} - \boldsymbol{C}_{prim}\boldsymbol{\eta})'(\sigma^2 \boldsymbol{I}_{m+n})^{-1}(Z_{prim} - \boldsymbol{C}_{prim}\boldsymbol{\eta})\} \\ &\times exp\{-\frac{1}{2}\boldsymbol{\theta}'(\lambda_Y \boldsymbol{I}_{q+2})^{-1}\boldsymbol{\theta}\}(\sigma^2)^{-q/2} exp\{-\frac{1}{2}\boldsymbol{\eta}'(\sigma^2 k \boldsymbol{I}_q)^{-1}\boldsymbol{\eta}\}(\sigma^2)^{-(a_0+1)} exp\{-\frac{b_0}{\sigma^2}\} \end{aligned}$$

where $\hat{Z}_{sup} = E[\boldsymbol{U}_{sup}\tilde{\boldsymbol{\gamma}}]$, the posterior mean of $\boldsymbol{U}_j\tilde{\boldsymbol{\gamma}}$ from stage 1.

A.3.2.4 MCMC Algorithm

A.3.2.4.1 Stage 1

1. Draw $\boldsymbol{\theta}_X^{(t+1)}$ from $N_{q+p}(((1+\frac{1}{\lambda_X})\boldsymbol{W}'_{X,sup}\boldsymbol{W}_{X,sup})^{-1}\boldsymbol{W}'_{X,sup}X^{*(t)}, ((1+\frac{1}{\lambda_X})\boldsymbol{W}'_{X,sup}\boldsymbol{W}_{X,sup})^{-1})$
2. Draw $X_{sup}^{*(t+1)}$ from a truncated normal distribution with mean $\boldsymbol{W}_{X,sup}\boldsymbol{\theta}_X^{(t+1)}$ and variance 1 as described in *Albert and Chib* (1993).

A.3.2.4.2 Stage 2

1. Draw $Y^{*(t+1)}$ from a truncated normal distribution with mean $\boldsymbol{W}_Y^{(t)}\boldsymbol{\theta}_Y^{(t)}$ and variance 1
2. Draw $Z_{prim}^{(t+1)}$ from $N(\left(\frac{1}{\sigma^{2(t)}} + \tilde{\boldsymbol{\xi}}^{(t)2}\right)^{-1}\left((Y_{prim}^{*(t+1)} - \beta^{(t)}X_{prim} - \boldsymbol{C}_{prim}\boldsymbol{\xi}^{(t)})\tilde{\boldsymbol{\xi}}^{(t)2} + \frac{1}{\sigma^{2(t)}}\boldsymbol{C}_{prim}\boldsymbol{\eta}^{(t)}\right), \left(\frac{1}{\sigma^{2(t)}} + \tilde{\boldsymbol{\xi}}^{(t)2}\right)^{-1})$
3. Draw $\sigma^{2(t+1)}$ from $IG(a_0 + \frac{m+n}{2}, b_0 + 1/2(Z^{(t+1)\prime}Z^{(t+1)} - Z^{(t+1)\prime}\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{C} + \frac{1}{k}\boldsymbol{I}_q)^{-1}\boldsymbol{C}'Z^{(t+1)}))$
4. Draw $\boldsymbol{\eta}^{(t+1)}$ from $N_q((\boldsymbol{C}'\boldsymbol{C} + \frac{1}{k}\boldsymbol{I}_q)^{-1}\boldsymbol{C}'Z^{(t+1)}, \sigma^{2(t+1)}(\boldsymbol{C}'\boldsymbol{C} + \frac{1}{k}\boldsymbol{I}_q)^{-1})$
5. Draw $\boldsymbol{\theta}_Y^{(t+1)}$ from $N_{q+2}((\boldsymbol{W}^{(t+1)\prime}\boldsymbol{W}^{(t+1)} + \frac{1}{\lambda_Y}\boldsymbol{I}_{q+2})^{-1}\boldsymbol{W}^{(t+1)\prime}Y^{*(t+1)}, (\boldsymbol{W}^{(t+1)\prime}\boldsymbol{W}^{(t+1)} + \frac{1}{\lambda_Y}\boldsymbol{I}_{q+2})^{-1})$

A.3.3 Data Generating Mechanism for Simulations

The specific data generating mechanism for the simulation scenarios is outlined below.

1. Generate $\boldsymbol{C}_{prim,-1}$ from $N_q(\boldsymbol{\mu}_{prim}, \Sigma_C)$ and $\boldsymbol{C}_{sup,-1}$ from $N_q(\boldsymbol{\mu}_{sup}, \Sigma_C)$
2. Generate $\boldsymbol{U}_1|\boldsymbol{C}$ from $N_{p1}(\boldsymbol{C}\boldsymbol{\zeta}_l, \tau_l \boldsymbol{I}_{p1})$
3. Generate $\boldsymbol{U}_2$ from $N_{p2}(\boldsymbol{0}, \Sigma_U)$
4. Generate X from Bin$(n+m, px)$ where $px = \Phi((\boldsymbol{C}, \boldsymbol{U}) \begin{pmatrix} \boldsymbol{\gamma}_k \\ \tilde{\boldsymbol{\gamma}}_k \end{pmatrix})$ and Φ is the cumulative distribution function of the standard normal distribution.
5. Generate Y from Bin$(n+m, py)$ where $py = \Phi((X, \boldsymbol{C}, \boldsymbol{U}) \begin{pmatrix} \beta_k \\ \boldsymbol{\xi}_k \\ \tilde{\boldsymbol{\xi}}_k \end{pmatrix})$ and Φ is the cumulative distribution function of the standard normal distribution.

Recall that the first column of $\boldsymbol{C}$ is the intercept. $\boldsymbol{\mu}_{sup} = \boldsymbol{0}$ and $\boldsymbol{\mu}_{prim} = \boldsymbol{1}$. $\Sigma_{\boldsymbol{C}}$ and $\Sigma_{\boldsymbol{U}}$ have auto-regressive correlation structures (AR1) with $\rho = 0.3$. l corresponds to the correlation between $\boldsymbol{C}$ and $\boldsymbol{U}$ – low ($l = 1$) or moderate($l = 2$) – and k corresponds to the relative 'importance' of $\boldsymbol{C}$ and $\boldsymbol{U}$ as confounders of the effect of X on Y – $\boldsymbol{C} = \boldsymbol{U}$ ($k = 1$), $\boldsymbol{C} < \boldsymbol{U}$ ($k = 2$) or $\boldsymbol{C} > \boldsymbol{U}$ ($k = 3$). $p1$ = 4 for all scenarios. Specific values of $\tau, \boldsymbol{\zeta}, \boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}, \beta, \boldsymbol{\xi}$ and $\tilde{\boldsymbol{\xi}}$ are as follows.

$$\tau_1 = 1$$

$$\tau_2 = 3$$

$$\boldsymbol{\zeta}_1 = \begin{pmatrix} 0.03 & 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\ 0.03 & 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\ 0.03 & 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\ 0.03 & 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \end{pmatrix}$$

$$\boldsymbol{\zeta}_2 = \begin{pmatrix} 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.4 & 0.4 & 0.4 & 0.4 & 0.4 & 0.4 \end{pmatrix}$$

$$\boldsymbol{\gamma}_1 = \begin{pmatrix} -1 & 0.2 & 0.2 & 0.1 & 0.1 & 0.05 & 0.05 \end{pmatrix}$$

$$\boldsymbol{\gamma}_2 = \begin{pmatrix} -1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.05 & 0.05 \end{pmatrix}$$

$$\boldsymbol{\gamma}_3 = \begin{pmatrix} -1 & 0.4 & 0.4 & 0.1 & 0.1 & 0.05 & 0.05 \end{pmatrix}$$

$$\tilde{\boldsymbol{\gamma}}_1 = \begin{pmatrix} 0.2 & 0.1 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.2 \end{pmatrix}$$

$$\tilde{\boldsymbol{\gamma}}_2 = \begin{pmatrix} 0.4 & 0.2 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.4 \end{pmatrix}$$

$$\tilde{\boldsymbol{\gamma}}_3 = \begin{pmatrix} 0.1 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.1 \end{pmatrix}$$

$$\boldsymbol{\xi}_1 = \begin{pmatrix} -2 & 0.2 & 0.05 & 0.1 & 0.1 & 0.2 & 0.05 \end{pmatrix}$$

$$\boldsymbol{\xi}_2 = \begin{pmatrix} -2 & 0.1 & 0.05 & 0.1 & 0.05 & 0.1 & 0.05 \end{pmatrix}$$

$$\boldsymbol{\xi}_3 = \begin{pmatrix} -2 & 0.4 & 0.05 & 0.1 & 0.1 & 0.4 & 0.05 \end{pmatrix}$$

$$\tilde{\boldsymbol{\xi}}_1 = \begin{pmatrix} 0.05 & 0.1 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.2 & 0.2 \end{pmatrix}$$

$$\tilde{\boldsymbol{\xi}}_2 = \begin{pmatrix} 0.05 & 0.2 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.4 & 0.4 \end{pmatrix}$$

$$\tilde{\boldsymbol{\xi}}_3 = \begin{pmatrix} 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.1 & 0.1 \end{pmatrix}$$

A.3.4 Acknowledgements

Support for the research was provided by EPA grants R834894 and RD83479801, NIH grants R21 ES020152, R01 ES019955, R01 ES019560 and R01 ES012054, NCI grant P01 CA134294-02 and HEI grant 4909. The contents of this work are solely the responsibility of the grantee and do not necessarily represent the official views of the USEPA. Further, USEPA does not endorse the purchase of any commercial products or services mentioned in the publication. Many thanks to Brent Coull and Matt Cefalu for their thoughtful discussion and insight.

References

Albert, J. H., and S. Chib (1993), Bayesian analysis of binary and polychotomous response data, *Journal of the American Statistical Association*, *88*(422), 669–679, doi:10.2307/2290350, ArticleType: research-article / Full publication date: Jun., 1993 / Copyright 1993 American Statistical Association.

Bang, H., and J. M. Robins (2005), Doubly robust estimation in missing data and causal inference models, *Biometrics*, *61*(4), 962973, doi:10.1111/j.1541-0420.2005.00377.x.

Dominici, F., R. D. Peng, C. D. Barr, and M. L. Bell (2010), Protecting human health from air pollution, *Epidemiology*, *21*(2), 187–194, doi:10.1097/EDE.0b013e3181cc86e8.

Gelman, A., and J. Hill (2007), *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press.

Greenland, S. (1993), Methods for epidemiologic analyses of multiple exposures: A review and comparative study of maximum-likelihood, preliminary-testing, and empirical-bayes regression, *Statistics in Medicine*, *12*(8), 717736, doi:10.1002/sim.4780120802.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999), Bayesian model averaging: A tutorial, *Statistical Science*, *14*(4), 382–401.

Lichtman, J. H., N. B. Allen, Y. Wang, E. Watanabe, S. B. Jones, and L. B. Goldstein (2009), Stroke patient outcomes in US hospitals before the start of the joint commission primary stroke center certification program, *Stroke*, *40*(11), 3574–3579, doi:10.1161/STROKEAHA.109.561472.

Little, R. (2011), Calibrated bayes, for statistics in general, and missing data in particular, *Statistical Science*, *26*(2), 162–174, doi:10.1214/10-STS318, mathematical Reviews number (MathSciNet): MR2858391; Zentralblatt MATH identifier: 06075150.

Little, R. J. A., and D. B. Rubin (2002), *Statistical Analysis with Missing Data, Second Edition*, 2 ed., Wiley-Interscience.

Madigan, D., J. York, and D. Allard (1995), Bayesian graphical models for discrete data, *International Statistical Review / Revue Internationale de Statistique*, *63*(2), 215–232, doi:10.2307/1403615, ArticleType: research-article / Full publication date: Aug., 1995 / Copyright 1995 International Statistical Institute (ISI).

McCandless, L. C. (2012), Discussion of adjustment uncertainty and propensity scores, *Biometrics*, *68*(3), 678680, doi:10.1111/j.1541-0420.2011.01733.x.

McCandless, L. C., P. Gustafson, and P. C. Austin (2009), Bayesian propensity score analysis for observational data, *Statistics in Medicine*, *28*(1), 94–112, doi:10.1002/sim.3460, PMID: 19012268.

McCandless, L. C., I. J. Douglas, S. J. Evans, and L. Smeeth (2010), Cutting feedback in bayesian regression adjustment for the propensity score, *The International Journal of Biostatistics*, *6*(2), doi:10.2202/1557-4679.1205.

McCandless, L. C., S. Richardson, and N. Best (2012), Adjustment for missing confounders using external validation data and propensity scores, *Journal of the American Statistical Association*, *107*(497), 40–51, doi:10.1080/01621459.2011.643739.

NCI (2013), SEER, http://seer.cancer.gov/.

Pope, G., J. Kautter, R. Ellis, A. Ash, J. Ayanian, L. iezzoni, M. Ingber, J. Levy, and J. Robst (2004), Risk adjustment of medicare capitation payments using the CMS-HCC model, *Quantitative Health Sciences Publications and Presentations*.

Raftery, A. E. (1995), Bayesian model selection in social research, *Sociological Methodology*, *25*, 111, doi:10.2307/271063.

Raftery, A. E., D. Madigan, and J. A. Hoeting (1997), Bayesian model averaging for linear regression models, *Journal of the American Statistical Association*, *92*(437), 179–191.

Robins, J. M., S. D. Mark, and W. K. Newey (1992), Estimating exposure effects by modelling the expectation of exposure conditional on confounders, *Biometrics*, *48*(2), 479–495, doi:10.2307/2532304, ArticleType: research-article / Full publication date: Jun., 1992 / Copyright 1992 International Biometric Society.

Rosenbaum, P. R., and D. B. Rubin (1983), The central role of the propensity score in observational studies for causal effects, *Biometrika*, *70*(1), 41–55, doi:10.1093/biomet/70.1.41.

Rubin, D. (1985), The use of propensity scores in applied bayesian inference, in *Bayesian Statistics*, vol. 2, pp. 463–472, Elsevier Science Publishers and Valencia University Press.

Rubin, D. B. (2007), The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials, *Statistics in Medicine*, *26*(1), 20–36, doi:10.1002/sim.2739, PMID: 17072897.

Rubin, D. B. (2008), For objective causal inference, design trumps analysis, *The Annals of Applied Statistics*, *2*(3), 808–840, ArticleType: research-article / Full publication date: Sep., 2008 / Copyright 2008 Institute of Mathematical Statistics.

Schneeweiss, S., J. A. Rassen, R. J. Glynn, J. Avorn, H. Mogun, and M. A. Brookhart (2009), High-dimensional propensity score adjustment in studies of treatment effects using health care claims data, *Epidemiology*, *20*(4), 512–522, doi:10.1097/EDE.0b013e3181a663cc.

Stuart, E. A. (2010), Matching methods for causal inference: A review and a look forward, *Statistical Science*, *25*(1), 1–21, doi:10.1214/09-STS313.

Sturmer, T., S. Schneeweiss, J. Avorn, and R. J. Glynn (2005), Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration, *American journal of epidemiology*, *162*(3), 279289.

Sturmer, T., S. Schneeweiss, K. J. Rothman, J. Avorn, and R. J. Glynn (2007), Performance of propensity score calibrationa simulation study, *American journal of epidemiology*, *165*(10), 11101118.

US EPA, O. (2012), AQS data for downloading, TTN AIRS AQS, US EPA, http://www.epa.gov/ttn/airs/airsaqs/detaildata/downloadaqsdata.htm, AQS data available for downloading.

Vansteelandt, S. (2012), Discussions, *Biometrics*, *68*(3), 675678, doi:10.1111/j.1541-0420.2011.01734.x.

Vedal, S., and J. D. Kaufman (2011), What does multi-pollutant air pollution research mean?, *American Journal of Respiratory and Critical Care Medicine*, *183*(1), 4–6, doi:10.1164/rccm.201009-1520ED.

Wang, C., G. Parmigiani, and F. Dominici (2012), Bayesian effect estimation accounting for adjustment uncertainty, *Biometrics*, *68*(3), 661671, doi:10.1111/j.1541-0420.2011.01731.x.

www.census.gov (2012), Reference maps - geography - U.S. census bureau, https://www.census.gov/geo/maps-data/maps/reference.html.

www.ncdc.noaa.gov (2012), National climatic data center (NCDC) | the world's largest active archive of weather and climate data producing and supplying data and publications for the world., http://www.ncdc.noaa.gov/.

Zigler, C. M., and F. Dominici (2013), Uncertainty in propensity score estimation: Bayesian methods for variable selection and model averaged causal effects, *Journal of the American Statistical Association*.

Zigler, C. M., K. Watts, R. W. Yeh, Y. Wang, B. A. Coull, and F. Dominici (2013), Model feedback in bayesian propensity score estimation, *Biometrics*.

Zou, H. (2006), The adaptive lasso and its oracle properties, *Journal of the American Statistical Association*, *101*(476), 1418–1429, doi:10.1198/016214506000000735.