

# Hierarchical Priors for Bias Parameters in Bayesian Adjustment for Unmeasured Confounding

Lawrence C. McCandless<sup>1</sup>

Paul Gustafson<sup>2</sup>

Adrian R. Levy<sup>3</sup>

Sylvia Richardson<sup>4</sup>

<sup>1</sup> Faculty of Health Sciences, Simon Fraser University, Canada. <sup>2</sup> Department of Statistics, University of British Columbia. <sup>3</sup> School of Population and Public Health, University of British Columbia. <sup>4</sup> Department of Epidemiology and Public Health, Imperial College London.

June 2009

Corresponding author:

Lawrence McCandless  
Assistant Professor of Biostatistics  
Faculty of Health Sciences  
Simon Fraser University  
8888 University Drive  
Burnaby BC V5A 1S6  
Canada  
mccandless@sfu.ca  
Tel: 778-782-8651  
[www.fhs.sfu.ca/portal\\_memberdata/lmccandless](http://www.fhs.sfu.ca/portal_memberdata/lmccandless)

## Abstract

Recent years have witnessed new innovation in Bayesian techniques to adjust for unmeasured confounding. A challenge with existing methods is that the user is often required to elicit prior distributions for high dimensional parameters that model competing bias scenarios. This can render the methods unwieldy. In this paper we propose a novel methodology to adjust for unmeasured confounding that derives default priors for bias parameters for observational studies with binary covariates. The confounding effects of measured and unmeasured variables are treated as exchangeable within a Bayesian framework. We model the joint distribution of covariates using a loglinear model with pairwise interaction terms. Hierarchical priors constrain the magnitude and direction of bias parameters. An appealing property of the method is that the conditional distribution of the unmeasured confounder follows a logistic model, giving a simple equivalence with previously proposed methods. We apply the method in a data example from pharmacoepidemiology and explore the impact of different priors for bias parameters on the analysis results.

**Keywords:** Unmeasured confounding; bias; observational studies; sensitivity analysis, Bayesian statistics, pharmacoepidemiology

**Running title:** Bayesian Sensitivity Analysis

# 1. Introduction

## 1.1 Unmeasured confounding in pharmacoepidemiology

Bias from unmeasured confounding figures prominently in pharmacoepidemiology, which is concerned with improving our understanding of the effectiveness and safety of medications. A typical pharmacoepidemiology study compares outcome response rates in patients who were prescribed a medication with those that were not. Study findings are often biased without careful adjustment for the factors that influence prescribing. Unfortunately, control of confounding is notoriously difficult because medication prescribing is intimately connected to the disease process that determines the study outcome. The myriad of patient characteristics that influence prescribing can act as powerful confounders and bias effect estimates in a manner that is difficult to predict. Epidemiologists call this *confounding by indication* because the confounders are the clinical indications for treatment (Schneeweiss, 2006).

In this paper, we revisit the example of an observational study of the effect of beta blocker therapy on mortality among British Columbia heart failure patients (McCandless, Gustafson and Levy, 2007, 2008). We have healthcare administrative data for 6969 patients discharged from hospital in 1999 and 2000 after treatment for heart failure. We followed them for one year and 1755 died. Interest lies on the association between beta blocker therapy and mortality, but the data only provide basic information on the many possible confounders. A total of 21 covariates are available in the data, including patient characteristics, comorbidities and prescribing of cardiovascular therapies. See Table 1 for a complete listing.

Let  $X$  and  $Y$  denote binary variables modeling the treatment and outcome variables respectively. We set  $X$  equal one if the patient was dispensed a beta blocker within thirty days of hospital discharge, and zero otherwise. Similarly, we let  $Y$  denote an indicator variable for death within one year of hospital discharge. We let  $\mathbf{C} = (C_1, \dots, C_p)$  denote the  $p = 21$  dimensional vector of covariates listed in Table 1. In pharmacoepidemiological studies of cardiovascular therapies, interest centers on confounding induced by the various patient illnesses. The vector  $\mathbf{C}$

includes  $q = 9$  comorbidity indicator variables measured at baseline including cerebrovascular disease (CVD), chronic obstructive pulmonary disorder (COPD), hyponatremia (HYPNAT), metastatic disorder (MTSTD), renal disease (MSRD), ventricular arrhythmia (VENTRAR), liver disease (MLD), malignancy (MALIG), and shock (HPTN).

To estimate the association between  $X$  and  $Y$  while adjusting for  $\mathbf{C}$ , we fit the logistic regression model

$$\text{logit}[Pr(Y = 1|X, \mathbf{C})] = \alpha + \beta X + \boldsymbol{\xi}^T \mathbf{C}. \quad (1)$$

Section A of Table 2 gives log odds ratio estimates and standard errors for the treatment effect  $\beta$  and selected components of the parameter vector  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_{21})$  corresponding to the association between the  $q = 9$  comorbidity variables and mortality. The quantity  $\beta$  models the association between beta blockers and mortality and is estimated as -0.32 with standard error 0.08, suggesting that beta blocker therapy reduces mortality. The corresponding odds ratio  $\exp(-0.32)=0.72$  roughly agrees with estimates reported from randomized trials of beta blockers and heart failure (Foody, Farrell and Krumholz, 2002).

Nonetheless, there are concerns about possible unmeasured confounding in this analysis. The analysis uses healthcare administrative data and it is unclear whether or not we have succeeded in adjusting adequately for patient illness and disease severity. In North America, patients with multiple comorbidities are less likely to be prescribed cardiovascular therapies (Glynn et al., 2001). When providing care for the elderly and very ill, physicians are less inclined to manage long term risk of cardiovascular events because they are focussed on treating more immediate health problems. Consequently, patients treated with beta blockers may have lower baseline mortality risk simply because they are in better health. Thus the point estimate of  $\beta$  may be biased due to confounding from unmeasured variables.

## *1.2 Bayesian Adjustment for Unmeasured Confounding and the Challenges of Prior Elicitation for Bias Parameters*

A typical sensitivity analysis for unmeasured confounding posits the existence of a unmeasured binary variable  $U$  which confounds the association between  $X$  and  $Y$ . Parallelling existing modelling frameworks (e.g. Rosenbaum and Rubin, 1983; Lin, Psaty and Kronmal, 1998; Yanagawa 1984, McCandless et al. 2007, 2008; Greenland, 2005), we can extend equation (1) to model the joint density  $P(Y, U|X, \mathbf{C}) = P(Y|X, \mathbf{C}, U)P(U|X, \mathbf{C})$  where

$$\text{logit}[P(Y = 1|X, \mathbf{C}, U)] = \alpha + \beta X + \boldsymbol{\xi}^T \mathbf{C} + \lambda U \quad (2)$$

$$\text{logit}[P(U = 1|X, \mathbf{C})] = \gamma_X X + \boldsymbol{\gamma}_C^T \mathbf{C}, \quad (3)$$

and  $\boldsymbol{\gamma} = (\gamma_X, \boldsymbol{\gamma}_C)$ . Equation (2) includes  $U$  as a covariate in the regression model for the outcome. The quantities  $(\lambda, \boldsymbol{\gamma})$  are *bias parameters* because they model information about unmeasured confounding. The parameter  $\lambda$  models the association between  $U$  and  $Y$ , while the quantity  $\boldsymbol{\gamma}$  captures the associations between  $U$  and  $(X, \mathbf{C})$ .

Equation (3) does not include a regression intercept term, and therefore assumes that the baseline prevalence of  $U$  is equal to 50%. In practice this has negligible impact on the results of sensitivity analysis because confounding from  $U$  is induced by *differences* in the prevalence of  $U$  over covariate levels. See McCandless et al. (2007), Rosenbaum and Rubin (1983) and Lin et al. (1998) for empirical results that investigate sensitivity to beliefs about the baseline prevalence of  $U$ .

The variable  $U$  is completely unmeasured. Consequently, the data provide no information about either  $\lambda$  or  $\boldsymbol{\gamma}$ , and the model is nonidentifiable. But nonidentifiability does not preclude Bayesian model fitting if additional sources of information are incorporated. Recent years have witnessed the development of numerous techniques for Bayesian adjustment for unmeasured or partially measured confounders. See for example Greenland (2005), Molitor et al. (2009), Jackson, Best and Richardson (2009) and McCandless et al. (2007, 2008). A Bayesian strategy would start by assigning proper prior distributions to model parameters that translate beliefs about the magnitude and direction of confounding by  $U$ . Bayes Theorem provides a mechanism for model fitting which synthesizes the data with prior information about bias. We study the

posterior distribution for the treatment effect  $\beta$  integrating over the unmeasured confounder  $U$ . Posterior credible intervals for the treatment effect account for uncertainty in the amount of bias from unmeasured confounding in addition to random error.

A difficulty with this estimation procedure is eliciting satisfactory prior distributions for the  $(p + 1) \times 1$  vector of bias parameters  $\gamma$ . Whereas the parameter  $\lambda$  is an easily interpretable scalar, the vector  $\gamma$  is more complex and describes the manner in which  $U$  is distributed within levels of  $X$  and  $\mathbf{C}$ . In many applications it is burdensome to obtain reasonable prior guesses for  $\gamma$ . The multiplicity of different bias scenarios also complicates sensitivity analysis where results are presented using large tables.

To mitigate this problem, virtually all published methods to adjust for unmeasured confounding assume that the unmeasured confounder is independent of measured confounders, conditional on treatment. See Lin et al. (1998), McCandless et al. (2007, 2008), Greenland (2005), Rosenbaum and Rubin (1983) for examples and Hernán and Robins (1999), VanderWeele (2008), Fewell, Smith and Sterne (2007), Schneeweiss (2006) for discussion. In the beta blocker data example, the conditional independence assumption is written as  $U \perp\!\!\!\perp \mathbf{C} | X$  and forces  $\gamma_{\mathbf{C}} = \mathbf{0}$  in equation (3). Intuitively, we assume zero correlation between  $U$  and  $\mathbf{C}$  within treatment levels. To model confounding from  $U$ , the analyst is only required to elicit plausible values for  $\lambda$  and  $\gamma_X$ .

VanderWeele (2008) and Hernán and Robins (1999) argue that it is unrealistic to assume zero correlation between measured and unmeasured confounders (i.e. force  $\gamma_{\mathbf{C}} = \mathbf{0}$ ). Furthermore, epidemiologists have shown that such assumptions give inferences from sensitivity analysis are overly conservative (Fewell et al., 2007; Schneeweiss 2006; Joffe 2001). Fewell et al. (2007) study the bias induced from unmeasured confounders that correlated with measured confounders. They demonstrate that high correlations tend to reduced unmeasured confounding. Epidemiologists are generally careful to collect data on confounders. When adjusting for measured variables, they may inadvertently adjust for a constellation of unmeasured variables

because they are correlated with one another. Covariate adjustment induces homogeneity in the distribution of unmeasured variables across treatment levels, and this reduces bias. Thus one can argue that forcing  $\gamma_{\mathbf{C}} = \mathbf{0}$  in sensitivity analysis may exaggerate the uncertainty from bias.

Returning to the beta blocker example, we attempt to elicit judgments about plausible values for  $\lambda$  and  $\gamma$ . Section B of Table 2 presents the results from fitting a loglinear model by maximum likelihood to the  $2 \times 2 \times \dots \times 2$  contingency table of cell counts over all combinations of  $X$  and the  $q = 9$  binary comorbidity variables that are included in  $\mathbf{C} = (C_1, \dots, C_{21})$ . The regression model includes 10 main effects and all  $\binom{10}{2} = 45$  pairwise interactions. Section B contains point estimates and standard errors of the interaction terms in the loglinear model, which correspond to pairwise log odds ratios for associations between combinations of variables. Elements denoted “NA” indicate terms that must be dropped from the model due to sparsity in order to obtain a valid maximum likelihood estimator. Section C of Table 2 gives the marginal prevalences of the variables.

Table 2 suggests that the comorbidity variables are confounders for the effect of  $X$  on  $Y$ , and furthermore, that they are correlated with one another. Most of the variables show associations with  $X$  and  $Y$ . Furthermore, evidence from the literature indicates that they are predictors of mortality in heart failure patients and they influence prescribing of cardiovascular therapies in Canada (Polanczyk et al., 1998; Glynn et al., 2001 ). Therefore they induce confounding. But the comorbidities are also correlated with one another. Figure 1 plots the empirical distribution of the log odds ratios in Section B of Table 2. The sample mean  $\hat{\mu}_{\gamma}$  is equal to 0.71 and this means that, on average, the log odds ratio estimates are greater than zero. In the beta blocker data there is a segment of patients that very sick with numerous comorbidities.

Table 2 illustrates that it may not be make sense to force  $\gamma_{\mathbf{C}} = \mathbf{0}$  in sensitivity analysis. In the beta-blocker data we are concerned about confounding from unmeasured indications of disease. In formulating judgments about the unmeasured  $U$ , it is possible that  $U$  is correlated with  $\mathbf{C}$ . Thus analytic adjustment for  $\mathbf{C}$  may remove much of the unmeasured confounding.

But how much does it matter? And how are we to formulate a prior distribution for the bias parameter  $\gamma$ ? Are we being conservative by forcing  $\gamma_{\mathbf{C}} = \mathbf{0}$  in equation (3)? If so then we might expect that admitting uncertainty in  $\gamma_{\mathbf{C}}$  would decrease the posterior variance of the treatment effect  $\beta$ . The vector  $\gamma_{\mathbf{C}}$  is high dimensional, and we suspect that the components of  $\gamma_{\mathbf{C}}$  are non-zero. Nonetheless it is not obvious how to formulate plausible judgments about the relationship between  $U$  and each of  $C_1, \dots, C_p$ .

### 1.3 Plan of the Paper

In a recent series of papers, McCandless et al. (2007, 2008) and Gustafson et al. (2009) describe procedures for assigning default prior to bias parameters to adjust for unmeasured confounding. By *default* we mean that the analyst need only supply modest prior input about the nature of unmeasured confounding. The analysis techniques model the confounding effects of  $(U, \mathbf{C})$  on  $X$  and  $Y$  as exchangeable within a hierarchical Bayesian framework. The bias induced by  $U$  and  $\mathbf{C}$  are assumed to be similar in magnitude. Thus the analyst can learn about unmeasured confounding from measured confounders.

The approach of McCandless et al. (2008) ignores the bias parameter  $\gamma_{\mathbf{C}}$  completely and focuses on the relationship between  $Y$ ,  $X$  and  $U$ . In contrast, Gustafson et al. (2008), proposes a methodology that assigns priors for the correlation between measured and unmeasured confounders. The approach is applicable in settings with continuous covariates, such as in Fewell et al. (2007), but it cannot be used directly in the beta blocker data, which has binary confounders.

The present article describes a new methodology that adapts the work of Gustafson et al. (2009) to accommodate studies with binary covariates. We model the joint distribution of  $(X, \mathbf{C}, U)$  using a loglinear model with pairwise interaction. Hierarchical priors allow the investigator to borrow information from  $\mathbf{C}$  in order to learn about bias from  $U$ . The method has the appealing property that conditioning on  $(X, \mathbf{C})$  yields a logistic model for unmeasured confounding that is identical to that of McCandless et al. (2007, 2008) and Lin et al. (1998).

Section 2 describes the method including the model, prior distributions and posterior computation using Markov chain Monte Carlo (MCMC). In Section 3, we apply the method to the beta blocker data. A key objective of this article is to investigate the impact of the prevailing approach to sensitivity analysis which assumes zero correlation between measured and unmeasured confounders. We contrast the results using degenerate zero mass priors that force  $\gamma_{\mathbf{C}} = \mathbf{0}$ . Following the logic of Fewell et al. (2007) and Schneeweiss (2006), we illustrate that when  $U$  and  $\mathbf{C}$  are highly correlated, then confounding from  $U$  vanishes. Setting  $\gamma_{\mathbf{C}} = \mathbf{0}$  gives results that are more conservative. Section 4 concludes with a discussion.

## 2. Bayesian Adjustment for Unmeasured Confounding

### 2.1 Model

We model the joint probability density  $P(Y, X, \mathbf{C}, U) = P(Y|X, \mathbf{C}, U)P(X, \mathbf{C}, U)$  as

$$\text{logit}[Pr(Y = 1|X, \mathbf{C}, U)] = \alpha + \beta X + \boldsymbol{\xi}^T \mathbf{C} + \lambda U \quad (4)$$

$$P(X, \mathbf{C}, U) = \frac{\exp\{\phi_X X + \boldsymbol{\phi}_{\mathbf{C}}^T \mathbf{C} + \gamma_X XU + \boldsymbol{\gamma}_{\mathbf{C}}^T \mathbf{C} U + \boldsymbol{\rho}_{\mathbf{X}}^T \mathbf{C} X + \boldsymbol{\rho}_{\mathbf{C}}^T (\mathbf{C} \oplus \mathbf{C})\}}{Q(\boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\rho})} \quad (5)$$

where  $\boldsymbol{\phi} = (\phi_X, \boldsymbol{\phi}_{\mathbf{C}})$  and  $\boldsymbol{\rho} = (\boldsymbol{\rho}_{\mathbf{X}}, \boldsymbol{\rho}_{\mathbf{C}})$ . The quantities  $\boldsymbol{\rho}_{\mathbf{X}}$  and  $\boldsymbol{\phi}_{\mathbf{C}}$  are vectors of length  $p$  and  $\boldsymbol{\rho}_{\mathbf{C}}$  is a vector of length  $\binom{p}{2}$ . The quantity  $\mathbf{C} \oplus \mathbf{C}$  denotes the vector of length  $\binom{p}{2}$  of pairwise products among the  $p$  components of  $\mathbf{C} = (C_1, \dots, C_p)$ . In other words,

$$\begin{aligned} \mathbf{C} \oplus \mathbf{C} = & (C_1 C_2, C_1 C_3, \dots, C_1 C_p, \\ & C_2 C_3, C_2 C_4, \dots, C_2 C_p, \\ & C_{p-1} C_p). \end{aligned}$$

Equation (4) is identical to equation (2) and models the response surface over levels of measured and unmeasured confounders. Equation (5) is a loglinear model with main effects and pairwise interactions (Gelman et al., 2004). The parameters  $\gamma_x$  and  $\boldsymbol{\gamma}_{\mathbf{c}}$  model the associations between  $U$  and  $X$  and between  $U$  and  $\mathbf{C}$  in exactly the same manner as equation (3). In fact, if we take  $P(X, \mathbf{C}, U)$  from equation (5) and condition on  $(X, \mathbf{C})$ , then  $P(U|X, \mathbf{C})$  is equivalent

to equation (3) because of the well known connection between logistic and loglinear models (Gelman et al., 2004). This gives an appealing equivalence between equations (4) and (5) and the previously proposed models for unmeasured confounding given by McCandless et al. (2007, 2008), Lin et al. (1998). See also the model of Rosenbaum and Rubin (1983).

In equation (5), the parameters  $(\boldsymbol{\rho}_X, \boldsymbol{\rho}_C)$  model the pairwise associations among the components of  $(X, \mathbf{C})$ . The quantities  $(\phi_X, \phi_C)$  govern the baseline prevalences of  $(X, \mathbf{C})$ . In keeping with equation (3), we do not include a main effect for  $U$ , which implies that  $U$  has a baseline prevalence of 50%. The denominator  $Q(\boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\rho})$  is the constant of normalization and is a summation of the numerator of equation (5) over the support of the binary  $(U, X, \mathbf{C})$ , which is a set with  $2^{p+2}$  elements.

## 2.2 Prior Distributions

Building on reasoning by Gustafson et al. (2009) and McCandless et al. (2008), we treat the confounding effects of  $U$  and  $\mathbf{C}$  as exchangeable within a hierarchical Bayesian framework. For the response surface model in equation (4), we assign a flat prior to the intercept  $\alpha$  and

$$\begin{aligned} \lambda, \beta, \boldsymbol{\xi} &\sim \text{N}\{0, \sigma_\lambda^2\} \\ \sigma_\lambda^2 &\sim \text{Inv-}\chi^2\{\nu_0 = 10^{-3}, \sigma_0^2 = 10^{-3}\}, \end{aligned} \tag{6}$$

in order to reflect the notion that the effects of  $(X, \mathbf{C}, U)$  on  $Y$  are similar in magnitude. The LHS of equation (6) refers implicitly to the *components* of  $\boldsymbol{\xi}$ , and  $\text{Inv-}\chi^2\{.\}$  is an inverse  $\chi^2$  distribution. The variance parameter  $\sigma_\lambda^2$  shares information between  $(X, \mathbf{C})$  and  $U$ . If  $\sigma_\lambda^2$  is small, then this shrinks the posterior for  $\lambda$  towards zero giving less pessimistic conclusions about unmeasured confounding.

In equation (5), we assign flat priors to the prevalence parameters  $\boldsymbol{\phi}$ . Eliciting a prior distribution for  $\boldsymbol{\gamma} = (\gamma_X, \boldsymbol{\gamma}_C)$  is more challenging because it includes  $p+1$  different parameters. Building on the discussion of Table 2, we use the estimates of  $\boldsymbol{\rho}$  to guide the prior for  $\boldsymbol{\gamma}$ . We

assign

$$\begin{aligned} \boldsymbol{\gamma}, \boldsymbol{\rho} &\sim \text{N}\{\boldsymbol{\mu}_\gamma, \boldsymbol{\sigma}_\gamma^2\} \\ (\boldsymbol{\mu}_\gamma, \boldsymbol{\sigma}_\gamma^2) &\propto \text{N-Inv-}\chi^2\{\boldsymbol{\mu}_0 = 0, \boldsymbol{\sigma}_0^2/\kappa_0 = 10^3; \nu_0 = 10^{-3}, \boldsymbol{\sigma}_0^2 = 10^{-3}\}, \end{aligned} \tag{7}$$

where the LHS of equation (7) refers to the components of  $\boldsymbol{\gamma}$  and  $\boldsymbol{\rho}$ , which are vectors of length  $p$  and  $\binom{p}{2}+p$ , respectively. The density  $\text{N-Inv-}\chi^2\{\cdot\}$  is normal inverse  $\chi^2$  distribution (Gelman et al., 2004). This prior distribution models the pairwise associations among  $(X, \mathbf{C}, U)$  as exchangeable a priori. The parameter  $\boldsymbol{\mu}_\gamma$  models the mean level of correlation between the measured and unmeasured confounders on the log odds scale, while  $\boldsymbol{\sigma}_\gamma^2$  captures the dispersion. Recall that Figure 1 gives a sense of the magnitude of  $\boldsymbol{\gamma}$  for the beta blocker data. Because the associations among comorbidities are large, we reason that  $\boldsymbol{\gamma}_{\mathbf{C}}$  may be non-zero. The hierarchical model in in equation (7) formalizes this idea. It uses estimates of  $\boldsymbol{\rho}$  to guide the prior for  $\boldsymbol{\gamma}$ . On a more intuitive level, we can view equation (7) as using the empirical distribution in Figure 1 as an independent prior for each component of  $\boldsymbol{\gamma}$ .

An advantage of the prior formulation in equation (7) is that it stabilize estimates for  $\boldsymbol{\rho}$ . In Section B of Table 2, we saw that maximum likelihood estimation may fail because of sparse cell counts in the loglinear model for the joint distribution of  $(X, \mathbf{C})$ . The variance parameter  $\boldsymbol{\sigma}_\gamma^2$  shrinks the interaction estimates towards a common mean and this improves estimation. Further illustration is presented in Section 3.

### 2.3 Model Fitting and Computation

Denote the data as  $data = \{(Y_i, X_i, \mathbf{C}_i); i = 1, \dots, n\}$ . If  $U$  were measured, then the likelihood function would be

$$\begin{aligned} L(\alpha, \beta, \boldsymbol{\xi}, \lambda, \boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\rho}) &\propto \prod_{i=1}^n P(Y_i|X_i, \mathbf{C}_i, U_i)P(X_i, \mathbf{C}_i, U_i) \\ &= \prod_{i=1}^n \left[ \frac{\exp\{Y_i(\alpha + \beta X_i + \boldsymbol{\xi}^T \mathbf{C}_i + \lambda U_i)\}}{1 + \exp\{\alpha + \beta X_i + \boldsymbol{\xi}^T \mathbf{C}_i + \lambda U_i\}} \times \right. \\ &\quad \left. \frac{\exp\{\phi_X X_i + \boldsymbol{\phi}_{\mathbf{C}}^T \mathbf{C}_i + \gamma_X X_i U_i + \boldsymbol{\gamma}_{\mathbf{C}}^T \mathbf{C}_i U_i + \boldsymbol{\rho}_X^T \mathbf{C}_i X_i + \boldsymbol{\rho}_{\mathbf{C}}^T (\mathbf{C}_i \oplus \mathbf{C}_i)\}}{Q(\boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\rho})} \right]. \end{aligned}$$

Because  $U$  is not measured, the likelihood for the observed data is obtained by integrating over the binary  $U$ . We obtain

$$\begin{aligned}
L(\alpha, \beta, \boldsymbol{\xi}, \lambda, \boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\rho}) &\propto \prod_{i=1}^n \left[ P(Y_i|X_i, \mathbf{C}_i, U=0)P(X_i, \mathbf{C}_i, U=0) + \right. \\
&\quad \left. P(Y_i|X_i, \mathbf{C}_i, U=1)P(X_i, \mathbf{C}_i, U=1) \right] \\
&= \prod_{i=1}^n \left[ \frac{\exp\{Y_i(\alpha + \beta X_i + \boldsymbol{\xi}^T \mathbf{C}_i)\}}{1 + \exp\{\alpha + \beta X_i + \boldsymbol{\xi}^T \mathbf{C}_i\}} \times \right. \\
&\quad \frac{\exp\{\phi_X X_i + \boldsymbol{\phi}_C^T \mathbf{C}_i + \boldsymbol{\rho}_X^T \mathbf{C}_i X_i + \boldsymbol{\rho}_C^T (\mathbf{C}_i \oplus \mathbf{C}_i)\}}{Q(\boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\rho})} + \\
&\quad \frac{\exp\{Y_i(\alpha + \beta X_i + \boldsymbol{\xi}^T \mathbf{C}_i + \lambda)\}}{1 + \exp\{\alpha + \beta X_i + \boldsymbol{\xi}^T \mathbf{C}_i + \lambda\}} \times \\
&\quad \left. \frac{\exp\{(\phi_X + \gamma_X)X_i + (\boldsymbol{\phi}_C + \boldsymbol{\gamma}_C)^T \mathbf{C}_i + \boldsymbol{\rho}_X^T \mathbf{C}_i X_i + \boldsymbol{\rho}_C^T (\mathbf{C}_i \oplus \mathbf{C}_i)\}}{Q(\boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\rho})} \right]. \quad (8)
\end{aligned}$$

This likelihood function is readily computed over points in the parameter space. We can calculate acceptance ratios for Metropolis updating to implement Bayesian inferences. One computational challenge is that the denominator  $Q(\boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\rho})$  in the loglinear model becomes prohibitively expensive to compute for large  $p$ . In practice this is not a problem for  $p$  less than 15, and alternatively, fast approximations for loglinear model fitting are available (Besag 1977).

We sample from the posterior distribution  $P(\alpha, \beta, \boldsymbol{\xi}, \lambda, \boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\rho}, \mu_\gamma, \sigma_\gamma^2, \sigma_\lambda^2 | data)$  using Metropolis updating in blocks (Gelman et al., 2004). We update sequentially from the densities

$$[\alpha, \beta, \boldsymbol{\xi} | \cdot] \quad [\lambda | \cdot] \quad [\boldsymbol{\phi}, \boldsymbol{\rho} | \cdot] \quad [\boldsymbol{\gamma} | \cdot] \quad [\mu_\gamma, \sigma_\gamma^2 | \cdot] \quad [\sigma_\lambda^2 | \cdot],$$

which are conditional on the data and remaining model parameters. We update  $(\alpha, \beta, \boldsymbol{\xi})$ ,  $\lambda$ ,  $(\boldsymbol{\phi}, \boldsymbol{\rho})$  and  $\boldsymbol{\gamma}$  using a random walk with proposal distributions that are scalar or multivariate  $t$ -distributed with small degrees of freedom and with scale matrix equal to the identity matrix multiplied by a tuning parameter that is set by trial MCMC runs. Updating  $\sigma_\lambda^2$  and  $(\mu_\gamma, \sigma_\gamma^2)$  is straightforward because they are conditionally conjugate under an  $\text{Inv-}\chi^2$  prior (Gelman et al., 2004).

Updating the bias parameters  $\lambda$  and  $\boldsymbol{\gamma}$  is challenging because the data reveal little about either quantity. As discussed in McCandless et al. (2007, 2008) and Gustafson (2009), the

model for unmeasured confounding in equations (4) and (5) is nonidentifiable. Different points in the parameter space give identical likelihood functions for the data. This can lead to slow MCMC mixing of  $(\lambda, \boldsymbol{\gamma})$ . Furthermore, the hierarchical priors in equations (6) and (7) have a marginal multivariate  $t$  distribution. See McCandless et al. (2008) for details. This distribution has heavy tails that exacerbate sampler convergence in nonidentifiable models. For example, during posterior updating the pair  $(\lambda, \sigma_\lambda^2)$  are correlated and may drift away from zero.

One pragmatic solution to speed convergence is to estimate  $\sigma_\lambda^2$  and  $(\mu_\gamma, \sigma_\gamma^2)$  beforehand using the results of Sections A and B of Table 2, and then substitute the estimates into the prior distributions of equation (6) and (7). We compute the sample mean and variance of the regression coefficients that are obtained from maximum likelihood estimation. This approach chooses the shrinkage parameter from the data in a manner akin to empirical Bayes. It trades the marginal  $t$  prior for  $\lambda$  and  $\boldsymbol{\gamma}$  with a Gaussian prior that has lighter tails, rapidly improving sampler convergence.

### 3. Analysis Results for the Beta Blocker Data

#### 3.1 Full Bayesian Analysis

We fit the model in equations (4) and (5) to the beta blocker data and estimate the association between  $X$  and  $Y$  while adjusting for  $\mathbf{C}$  and the unmeasured confounder  $U$ . In the beta blocker data, we are concerned about confounding due to unmeasured indications of illness. For this reason, we fix  $\mathbf{C}$  to include only the  $q = 9$  comorbidity variables listed in Table 2 because they are informative about patient illnesses. The remaining  $p - q = 21 - 9 = 12$  covariates are not measures of disease (e.g. age and gender) and they are included via a separate linear terms in the regression model for the outcome. In other words, we substitute equation (4) with

$$\text{logit}[Pr(Y = 1|U, \mathbf{C}, \tilde{\mathbf{C}})] = \alpha + \beta X + \boldsymbol{\xi}^T \mathbf{C} + \tilde{\boldsymbol{\xi}}^T \tilde{\mathbf{C}} + \lambda U, \quad (9)$$

where “ $\dots + \tilde{\boldsymbol{\xi}}^T \tilde{\mathbf{C}} + \dots$ ” refers to the 12 non-disease covariates, denoted  $\tilde{\mathbf{C}}$ , and we assign flat priors to  $\tilde{\boldsymbol{\xi}}$ . Furthermore, we keep the model for  $P(X, \mathbf{C}, U)$  exactly as written in equation (5)

and exclude a model for the distribution of covariates  $\tilde{C}$ .

We run a single MCMC chain of length 1000 000 after 100 000 burn-in iterations. As discussed in Section 2.3, sampler convergence of the bias parameters  $(\lambda, \gamma)$  is not ideal because of nonidentifiability, and long simulation runs are required to calculate posterior summaries. To improve convergence we estimate the hyperparameters in equations (6) and (7). We obtain the sample mean  $\hat{\mu}_\gamma = 0.71$  and standard deviation  $\hat{\sigma}_\gamma = 0.93$  by averaging the log odds ratios in Section B of Table 2 (see Figure 1). The sample standard deviation  $\hat{\sigma}_\lambda = 0.57$  is calculated from the log odds ratios in Section A. These values are substituted into the prior so that MCMC updating of  $(\mu_\gamma, \sigma_\gamma^2)$  and  $\sigma_\lambda^2$  is not required. Sampler convergence is assessed using separate simulation runs with overdispersed starting values and the diagnostics tools included in the R package CODA (R Development Core Team, 2004).

To illustrate the impact of hierarchical priors on sampler convergence, Figure 2 show mixing of selected components of  $\rho_C$ . The top figures are generated from the Bayesian analysis output. The bottom figures shows mixing of the same parameters after running a second additional analysis which fits the loglinear model in equation (5) all by itself, with  $U = 0$ ,  $\gamma = \mathbf{0}$  and flat priors on  $\phi$  and  $\rho$ . This second analysis is an identical Bayesian version of the maximum likelihood analysis used to generate Section B of Table 2. In Figure 2, we see that hierarchical priors improve mixing with more stable trace plots. The posterior distributions are shrunk towards the common mean parameter  $\hat{\mu}_\gamma = 0.71$ .

The results of the full Bayesian analysis are given in Table 1. The first column under the heading “Naive Analysis” gives benchmark results for comparing different analyses. The Naive Analysis consists of logistic regression of the outcome on treatment and covariates ignoring unmeasured confounding (i.e. fitting equation (9) all by itself and forcing  $U = 0$ ). The regression coefficient for the beta blocker effect parameter  $\beta$  is -0.32 with 95% credible interval (-0.48, -0.16), which is identical the results given in Section A of Table 2. The remaining rows in the table give estimates of the comorbidity effects  $\xi$  and the covariate effects  $\tilde{\xi}$ .

The second column with heading “Exchangeable Correlation” gives the results of fitting the full Bayesian model of equations (4) and (5) with  $\gamma_c$  non-zero and exchangeable with  $(\boldsymbol{\rho}, \gamma_X)$  according to the hierarchical prior of equation (7). The log odds ratio for the beta blocker effect parameter  $\beta$  is -0.31 95% CI (-0.64, 0.08), but this interval estimate is much wider compared to the Naive Analysis with length 0.72 versus 0.32. The reason is because the interval incorporates uncertainty from unmeasured confounding. Similar results are given in McCandless et al. (2007, 2008).

### 3.2 Assessing Prior Sensitivity for $\gamma_C$

Our modelling framework gives the opportunity to study the role of  $\gamma_C$  in adjustment for unmeasured confounding. One issue is assessing the impact of assuming that  $\gamma_C = \mathbf{0}$ . Recall many sensitivity analysis techniques assume zero correlation between measured and unmeasured confounders in order to reduce the burden of prior elicitation. To explore prior sensitivity in the beta blocker data, we change equation (7) and instead assign the prior

$$\begin{aligned} \gamma_C &= \mathbf{0} \\ \gamma_X, \boldsymbol{\rho} &\sim \text{N}\{\mu_\gamma, \sigma_\gamma^2\} \\ (\mu_\gamma, \sigma_\gamma^2) &\propto \text{N-Inv-}\chi^2\{\mu_0 = 0, \sigma_0^2/\kappa_0 = 10^3; \nu_0 = 10^{-3}, \sigma_0^2 = 10^{-3}\}, \end{aligned}$$

The results are presented in the third column of Table 1 under the heading “Zero Correlation”, meaning zero correlation between measured and unmeasured confounders within levels of treatment. Because this prior permits  $\gamma_X$  to be non-zero, this implies that  $U$  is still a confounder for the effect of  $X$  on  $Y$  because  $U$  is associated with  $X$ . The magnitude of the bias parameter  $\gamma_X$  is dictated by the hyperparameters  $(\hat{\mu}_\gamma, \hat{\sigma}_\gamma^2)$  from Figure 1.

Table 1 reveals that fixing  $\gamma_C = \mathbf{0}$  in the beta blocker data increases the posterior uncertainty about unmeasured confounding. The credible interval for the beta blocker effect in the Zero Correlation case is 10% wider relative to Exchangeable Correlation, with 95% CI (-0.67, 0.12) versus (-0.64, 0.08). Both analyses acknowledge uncertainty from unmeasured confound-

ing, but only one analysis posits that adjustment for  $\mathbf{C}$  may reduce bias because of correlation between  $\mathbf{C}$  and  $U$ . Because the prior mean for  $\gamma_{\mathbf{C}}$  is set at  $\hat{\mu}_{\gamma} = 0.71$ , this implies that the analysis learns from dependence among the covariates  $(X, \mathbf{C})$  about the magnitude of correlation between  $\mathbf{C}$  and  $U$ .

To illustrate more clearly, we can repeat the analysis while toying with fixed values for  $\gamma_{\mathbf{C}}$ . Figure 3 demonstrates what happens to interval estimates for the beta blocker effect  $\beta$  when we set  $\gamma_{\mathbf{C}}$  equal to  $\mathbf{0}, \mathbf{1}, \dots, \mathbf{5}$ . The grey shaded region refers to the width and positioning of the Naive Analysis interval estimate for  $\beta$ . When  $\gamma_{\mathbf{C}}$  is large, the interval estimates collapse towards the shaded region and we obtain inferences that are identical to assuming that there is no unmeasured confounding. If  $U$  and  $\mathbf{C}$  are perfectly correlated within treatment groups, then it means that regression adjustment for  $\mathbf{C}$  eliminates bias from  $U$ . Within levels of  $\mathbf{C}$ , all subjects either have  $U$  or they do not. This forces conditional independence between  $U$  and  $X$  given  $\mathbf{C}$ , and therefore  $U$  is not a confounder. As an illustration, consider equation (3), which can be written  $P(U = 1|X, \mathbf{C}) = \frac{\exp\{\gamma_X X + \gamma_{\mathbf{C}}^T \mathbf{C}\}}{1 + \exp\{\gamma_X X + \gamma_{\mathbf{C}}^T \mathbf{C}\}}$ . When  $\gamma_{\mathbf{C}} \rightarrow \infty$  we have  $P(U = 1|X, \mathbf{C}) \rightarrow 1$  for all subjects who have non-zero  $\mathbf{C}$ . Thus  $U$  is conditionally independent of  $X$  given  $\mathbf{C}$ .

In Figure 3, the posterior means are shifted slightly towards zero relative to the Naive analysis. This finding is also visible in the Zero and Exchangeable correlation analyses of Table 1. The reason for this shift is because of the informative prior on  $\beta$  in equation (6). We have set  $\hat{\sigma}_{\lambda} = 0.57$ , which corresponds to the belief that *a priori* the log odds ratio  $\beta$  lies in the interval  $(-1.96 \times 0.57, 1.96 \times 0.57) = (-1.11, 1.11)$  with probability 95%. In contrast, the Naive interval is fit by maximum likelihood, which presumes a flat prior on  $\beta$ .

An additional feature in the Exchangeable Correlation analysis is the interval estimates for the log odds ratios  $\boldsymbol{\xi}$  which govern the effect of comorbid illnesses on mortality are wider than the other analyses. In contrast, the inferences for  $\tilde{\boldsymbol{\xi}}$  in Table 1 are all essentially identical. The reason is because when  $\gamma_{\mathbf{C}}$  is non-zero, then this implies that  $U$  is a confounder for the effect

of  $\mathbf{C}$  on  $Y$ . We have an association between  $U$  and  $Y$  and also between  $U$  and  $\mathbf{C}$ . Because the magnitude and direction of these associations are unknown, this inflates the interval estimates for  $\boldsymbol{\xi}$ , but has no impact on estimation of  $\tilde{\boldsymbol{\xi}}$ . In adjusting for  $U$ , our loglinear model for  $P(X, \mathbf{C}, U)$  handles the variables  $X$  and  $\mathbf{C}$  in a manner that is symmetric, without assigning special status to  $X$  because it is the putative treatment of interest.

## 4. Discussion

Recent years have witnessed new innovation in Bayesian techniques to adjust for unmeasured confounding. A challenge with existing methods is that the user is often required to elicit prior distributions for high dimensional parameters that model competing bias scenarios. This can render the methods unwieldy. Similar problems are encountered in sensitivity analysis that use large tables to present analyses over a range of bias parameter inputs.

In this paper we propose a novel methodology to adjust for unmeasured confounding that derives default priors for bias parameters for observational studies with binary covariates. The confounding effects of measured and unmeasured variables are modelled as exchangeable within a hierarchical Bayesian framework. This operationalizes the intuitive idea that the confounding from measured variables may be informative about unmeasured variables. This estimation strategy is appealing in pharmacoepidemiology where confounding is largely the result of physician prescribing preferences based on patient characteristics. Careful measurement and adjustment for treatment indications may partially reduce residual confounding because of correlations among variables.

We apply the method in a data example from pharmacoepidemiology and study the consequences of assuming there is zero correlation between measured and unmeasured confounders within treatment groups. As intuition suggests, power correlations tend to reduce uncertainty about bias. Furthermore we can learn about such correlations from the data and then use them as an input in sensitivity analysis.

A challenge with our methodology is assessing the adequacy of prior distributions and modelling assumptions. Because  $U$  is totally unobserved it means that we cannot study the relationship between  $U$  and  $(Y, X, \mathbf{C})$ . Nonidentifiability will exacerbate the impact of the prior distribution on the analysis results (Gustafson et al, 2009). One direction for future research is to collect additional data for  $U$  in a subset of the population. For example, we could use external validation data from surveys. This approach is explored in the context of environmental epidemiology by McCandless, Richardson and Best (2009) and Molitor et al. (2009).

## References

- Besag, J. (1977). Efficiency of pseudo-likelihood estimation for simple Gaussian fields, *Biometrika* **64**, 616–618.
- Fewell, Z., Smith, D. and Sterne, J. A. C. (2007). The Impact of Residual and Unmeasured Confounding in Epidemiologic Studies: A Simulation Study, *American Journal of Epidemiology* **166**, 646–656.
- Foody, J. A. M., Farrell, M. H., and Krumholz, H. M. (2002).  $\beta$ -blocker therapy in heart failure: Scientific Review, *Journal of the American Medical Association* **278**, 883–889.
- Gelman A., Carlin J. B., Stern H. S., and Rubin D. B. (2004). *Bayesian Data Analysis, 2nd Edition*. Chapman Hall/CRC:New York.
- Glynn, R. J., Knight, E. L., Levin, R., and Avorn, J. (2001). Paradoxical relations of drug treatment with mortality in older persons, *Epidemiology* **12**, 682–689.
- Greenland, S. (2005). Multiple bias modelling for analysis of observational data (with discussion), *Journal of the Royal Statistical Society Series A* **168**, 267–306.
- Gustafson, P., McCandless, L. C., Levy, A. R., and Richardson, S. R. (2009). Simple Bayesian sensitivity analysis for mismeasured and unobserved confounders. (*Under Review*).

- Hernán, M. A., and Robins, J. M. (1999) Letter to the editor, *Biometrics* **55**, 1316–1317.
- Jackson, C., Best, N. G. and Richardson, S. (2009) Bayesian graphical models for regression on multiple datasets with different variables, *Biostatistics*, **10**, 335–51.
- Joffe, M. M.. (2000) Confounding by indication: the case of calcium channel blocker, *Pharmacoepidemiology and Drug Safety* **9**, 37–41.
- Lin, D. Y., Psaty, B. M. and Kronmal, R. A. (1998) Assessing the sensitivity of regression results to unmeasured confounders in observational studies, *Biometrics* **54**, 948–963.
- McCandless, L. C., Gustafson, P., and Levy, A. R. (2007). Bayesian sensitivity analysis for unmeasured confounding in observational studies, *Statistics in Medicine* **26**, 2331–47.
- McCandless, L. C., Gustafson, P., and Levy, A. R. (2008) A sensitivity analysis using information about measured confounders yielded improved assessments of uncertainty from unmeasured confounding, *Journal of Clinical Epidemiology* **61**, 247–255.
- McCandless, L. C., Richardson, S., and Best, N. B. (2009) Adjustment for unmeasured confounding using propensity scores (*under review*). Available at <http://www.sfu.ca/~lmccandl>.
- Molitor, N., Best, N. B., Jackson, C., and Richardson, S. (2009) Using Bayesian graphical models to model biases in observational studies and to combine multiple data sources: Application to low birthweight and water disinfection by-products, *Journal of the Royal Statistical Society Series A* **172**, 615–37.
- Polanczyk, C. A., Rohde, L. E., Philbin, E. A., and Di Salvo, T. G. (1998). A new casemix adjustment index for hospital mortality among patients with congestive heart failure, *Medical Care* **36**, 1489–99.
- R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, 2004. ISBN 3-900051-00-3. URL <http://www.R-project.org>.

- Rosenbaum, P. R. and Rubin, D. B. (1983) Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome, *Journal of the Royal Statistical Society Series B* **45**, 212–218.
- Schneeweiss, S. (2006). Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics, *Pharmacoepidemiology and Drug Safety* **15**, 291–303.
- VanderWeele, T. J. (2008) Sensitivity analysis: Distributional assumptions and confounding assumptions, *Biometrics* **64**, 645–649.
- Yanagawa, T. (1984) Case-control studies: Assessing the effect of a confounding factor, *Biometrika* **71**, 191–94.

Table 1: Adjusted log odds ratios (95% interval estimates) for the associations between covariates and mortality.

	Naive Analysis		Bayesian Sensitivity Analysis	
	Correlation between $U$ and $C$ given $X$		Exchangeable Correlation	Zero Correlation
	$(\gamma_C \text{ exchangeable with } \rho, \gamma_X)$		$(\gamma_C \text{ Fixed at Zero})$	
Beta blocker use	-0.32 (-0.48, -0.16)	-0.31 (-0.64, 0.08)	-0.29 (-0.67, 0.12)	
<i>Comorbid conditions</i>				
CVD	0.36 (-0.30, 1.03)	0.25 (-0.49, 0.91)	0.26 (-0.42, 0.88)	
COPD	0.07 (-0.20, 0.34)	0.04 (-0.33, 0.42)	0.06 (-0.22, 0.32)	
HYPNAT	0.10 (-0.20, 0.41)	0.06 (-0.33, 0.43)	0.10 (-0.20, 0.39)	
MTSTD	1.79 (1.11, 2.48)	1.42 (0.75, 2.10)	1.51 (0.87, 2.12)	
MSRD	0.63 (0.43, 0.83)	0.56 (0.19, 0.89)	0.62 (0.42, 0.82)	
VENTRAR	0.35 (-0.29, 1.00)	0.23 (-0.38, 0.87)	0.22 (-0.41, 0.85)	
MLD	0.53 (0.01, 1.04)	0.41 (-0.09, 0.93)	0.42 (-0.01, 0.84)	
MALIG	1.06 (0.68, 1.44)	0.91 (0.44, 1.35)	0.97 (0.60, 1.36)	
HPTN	-0.01 (-0.58, 0.56)	-0.03 (-0.61, 0.49)	-0.02 (-0.51, 0.48)	
<i>Patient characteristics</i>				
Female sex	-0.29 (-0.4, -0.17)	-0.30 (-0.42, -0.18)	-0.30 (-0.42, -0.18)	
Age				
<65	1	1	1	
65-74	0.38 (0.15, 0.61)	0.38 (0.15, 0.62)	0.39 (0.15, 0.63)	
75-84	0.81 (0.60, 1.02)	0.82 (0.61, 1.04)	0.83 (0.60, 1.05)	
85+	1.21 (0.99, 1.43)	1.23 (1.01, 1.47)	1.24 (1.00, 1.48)	
<i>Characteristics of hospitalization</i>				
Transferred admission	0.72 (0.55, 0.88)	0.74 (0.56, 0.91)	0.73 (0.56, 0.91)	
Hospital LOS	0.08 (0.05, 0.12)	0.09 (0.05, 0.13)	0.09 (0.05, 0.13)	
<i>Heart Failure medications</i>				
Digoxin	0.04 (-0.09, 0.17)	0.03 (-0.10, 0.17)	0.03 (-0.09, 0.17)	
Diuretics	-0.11 (-0.25, 0.03)	-0.11 (-0.26, 0.03)	-0.11 (-0.26, 0.03)	
CCB	-0.16 (-0.33, 0.01)	-0.16 (-0.33, 0.01)	-0.16 (-0.34, 0.00)	
ACE inhibitor	-0.13 (-0.26, 0.00)	-0.13 (-0.27, 0.00)	-0.14 (-0.27, -0.01)	
ARB	-0.07 (-0.43, 0.29)	-0.09 (-0.47, 0.28)	-0.08 (-0.45, 0.26)	
Statin	-0.16 (-0.38, 0.06)	-0.17 (-0.40, 0.05)	-0.16 (-0.40, 0.06)	

Table 2: Description of the confounding induced by the  $q = 9$  comorbidity indicator variables in the association between  $X$  and  $Y$ .

**Section A:** Log odds ratio estimates (standard errors) of the treatment effect  $\beta$  and selected components of  $\xi$ .

Beta Blocker	CVD	COPD	HYPNAT	MTSTD	MSRD	VENTRAR	MLD	MALIG	HPTN
-0.32 (0.08)	0.4 (0.3)	0.1 (0.1)	0.1 (0.2)	1.8 (0.3)	0.6 (0.1)	0.4 (0.3)	0.5 (0.3)	1.1 (0.2)	0.0 (0.3)

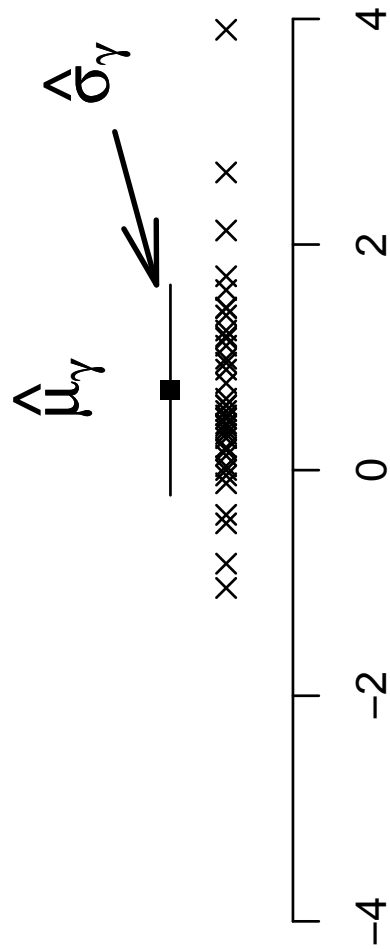
**Section B:** Log odds ratio estimates (standard errors) of the  $\binom{10}{2} = 45$  pairwise associations among treatment  $X$  and the  $q = 9$  binary comorbidity indicator variables. “NA” denotes interaction terms that were dropped from the loglinear model in order to obtain a valid maximum likelihood estimate vector.

Beta Blocker	.	0.5 (0.3)	-0.8 (0.2)	0.0 (0.2)	-0.5 (0.5)	0.2 (0.1)	0.4 (0.4)	-1.0 (0.5)	0.0 (0.3)	-0.1 (0.4)
CVD	.	.	1.4 (0.5)	1.0 (0.7)	NA	1.2 (0.4)	NA	NA	1.2 (1.0)	2.6 (0.6)
COPD	.	.	.	0.3 (0.3)	0.0 (0.8)	-0.4 (0.3)	0.6 (0.7)	0.4 (0.5)	0.1 (0.5)	0.6 (0.5)
HYPNAT	.	.	.	.	0.5 (0.8)	1.0 (0.2)	0.9 (0.7)	0.3 (0.7)	0.5 (0.5)	1.6 (0.4)
MTSTD	.	.	.	.	.	1.1 (0.7)	NA	NA	3.9 (0.3)	NA
MSRD	.	.	.	.	.	.	1.2 (0.4)	0.8 (0.4)	0.2 (0.4)	0.4 (0.4)
VENTRAR	.	.	.	.	.	.	.	1.4 (1.0)	1.7 (0.7)	2.1 (0.8)
MLD	.	.	.	.	.	.	.	.	1.4 (0.6)	NA
MALIG	.	.	.	.	.	.	.	.	.	0.6 (1.0)
HPTN	.	.	.	.	.	.	.	.	.	.

**Section C:** Prevalences of  $X$  and the  $q = 9$  comorbidity indicator variables, given as counts from the 6969 study participants.

1295	45	310	224	48	530	45	74	130	67
------	----	-----	-----	----	-----	----	----	-----	----

Figure 1: Sample mean ( $\hat{\mu}_\gamma = 0.71$ ) and standard deviation ( $\hat{\sigma}_\gamma = 0.93$ ) of the log odds ratios estimates listed in Section B of Table 1.



Log Odds Ratios for Associations Among X, C<sub>1</sub>, ... C<sub>p</sub>

Figure 2: MCMC trace plots of selected components of  $\rho_C$ . Horizontal lines indicates the common mean parameter  $\hat{\mu}_\gamma = 0.71$ , in the hierarchical prior of equation (7).

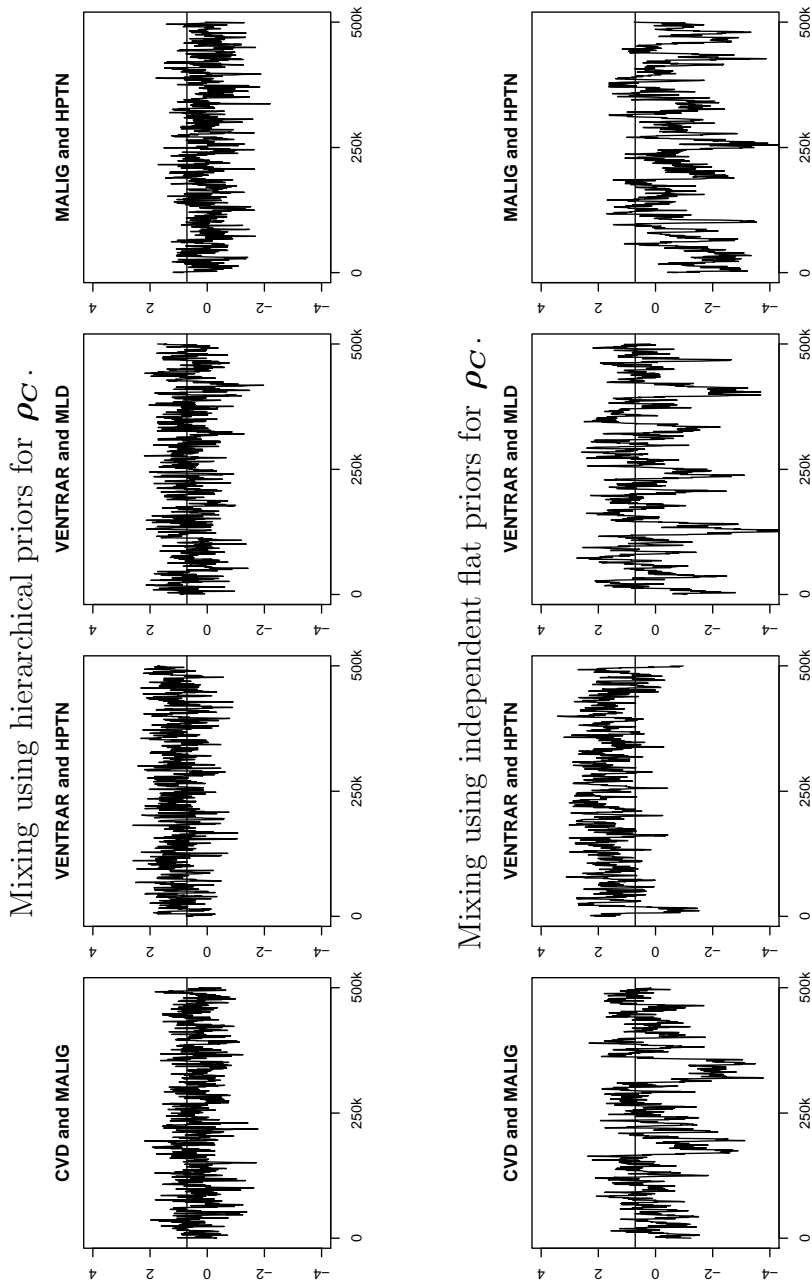


Figure 3: 95% credible intervals for the beta blocker effect  $\beta$  when we repeat the Bayesian analysis with  $\gamma_c$  held fixed at  $0, 1, \dots, 5$ .

