Hierarchical Priors for Bias Parameters in Bayesian Sensitivity Analysis for Unmeasured Confounding

Lawrence C. McCandless[1]
Paul Gustafson[2]
Adrian R. Levy[3]
Sylvia Richardson[4]

[1] Faculty of Health Sciences, Simon Fraser University, Canada. [2] Department of Statistics, University of British Columbia. [3] Department of Community Health and Epidemiology, Dalhousie University. [4] Department of Epidemiology and Public Health, Imperial College London.

September 2010

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

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: bias; observational studies; 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 [1].

In this paper, we illustrate the problem of unmeasured confounding using the data example of McCandless, Gustafson and Levy [2,3]. The authors conducted a retrospective cohort study to estimate the effect of beta blocker therapy on mortality in heart failure patients living in British Columbia. We have healthcare administrative data for 6969 persons 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, disease indicator variables 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 $\boldsymbol{C} = (C_1, \ldots, C_p)$ denote the $p = 21$ dimensional vector of covariates listed in Table 1. In pharmacoepidemiological studies of cardiovascular ther-

apies, interest centers on confounding induced by the various patient illnesses. The vector $\boldsymbol{C}$ includes $q = 9$ disease 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), cancer (CAN), and cardiogenic shock (CARS).

To estimate the association between beta blocker therapy and mortality while adjusting for confounding, we fit a logistic regression of Y on X and $\boldsymbol{C}$. The results are presented in the first column of Table 1 under the heading "Naive Analysis". The table displays regression coefficients, which are log odds ratios, with 95% interval estimates. The regression coefficient for the treatment effect X is estimated as -0.32 with 95% interval estimates (-0.48, -0.16), suggesting that beta blocker therapy reduces mortality. The corresponding odds ratio $\exp(-0.32) = 0.72$ agrees closely with estimates reported from randomized trials of beta blockers and heart failure. In a scientific review of meta-analyses of randomized trials, Foody, Farrell and Krumholz [4], found that beta blocker use is associated with a consistent 30% reduction in mortality compared to placebo.

Nonetheless, there are concerns about unmeasured confounding. The problem is that the probability of being prescribed a beta blocker is influenced by severity of heart disease in the patient, which in turn affects risk of death. This analysis uses healthcare administrative data and it is unclear whether or not we can adequately measure and adjust for severity of heart disease. The data contain no detailed clinical information on the factors that influence prescribing, which are recorded on medical charts. For example, one unmeasured confounder is the class of heart failure. This is an ordinal variable with four categories that indicates the severity of heart failure. Another unmeasured confounder is ejection fraction, which is a measure of heart function. Both of these variables are important predictors of mortality and treatment (Foody et al., 2002). They can either increase or decrease the probability of receiving a beta blocker, depending on the preferences of the prescribing physician. See Glynn et al. [5] for review of

how cardiovascular therapies are prescribed in North America.

1.2 Bayesian Sensitivity Analysis 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. Paralleling existing modelling frameworks (e.g. [2,3,6-9]), we model the probability density $P(Y, U|X, \boldsymbol{C}) = P(Y|X, \boldsymbol{C}, U)P(U|X, \boldsymbol{C})$ where

$$\text{logit}[P(Y = 1|X, \boldsymbol{C}, U)] = \beta_0 + \beta_{XY}X + \boldsymbol{\beta_{CY}}^T \boldsymbol{C} + \beta_{UY}U \quad (1)$$

$$\text{logit}[P(U = 1|X, \boldsymbol{C})] = \gamma_U + \gamma_{XU}X + \boldsymbol{\gamma_{CU}}^T \boldsymbol{C}. \quad (2)$$

See Table 2 for a detailed explanation of the variables and parameters. Equation (1) includes U as a missing covariate in the regression model for the outcome. Equation (2) characterizes the distribution of the missing confounder. The quantity β_{XY} is the parameter of primary interest and is the causal log odds ratio for the effect of X on Y conditional on $(\boldsymbol{C}, U)$. Provided that all models are correctly specified and there are no *additional* unmeasured confounders, then the parameter β_{XY} has a causal interpretation. The quantities $\beta_{UY}, \gamma_U, \gamma_{XU}$ and $\boldsymbol{\gamma_{CU}}$ are *bias parameters* because they determine the magnitude of unmeasured confounding. The parameter β_{UY} governs the association between U and Y, conditional on $(X, \boldsymbol{C})$, while the parameters γ_{XU} and $\boldsymbol{\gamma}_{CU}$ captures the associations between U and $(X, \boldsymbol{C})$. The quantity $\exp(\gamma_U)/(1+\exp(\gamma_U))$ is the prevalence of $U = 1$ when $X = 0$ and $C_1, \ldots, C_{21} = 0$. See Table 2 for details.

The variable U is completely unmeasured. Consequently, the data provide no information about the relationship between U and the measured variables Y, X and $\boldsymbol{C}$, 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 [2,3,9-11]. 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 β_{XY} 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 Bayesian sensitivity analysis is eliciting prior distributions for the bias parameters. In particular, the quantities γ_U, γ_{XU}, $\boldsymbol{\gamma_{CU}}$ consist of $p+2$ different parameters that characterize how U is distributed within levels of X and $\boldsymbol{C}$. In many applications, it is burdensome to obtain reasonable prior guesses for $\boldsymbol{\gamma}_{CU}$, which describes the association between $\boldsymbol{C}$ and U given X. An additional problem with using equations (1) and (2) for sensitivity analysis is that there are many combinations of bias parameters that are equally plausible. This can make it difficult to display results without presenting many tables.

To mitigate this problem, virtually all sensitivity analysis techniques assume that the unmeasured confounder is independent of measured confounders, conditional on treatment. Mathematically, we write $U \perp\!\!\!\perp \boldsymbol{C}|X$, where "$\perp\!\!\!\perp$" denotes conditional independence. See [1-3,6,8,9] for examples and [1,12-14] for discussion. In equations (1) and (2), this assumption forces $\boldsymbol{\gamma_{CU}} = \mathbf{0}$, where $\mathbf{0}$ is a zero vector of length $p+1$, and then explores sensitivity for the remaining bias parameters β_{UY}, γ_{XU} and γ_U.

VanderWeele [12] and Hernán and Robins [13] argue that it is unrealistic to assume that $U \perp\!\!\!\perp \boldsymbol{C}|X$. Furthermore, epidemiologists argue that such assumptions give inferences from sensitivity analysis that are too *pessimistic* [1,14]. In a simulation study, Fewell et al. [14] demonstrate that high correlations between measured and unmeasured confounders tends to *reduce* bias from unmeasured confounding. Intuitively, the reason is because adjusting for measured variables may control for unmeasured variables because they are correlated with one another. This reasoning suggests that forcing $\boldsymbol{\gamma_{CU}} = \mathbf{0}$ for convenience may actually *exaggerate*

the sensitivity of the analysis results to unmeasured confounding.

1.3 Correlations Between Measured and Unmeasured Confounders in the Beta Blocker Data

Returning to the beta blocker example, we attempt to elicit judgments about plausible values for the bias parameters β_{UY}, γ_U, γ_{XU} and $\boldsymbol{\gamma_{CU}}$. Table 3 describes the confounding induced by the $q = 9$ disease indicator variables listed in Table 1. In Section A we list the conditional log odds ratios for the association between each variable and mortality by copying and pasting from the Naive analysis column of Table 1. Section B describe the pairwise conditional associations among the variables X and the $q = 9$ disease indicator variables. In Section B, we fit a loglinear model by maximum likelihood to the $2 \times 2 \times \ldots \times 2$ contingency table of cell counts over all combinations of X and the $q = 9$ binary disease indicator variables that are included in $\boldsymbol{C} = (C_1, \ldots, C_{21})$. The regression model includes 10 main effects and all $\binom{10}{2} = 45$ pairwise interactions. Section B contains point estimates and standard errors of coefficients of the interaction terms in the loglinear model. These coefficients correspond to conditional log odds ratios for pairwise associations between variables [15]. Elements denoted "NA" indicate terms that were dropped from the model due to sparsity in order to obtain a valid maximum likelihood estimator. Section C of Table 3 gives the prevalences of the disease variables.

Table 3 suggests that the disease indicator 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 (Section A). Furthermore, evidence from the literature indicates that they are predictors of mortality in heart failure patients and they influence prescribing of cardiovascular therapies [5,16]. Therefore they induce confounding. But the disease variables are also correlated with one another. In Section B of Table 3, most of the log odds ratios are greater than zero. Figure 1 plots the log odds ratios. The sample mean is equal to 0.71, which gives an average odds ratio of exp(0.71)=2.05. This suggests that in the beta blocker data, patients

who have one disease are also likely to have other diseases. In other words, the confounders are correlated with one another.

The missing confounder U is a binary indicator of the severity of heart disease, such as ejection fraction or class of heart failure. Both of these quantities are measures of heart function. In formulating judgments about U, it is possible that U is correlated with $\boldsymbol{C}$. Vassan et al. [17] studied ejection fraction in heart failure patients and showed that patients with low ejection fraction are more likely to have diabetes, hypertension, high blood pressure and other chronic illnesses. This suggests that adjustment for $\boldsymbol{C}$ in the Naive analysis of Table 1 may also control for confounding from U. Therefore, if we do a sensitivity analysis assuming that $\boldsymbol{\gamma_C} = \mathbf{0}$ (i.e. assuming $U \perp\!\!\!\perp \boldsymbol{C}|X$) then this may exagerate the bias from U.

Thus we are faced with a conundrum: One the one hand it seems unrealistic and possibly harmful to assume that $\boldsymbol{\gamma_C} = \mathbf{0}$ in sensitivity analysis. But on the other hand it is not clear how to assign a prior for $\boldsymbol{\gamma_C}$ because it is a p-dimensional vector and there is only limited information available about U.

1.4 Plan of the Paper

One way to elicit priors for the bias parameters is to assume that the confounding effects of measured and unmeasured confounders are exchangeable in a Bayesian analysis. In other words, to assume that the confounding induced by U is similar in magnitude to the confounding induced by $\boldsymbol{C}$. The assumption of exchangeability is a strong one, however it is has been used previously in epidemiology to form *qualitative* judgments about unmeasured confounding. For example, in a 2002 review paper on confounding by indication, Joffe [18] writes that "... one can learn about unmeasured confounders and confounding from measured factors. The argument is sometimes advanced that if adjustment for known covariates fails to change the measure of effect, there must be little residual confounding.... When control for measured factors reveals confounding, it is then more likely that there is residual confounding." This logic rests on the assumption

that the measured and unmeasured confounders are similar. If the investigator collects enough covariate information on the patients in the study, then this can be used to characterize the bias that would be produced from a confounder that was missing. Other examples of this reasoning from pharmacoepidemiology are described by Schneeweiss [1].

McCandless et al. [3] and Gustafson et al. [19] describe Bayesian methods that assume exchangeability in the confounding effects of measured and unmeasured confounders. In the paper of McCandless et al. [3], the authors analyze the beta blocker data, but they ignore the bias parameter $\boldsymbol{\gamma_{CU}}$ altogether because of the difficulties of prior elicitation. Gustafson et al. [19] consider the specific case where all of the measured covariates are continuous and are assumed to have a multivariate Gaussian distribution. Their approach involves estimating the covariance matrix of the covariates and then using it to construct a prior distribution for bias parameters. However, the method of Gustafson et al. [19] cannot be used with binary covariates, as is the case with the beta blocker data.

In this article, we propose a new method for that accommodates observational studies with binary covariates. We model the joint distribution of $(X, \boldsymbol{C}, U)$ using a loglinear model with pairwise interactions. Hierarchical priors borrow information from $\boldsymbol{C}$ in order to learn about bias from U. The method has the appealing property that conditioning on $(X, \boldsymbol{C})$ yields a logistic model for unmeasured confounding that is identical to that of McCandless et al. [2,3] and Lin et al. [8]. Section 2 describes the method including the model, prior distributions and posterior computation using Markov chain Monte Carlo (MCMC). The exchangeability assumption is a strong one, and we discuss the plausibility and generality of our method in Section 2.2. 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 study the results when using degenerate zero mass priors that force $\boldsymbol{\gamma_{CU}} = \mathbf{0}$. Following the logic of Schneeweiss [1] and Fewell et al. [14], we illustrate that if U and $\boldsymbol{C}$ are highly correlated, then confounding from U

tends to diminish. In the beta blocker data, setting $\boldsymbol{\gamma_{CU}} = \mathbf{0}$ for convenience gives conclusions that are too pessimistic. Section 4 concludes with a discussion.

2. Bayesian Adjustment for Unmeasured Confounding

2.1 Model

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

$$\begin{aligned} \text{logit}[Pr(Y=1|X, \boldsymbol{C}, U)] &= \beta_0 + \beta_{XY}X + \boldsymbol{\beta}_{CY}^T \boldsymbol{C} + \beta_{UY}U \qquad (3) \\ P(X, \boldsymbol{C}, U) &= \frac{1}{Q(\gamma_X, \boldsymbol{\gamma_C}, \gamma_U, \gamma_{XU}, \boldsymbol{\gamma_{CU}}, \boldsymbol{\gamma_{CX}}, \boldsymbol{\gamma_{C\oplus C}})} \times \\ & \quad \exp\{\gamma_X X + \boldsymbol{\gamma_C}^T \boldsymbol{C} + \gamma_U U + \\ & \quad \gamma_{XU}XU + \boldsymbol{\gamma_{CU}}^T \boldsymbol{C}U + \boldsymbol{\gamma_{CX}}^T \boldsymbol{C}X + \boldsymbol{\gamma_{C\oplus C}}^T(\boldsymbol{C} \oplus \boldsymbol{C})\}. \qquad (4) \end{aligned}$$

Equation (3) is identical to equation (1) and models the log odds of the outcome as a function of X, $\boldsymbol{C}$ and U. Equation (4) is a loglinear model for the joint distribution of $(X, \boldsymbol{C}, U)$ with main effects and pairwise interactions [15]. The denominator $Q(\gamma_X, \boldsymbol{\gamma_C}, \gamma_U, \gamma_{XU}, \boldsymbol{\gamma_{CU}}, \boldsymbol{\gamma_{CX}}, \boldsymbol{\gamma_{C\oplus C}})$ is the constant of normalization and is a summation of the numerator of equation (4) over the support of the binary $(U, X, \boldsymbol{C})$, which is a set with 2^{p+2} elements.

See Table 2 for a description of variables and parameters. The quantities γ_X, $\boldsymbol{\gamma_C}$, γ_U are the main effects of $(X, U, \boldsymbol{C})$ in the loglinear model, whereas γ_{XU}, $\boldsymbol{\gamma_{CU}}$, $\boldsymbol{\gamma_{CX}}$ and $\boldsymbol{\gamma_{C\oplus C}}$ govern the interaction terms. The quantity $\boldsymbol{C} \oplus \boldsymbol{C}$ denotes the vector of length $\binom{p}{2}$ of pairwise products among the p components of $\boldsymbol{C} = (C_1, \ldots, C_p)$. In other words,

$$\begin{aligned} \boldsymbol{C} \oplus \boldsymbol{C} &= (C_1C_2,\ C_1C_3\ , \ldots, C_1C_p, \\ & \quad C_2C_3,\ C_2C_4,\ \ldots, C_2C_p, \\ & \quad C_{p-1}C_p). \end{aligned}$$

The parameter $\boldsymbol{\gamma_{C\oplus C}}$ is a vector of regression coefficients for the interaction terms $\boldsymbol{C} \oplus \boldsymbol{C}$, and it capture the pairwise conditional associations between components of $\boldsymbol{C}$.

There is a well-known connection between logistic and loglinear models through conditioning [15]. The parameters γ_{XU} and $\boldsymbol{\gamma_{CU}}$ are conditional log odds ratios for the association between

$(X, \boldsymbol{C})$ and U. If we take $P(X, \boldsymbol{C}, U)$ from equation (4) and condition on $(X, \boldsymbol{C})$, then $P(U|X, \boldsymbol{C})$ obeys equation (2). We have

$$\begin{aligned}
\text{logit}[P(U=1|X, \boldsymbol{C})] &= \log\left[\frac{P(U=1|X, \boldsymbol{C})}{P(U=0|X, \boldsymbol{C})}\right] \\
&= \log\left[\frac{P(U=1, X, \boldsymbol{C})}{P(U=0, X, \boldsymbol{C})}\right] \\
&= \log\big[\exp\{\gamma_X X + \boldsymbol{\gamma_C}^T \boldsymbol{C} + \gamma_U + \\
&\quad \gamma_{XU} X + \boldsymbol{\gamma_{CU}}^T \boldsymbol{C} + \boldsymbol{\gamma_{CX}}^T \boldsymbol{C} X + \boldsymbol{\gamma_{C\oplus C}}^T (\boldsymbol{C} \oplus \boldsymbol{C})\} / \\
&\quad \exp\{\gamma_X X + \boldsymbol{\gamma_C}^T \boldsymbol{C} + \boldsymbol{\gamma_{CX}}^T \boldsymbol{C} X + \boldsymbol{\gamma_{C\oplus C}}^T (\boldsymbol{C} \oplus \boldsymbol{C})\}\big]. \\
&= \gamma_U + \gamma_{XU} X + \boldsymbol{\gamma_{CU}}^T \boldsymbol{C}.
\end{aligned}$$

This gives an appealing equivalence between our proposed model and previously proposed models for unmeasured confounding given by McCandless et al. [2,3], Lin et al. [8]. See also the model of Rosenbaum and Rubin [6].

2.2 Prior Distributions

Suppose that $\theta_1, \theta_2, \ldots, \theta_J$ are a collection of J unknown parameters. In Bayesian analysis, we say that $\theta_1, \theta_2, \ldots, \theta_J$ are *exchangeable* in their joint distribution if $P(\theta_1, \theta_2, \ldots, \theta_J)$ is invariant to the permutation of the indices $(1, \ldots, J)$ [20]. An exchangeable prior distribution is plausible if, based on the available information, we are unable to distinguish one parameter from another. Gelman et al. [20] writes that "In practice, ignorance implies exchangeability. Generally, the less we know about a problem, the more confidently we can make claims about of exchangeability".

Now suppose that $\theta_1, \theta_2, \ldots, \theta_J$ are exchangeable. Then following standard principles of Bayesian analysis, we can apply de Finetti's Theorem, which states that in the limit as $J \to \infty$, then under standard regularity conditions any exchangeable distribution for $\theta_1, \theta_2, \ldots, \theta_J$ can be expressed as an independent and identically distributed mixture of random variables conditional on some latent variable [20]. In other words, $\theta_1, \theta_2, \ldots, \theta_J$ can be modelled as a random sample from a distribution. Technically, the theorem does not apply for finite J . See Bernardo and Smith [21] for further discussion of exchangeability.

Building on the discussion of unmeasured confounding in Section 1, we model the confounding effects of U and $\boldsymbol{C}$ as exchangeable. For the outcome model, we assign a diffuse normal prior to β_0 with mean zero and variance 10^3, and we assign

$$\begin{aligned} \beta_{XY}, \boldsymbol{\beta_{CY}}, \beta_{UY} &\overset{\text{IID}}{\sim} \text{N}(0, \sigma^2_\beta) \\ \sigma^2_\beta &\sim \text{Inv-}\chi^2\left(10^{-3}, 10^{-3}\right). \end{aligned} \tag{5}$$

The left hand side (LHS) of equation (5) refers to the individual components of β_{XY}, $\boldsymbol{\beta_{CY}}$, β_{UY}, and Inv-$\chi^2\{.\}$ is an inverse χ^2 distribution with degrees of freedom 10^{-3} and scale parameter 10^{-3}. This choice of hyperparameters gives priors that are proper but uninformative. Equation (5) models the conditional associations between $(X, \boldsymbol{C}, U)$ and Y as exchangeable. The variance parameter σ^2_β shares information between $\boldsymbol{C}$ and U. If σ^2_β is small, then this shrinks the posterior for the bias parameter β_{UY} towards zero to reflect that there is less unmeasured confounding.

Eliciting a prior for γ_{XU} and $\boldsymbol{\gamma_{CU}}$ is more challenging because the parameters describe the manner in which U is distributed within levels of X and $\boldsymbol{C}$. As described in Section 2.1, a well-known property of log linear models is the equivalence with logistic regression that arises through conditioning. The regression coefficients γ_{XU}, $\boldsymbol{\gamma_{CU}}$, $\boldsymbol{\gamma_{CX}}$, $\boldsymbol{\gamma_{C\oplus C}}$ are conditional log odds ratios for pairwise associations among the components of $(X, \boldsymbol{C}, U)$. Consequently, we assign

$$\begin{aligned} \gamma_{XU}, \boldsymbol{\gamma_{CU}}, \boldsymbol{\gamma_{CX}}, \boldsymbol{\gamma_{C\oplus C}} &\overset{\text{IID}}{\sim} \text{N}(\mu_\gamma, \sigma^2_\gamma) \\ \mu_\gamma | \sigma^2_\gamma &\sim N(0, \sigma^2_\gamma) \\ \sigma^2_\gamma &\sim \text{Inv-}\chi^2\left(10^{-3}, 10^{-3}\right), \end{aligned} \tag{6}$$

where the LHS of equation (6) refers to the individual components of γ_{XU}, $\boldsymbol{\gamma_{CU}}$, $\boldsymbol{\gamma_{CX}}$, $\boldsymbol{\gamma_{C\oplus C}}$. Equation (6) models the pairwise conditional associations among $(X, \boldsymbol{C}, U)$ as exchangeable. The parameter μ_γ is the mean log odds ratio and σ^2_γ is the variance.

Finally, we assign priors to the remaining model parameters γ_X, $\boldsymbol{\gamma_C}$ and γ_U, which are the main effects in the loglinear model of equation (4). The quantity $\exp(\gamma_U)/(1+\exp(\gamma_U))$ is the

prevalence of $U = 1$ when $X = 0$ and $C_1, \ldots, C_{21} = 0$. We model the quantities γ_X, $\boldsymbol{\gamma_C}$ and γ_U as exchangeable and assign

$$\begin{aligned} \gamma_X, \boldsymbol{\gamma_C}, \gamma_U &\overset{\text{IID}}{\sim} \text{N}(\mu_0, \sigma_0^2) \\ \mu_0|\sigma_0^2 &\sim N(0, \sigma_0^2) \\ \sigma_0^2 &\sim \text{Inv-}\chi^2\left(10^{-3}, 10^{-3}\right). \end{aligned} \tag{7}$$

2.3 Model Fitting and Computation

Denote the data as $data = \{(Y_i, X_i, \boldsymbol{C}_i); i = 1, \ldots, n\}$. If $U_1, \ldots, U_n$ were measured, then the likelihood function would be

$$\begin{aligned} &\prod_{i=1}^{n} P(Y_i|X_i, \boldsymbol{C}_i, U_i)P(X_i, \boldsymbol{C}_i, U_i) = \\ &= \prod_{i=1}^{n} \left[\frac{\exp\{Y_i(\beta_0 + \beta_{XY}X_i + \boldsymbol{\beta_{CY}}^T \boldsymbol{C}_i + \beta_{UY}U_i)\}}{1 + \exp\{\beta_0 + \beta_{XY}X_i + \boldsymbol{\beta_{CY}}^T \boldsymbol{C}_i + \beta_{UY}U_i\}} \times \right. \\ &\left. \frac{\exp\{\gamma_X X_i + \boldsymbol{\gamma_C}^T \boldsymbol{C}_i + \gamma_U U_i + \gamma_{XU}X_iU_i + \boldsymbol{\gamma_{CY}}^T \boldsymbol{C}_i U_i + \boldsymbol{\gamma_{CX}}^T \boldsymbol{C}_i X_i + \boldsymbol{\gamma_C}^T(\boldsymbol{C}_i \oplus \boldsymbol{C}_i)\}}{Q(\gamma_X, \boldsymbol{\gamma_C}, \gamma_U, \gamma_{XU}, \boldsymbol{\gamma_{CU}}, \boldsymbol{\gamma_{CX}}, \boldsymbol{\gamma_{C\oplus C}})} \right]. \end{aligned}$$

Because U is unmeasured, the likelihood for the observed data is obtained by integrating over the binary U. We obtain

$$\begin{aligned} &L(\beta_0, \beta_{XY}, \boldsymbol{\beta_{CY}}, \beta_{UY}, \gamma_X, \boldsymbol{\gamma_C}, \gamma_U, \gamma_{XU}, \boldsymbol{\gamma_{CU}}, \boldsymbol{\gamma_{CX}}, \boldsymbol{\gamma_{C\oplus C}}) \\ &= \prod_{i=1}^{n} \Big[P(Y_i|X_i, \boldsymbol{C}_i, U = 0)P(X_i, \boldsymbol{C}_i, U = 0) + \\ &\quad P(Y_i|X_i, \boldsymbol{C}_i, U = 1)P(X_i, \boldsymbol{C}_i, U = 1) \Big] \\ &= \prod_{i=1}^{n} \left[\frac{\exp\{Y_i(\beta_0 + \beta_{XY}X_i + \boldsymbol{\beta_{CY}}^T \boldsymbol{C}_i)\}}{1 + \exp\{\beta_0 + \beta_{XY}X_i + \boldsymbol{\beta_{CY}}^T \boldsymbol{C}_i\}} \times \right. \\ &\quad \frac{\exp\{\gamma_X X_i + \boldsymbol{\gamma_C}^T \boldsymbol{C}_i + \boldsymbol{\gamma_{CX}}^T \boldsymbol{C}_i X_i + \boldsymbol{\gamma_{C\oplus C}}^T(\boldsymbol{C}_i \oplus \boldsymbol{C}_i)\}}{Q(\gamma_X, \boldsymbol{\gamma_C}, \gamma_U, \gamma_{XU}, \boldsymbol{\gamma_{CU}}, \boldsymbol{\gamma_{CX}}, \boldsymbol{\gamma_{C\oplus C}})} + \\ &\quad \frac{\exp\{Y_i(\beta_0 + \beta_{XY}X_i + \boldsymbol{\beta_{CY}}^T \boldsymbol{C}_i + \beta_{UY})\}}{1 + \exp\{\beta_0 + \beta_{XY}X_i + \boldsymbol{\beta_{CY}}^T \boldsymbol{C}_i + \beta_{UY}\}} \times \\ &\quad \left. \frac{\exp\{(\gamma_X + \gamma_{XU})X_i + (\boldsymbol{\gamma_C} + \boldsymbol{\gamma_{CU}})^T \boldsymbol{C}_i + \boldsymbol{\gamma_{CX}}^T \boldsymbol{C}_i X_i + \boldsymbol{\gamma_{C\oplus C}}^T(\boldsymbol{C}_i \oplus \boldsymbol{C}_i)\}}{Q(\gamma_X, \boldsymbol{\gamma_C}, \gamma_U, \gamma_{XU}, \boldsymbol{\gamma_{CU}}, \boldsymbol{\gamma_{CX}}, \boldsymbol{\gamma_{C\oplus C}})} \right]. \end{aligned} \tag{8}$$

The posterior distribution is

$$
\begin{aligned}
&P(\beta_0, \beta_{XY}, \boldsymbol{\beta_{CY}}, \beta_{UY}, \gamma_X, \boldsymbol{\gamma_C}, \gamma_U, \gamma_{XU}, \boldsymbol{\gamma_{CU}}, \boldsymbol{\gamma_{CX}}, \boldsymbol{\gamma_{C\oplus C}}, \sigma^2_\beta, \mu_\gamma, \sigma^2_\gamma, \mu_0, \sigma^2_0 | data) \\
&\quad \propto \quad L(\beta_0, \beta_{XY}, \boldsymbol{\beta_{CY}}, \beta_{UY}, \gamma_X, \boldsymbol{\gamma_C}, \gamma_U, \gamma_{XU}, \boldsymbol{\gamma_{CU}}, \boldsymbol{\gamma_{CX}}, \boldsymbol{\gamma_{C\oplus C}}) \times \\
&\qquad P(\beta_0) P(\beta_{XY}, \boldsymbol{\beta_{CY}}, \beta_{UY} | \sigma^2_\beta) P(\gamma_{XU}, \boldsymbol{\gamma_{CU}}, \boldsymbol{\gamma_{CX}}, \boldsymbol{\gamma_{C\oplus C}} | \mu_\gamma, \sigma^2_\gamma) P(\gamma_X, \boldsymbol{\gamma_C}, \gamma_U | \mu_0, \sigma^2_0) \times \qquad (9) \\
&\qquad P(\sigma^2_\beta) P(\mu_\gamma | \sigma^2_\gamma) P(\sigma^2_\gamma) P(\mu_0 | \sigma^2_0) P(\sigma^2_0) \qquad (10)
\end{aligned}
$$

where lines (9) and (10) refer to the prior distributions for model parameters.

We sample from the posterior distribution using MCMC and the Metropolis Hastings algorithm [20]. We update sequentially from the conditional densities

$$[\beta_0, \beta_{XY}, \boldsymbol{\beta_{CY}}, \beta_{UY} | \cdot] \quad [\gamma_x, \boldsymbol{\gamma_C}, \gamma_U | \cdot] \quad [\gamma_{XU}, \boldsymbol{\gamma_{CU}}, \boldsymbol{\gamma_{CX}}, \boldsymbol{\gamma_{C\oplus C}} | \cdot]$$

$$[\sigma^2_\beta | \cdot] \quad [\mu_\gamma, \sigma^2_\gamma | \cdot] \quad [\mu_0, \sigma^2_0 | \cdot],$$

where "[|.]" means conditional on the data and remaining model parameters. We update $(\beta_0, \beta_{XY}, \boldsymbol{\beta_{CY}}, \beta_{UY})$, $(\gamma_x, \boldsymbol{\gamma_C}, \gamma_U)$ and $(\gamma_{XU}, \boldsymbol{\gamma_{CU}}, \boldsymbol{\gamma_{CX}}, \boldsymbol{\gamma_{C\oplus C}})$ using a multivariate random walk with proposal distributions that are 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 σ^2_β, $(\mu_\gamma, \sigma^2_\gamma)$ and (μ_0, σ^2_0) is straightforward in principle because they are conditionally conjugate and we can sample from the conditional distributions directly. See Section 3.1 for further discussion of computation and alternatives.

3. Analysis Results for the Beta Blocker Data

3.1 Full Bayesian Analysis

We fit the model in equations (3) and (4) to the beta blocker data and estimate the association between X and Y while adjusting for $\boldsymbol{C}$ and exploring sensitivity to the unmeasured confounder U. In the beta blocker data, we are concerned about confounding from unmeasured indications of disease. The variable U is a binary measure of the severity of heart disease, such as ejection

fraction or class of heart failure. For this reason, we fix $\boldsymbol{C}$ to include only the $q = 9$ disease indicator variables listed in Table 1 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 (3) with

$$\text{logit}[Pr(Y = 1|X, \boldsymbol{C}, \tilde{\boldsymbol{C}}, U)] \;=\; \beta_0 + \beta_{XY}X + \boldsymbol{\beta_{CY}}^T\boldsymbol{C} + \boldsymbol{\beta_{\tilde{C}Y}}^T\tilde{\boldsymbol{C}} + \beta_{UY}U, \qquad (11)$$

where "$\ldots + \boldsymbol{\beta_{\tilde{C}Y}}^T\tilde{\boldsymbol{C}} + \ldots$" refers to the 12 non-disease covariates, denoted $\tilde{\boldsymbol{C}}$, and we assign a prior to $\boldsymbol{\beta_{\tilde{C}Y}}$ that is independent normal with mean zero and variance 10^3. Furthermore, we keep the model for $P(X, \boldsymbol{C}, U)$ exactly as written in equation (4) and exclude a model for the distribution of covariates $\tilde{\boldsymbol{C}}$.

A computational challenge with our method is that sampling from the full set of parameters can be difficult because of nonidentifiability. As discussed in [2,3,19], the model for unmeasured confounding is not identifiable. This means is that different points in the parameter space give identical likelihood functions for the data. The data only minimally influence the posterior distribution for the bias parameters $\beta_{UY}, \gamma_U, \gamma_{XU}$, and $\boldsymbol{\gamma_{CU}}$, which model unmeasured confounding. Assigning informative prior distributions can improve computation, but there is still be very slow MCMC mixing, particularly if the sample size of the dataset is large. A further computational challenge is that the hierarchical prior distributions in equations (5), (6) and (7) have heavy tails. If we marginalize over the hyperparameters $(\sigma^2_\beta, \mu_\gamma, \sigma^2_\gamma, \mu_0, \sigma^2_0)$, then equation (9) assigns a t-distribution prior to each of the parameters in the sets $(\beta_{XY}, \boldsymbol{\beta_{CY}}, \beta_{UY})$,$(\gamma_{XU}, \boldsymbol{\gamma_{CU}}, \boldsymbol{\gamma_{CX}}, \boldsymbol{\gamma_{C\oplus C}})$, and $(\gamma_X, \boldsymbol{\gamma_C}, \gamma_U)$ [20]. Using prior distributions with heavy tails is problematic when fitting nonidentifiable models because the MCMC sampler may drift towards infinity.

One pragmatic solution is to speed MCMC convergence is by estimating the hyperparameters $(\sigma^2_\beta, \mu_\mu, \sigma^2_\gamma, \mu_0, \sigma^2_0)$ beforehand, and then plugging estimates into the priors in equation (9). This treats the hyperparameters as fixed and known during MCMC. It replaces t-distributions with

heavy tailed normals improving convergence. Furthermore, it reduces the dimension of the parameter space during MCMC, so that fewer parameters need to be updated at each iteration. However, a disadvantage of this approach is that it ignores uncertainty in the hyperparameters. This may give interval estimates for model parameters that are too narrow.

To illustrate, we calculate point estimates for the hyperparameters using Table 3. The quantity σ_β is the standard deviation of $\beta_{XY}, \boldsymbol{\beta_{CY}}, \beta_{UY}$. We estimate this quantity as the sample standard deviation of the log odds ratios in Section A of Table 3, which is equal to $\hat{\sigma}_\beta = 0.57$. To estimate the hyperparameters μ_γ and σ^2_γ, we compute the sample mean $\hat{\mu}_\gamma = 0.71$ and sample standard deviation $\hat{\sigma}_\gamma = 0.93$ of the log odds ratios in Section B of Table 3 (See also Figure 1). To estimate μ_0 and σ_0 we use Section C of Table 3. First, we compute the logits of prevalences of $\{X = 1\}$, $\{C_1 = 1\}, \{C_2 = 1\}, \ldots, \{C_q = 1\}$. So for example, the prevalence of COPD is $310/6969 = 0.044$, and the logit of the prevalence is $\log(0.044/(1 - 0.044)) = -3.06$. Similarly, the logit of the prevalence of cancer (CAN) is -3.96. We then compute the sample mean $\hat{\mu}_0 = -3.86$ and the sample standard deviation $\hat{\sigma}_0 = 1.22$ of the logits.

The point estimates $(\hat{\sigma}^2_\beta, \hat{\mu}_\gamma, \hat{\sigma}^2_\gamma, \hat{\mu}_0, \hat{\sigma}^2_0)$ are substituted into the priors in equation (9) so that updating of the hyperparameters is not required during MCMC. We run a single MCMC chain of length 1000000 after 100000 burn-in iterations. Sampler convergence is assessed using separate simulation runs with overdispersed starting values and the diagnostics tools included in the R package CODA [22].

The results are given in the second column of Table 1. The column has the heading "$U \not\perp\!\!\!\perp \boldsymbol{C} | X$" in order to indicate that the components of $\boldsymbol{\gamma_{CU}}$ are modelled as exchangeable with the components of $(\gamma_{XC}, \boldsymbol{\gamma_{CX}}, \boldsymbol{\gamma_{C\oplus C}})$ in equation (6), and therefore, that the analysis does not assume that $\boldsymbol{\gamma_{CU}} = \mathbf{0}$. The log odds ratio for the beta blocker effect parameter β_{XY} is -0.31 with 95% credible interval (-0.64, 0.07). This point estimate is nearly identical to that obtain from the NAIVE analysis. But the interval estimate is wider because the Bayesian analysis acknowledges uncertainty from unmeasured confounding. The prior distributions in

equation (5) assumes that the bias parameter β_{UY} has a prior mean zero and standard deviation $\hat{\sigma}_\beta = 0.57$. In other words, the sensitivity analysis assumes that U is associated with Y, given (X, C), and this association may either increase or decrease the probability of Y. Because the prior is symmetric at zero, this means that the estimated value for β_{XY} is similar to that of the Naive analysis, but the interval estimate is wider. Similar results are reported by McCandless et al. [2,3].

3.2 Assessing Prior Sensitivity for $\boldsymbol{\gamma_{CU}}$

Our modelling framework gives the opportunity to study the role of $\boldsymbol{\gamma_{CU}}$ in sensitivity analysis for unmeasured confounding. One issue is assessing the impact of the usual assumption that $\boldsymbol{\gamma_{CU}} = \mathbf{0}$. Recall from Section 1.2 that most sensitivity analysis techniques assume that measured and unmeasured confounders are uncorrelated (i.e. $U \perp\!\!\!\perp \boldsymbol{C}\,|X$) in order to reduce the burden of prior elicitation.

To study the effect of this assumption, we redo the Bayesian analysis in exactly the same way as Section 3.1, but change the prior in equation (6) to be

$$\begin{aligned} \boldsymbol{\gamma_{CU}} &= \mathbf{0} \\ \gamma_{XU}, \boldsymbol{\gamma_{CX}}, \boldsymbol{\gamma_{C\oplus C}} &\overset{\text{IID}}{\sim} \text{N}\{\hat{\mu}_\gamma, \hat{\sigma}^2_\gamma\}, \end{aligned} \tag{12}$$

where $\hat{\mu}_\gamma = 0.71$ and $\hat{\sigma}_\gamma = 0.93$. This sets each component $\boldsymbol{\gamma_{CU}}$ equal to zero and guarantees that $U \perp\!\!\!\perp \boldsymbol{C}\,|X$. However, it permits the other three bias parameters β_{UY}, γ_{UX} and γ_U to be non-zero. Thus using the prior in equation (12) instead of (6) allows U to be an unmeasured confounder for the effect of X and Y, despite the fact that $U \perp\!\!\!\perp \boldsymbol{C}\,|X$.

The results are presented in the third column of Table 1 under the heading "$U \perp\!\!\!\perp \boldsymbol{C}\,|X$". As intuition suggests, assuming that $\boldsymbol{\gamma_{CU}} = \mathbf{0}$ *increases* the posterior uncertainty about unmeasured confounding in the beta blocker data. The credible interval for the beta blocker effect in the third column ($U \perp\!\!\!\perp \boldsymbol{C}\,|X$) is nearly 15% wider than in the middle column where $U \not\perp\!\!\!\perp \boldsymbol{C}\,|X$. We have 95% interval estimates (-0.66, 0.15) versus (-0.64, 0.07). Both Bayesian analyses ac-

knowledge uncertainty from unmeasured confounding, but only the exchangeable analysis allows the possibility that U and $\boldsymbol{C}$ are correlated. The magnitude of this correlation is driven by the hyperparameter estimates $\hat{\mu}_\gamma = 0.71$ and $\hat{\sigma}_\gamma = 0.93$, which in turn are estimated from the joint distribution of X and $\boldsymbol{C}$.

To illustrate the prior sensitivity more clearly, we repeat the analysis while toying with fixed values for $\boldsymbol{\gamma_{CU}}$. Figure 2 illustrates what happens to the posterior distribution of β_{XY} when we set $\boldsymbol{\gamma_{CU}}$ equal to $\mathbf{0}$ versus $\mathbf{1}$ ... versus $\mathbf{5}$. The quantities $\mathbf{0}, \mathbf{1}, \ldots, \mathbf{5}$ denote vectors of length $p+1$. For example, $\mathbf{0} = (0, 0, \ldots, 0)$ and $\mathbf{1} = (1, 1, \ldots, 1)$. When $\boldsymbol{\gamma_{CU}} = \mathbf{5}$ then this means that the odds ratios for the conditional association between U and each of component $\boldsymbol{C}$ is equal to $\exp(5) = 148$, which corresponds to roughly perfect correlation between $\boldsymbol{C}$ and U. In Figure 2, the grey shaded region indicates the width and positioning of the Naive interval estimate for β_{XY}, which is (-0.48, -0.16). Additionally, each of the interval estimates in the figure is an interval estimate for β_{XY} that is calculated by doing a Bayesian sensitivity analysis with $\boldsymbol{\gamma_{CU}}$ locked at either $\mathbf{0}, \mathbf{1}, \ldots, \mathbf{5}$.

The key observation is that when $\boldsymbol{\gamma_{CU}}$ is large, the interval estimates collapse towards the shaded region and we obtain inferences that are essentially identical to assuming that there is no unmeasured confounding. If U and $\boldsymbol{C}$ are highly correlated, then this means that regression adjustment for $\boldsymbol{C}$ eliminates confounding from U, despite the fact that U is unmeasured. This occurs even though the prior distribution on the bias parameters β_{UY}, γ_{UX} and γ_U are non-zero. In other words, if an unmeasured confounder U is associated with the treatment and outcome, then it may nonetheless induce essentially no bias upon adjustment for $\boldsymbol{C}$, provided that $\boldsymbol{C}$ is sufficiently correlated with U.

In Figure 2, the interval estimates are shifted slightly towards zero compared to the shaded region. The reason is because of the informative prior on β_{XY} in equation (5). The prior has mean zero and variance $\hat{\sigma}_\beta = 0.57$, which tends to shrink point estimates for β_{XY} towards the zero. In contrast, the Naive interval estimate in Table 1 is computed by maximum likelihood,

which effectively presumes a flat prior on β_{XY}.

4. Discussion

Recent years have witnessed new innovation in Bayesian techniques to adjust for unmeasured confounding in observational studies. A challenge 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 for settings where the confounding effects of measured and unmeasured variables can be viewed as exchangeable within a Bayesian framework. Exchangeability captures the intuitive idea put forth by Joffe [18] that confounding from measured variables may be informative about unmeasured variables. Our method reduces the burden of prior elicitation in sensitivity analysis because it assigns priors to bias parameters without requiring that the analyst encode assumptions about each parameter individually.

Exchangeability is appealing in pharmacoepidemiology where confounding often results from a collection of homogeneous disease-related variables that influence prescribing. See Schneeweiss [1] and Joffe [18] for discussion of unmeasured confounding in pharmacoepidemiology. One example is observational studies of the effects of antidepressant drugs on adverse outcomes including infertility [23] and suicide [24]. An important confounder is depression, which is associated with antidepressant use. Because depression is difficult to measure and adjust for, it can produce a bias that is difficult to predict. However, one could argue that in typical scenarios the confounding induced by depression is indistinguishable, a priori, from the confounding of other disease-related variables that influence antidepressant use, such as cancer, alcohol abuse, sleep disorders, injury or other mental illness [24].

The assumption of exchangeability of measured and unmeasured confounders is a strong one. However it is only contingent on making judgments about the labelling of the indices of the parameters in the prior distribution (See Section 2.2). An exchangeable prior is plausible

for a set of parameters if, based on available information, we are unable to distinguish one parameter for another. For the beta blocker data, the unmeasured confounder U is either ejection fraction or class of heart failure, both of which are measures of the severity of heart disease. The confounding from U is poorly understood, and there is only limited information available in the literature to characterize the dependence between U and $\boldsymbol{C}$ (see for example [17]). A priori, we have no reason to believe that the confounding from U is any different from the confounding from the $q = 9$ disease indicator variables.

An advantage of our method is that it does not make the usual sensitivity analysis assumption that $U \perp\!\!\!\perp \boldsymbol{C} | X$ (see [1-3,6,8,9,12-14] for examples and discussion). We use a loglinear model for the joint distribution of $(X, \boldsymbol{C}, U)$, and assign a hierarchical prior distribution to the bias parameter $\boldsymbol{\gamma_{CU}}$, which models the dependence between the measured and unmeasured confounders, conditional on X. In Section 3, we show that when $\boldsymbol{\gamma_{CU}}$ is non-zero then this reduces the uncertainty from unmeasured confounding. In the beta blocker data, the convention of forcing $\boldsymbol{\gamma_{CU}} = \mathbf{0}$ for convenience gives results that are too pessimistic.

An important limitation of our method is that we substitute point estimates for the hyperparameters into the prior distribution in order to improve MCMC computation (See Section 3.1). In our experience, this substantially reduces computational time. However, it also ignores uncertainty in the hyperparameter estimates and may give interval estimates that are too narrow. This will be important if the number of measured confounders p is small. In this setting, it is difficult to characterize the patters of confounding in the dataset, and the uncertainty in hyperparameter estimates may be large. Nonetheless, pharmacoepidemiological studies often utilize healthcare administrative data with large sample sizes and rich covariate information on study participants.

References

1. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders

in epidemiologic database studies of therapeutics. *Pharmacoepidemiology and Drug Safety* 2006; **15**: 291–303.

2. McCandless LC, Gustafson P, Levy AR. Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Statistics in Medicine* 2007; **26**: 2331–47.

3. McCandless LC, Gustafson P, Levy AR. A sensitivity analysis using information about measured confounders yielded improved assessments of uncertainty from unmeasured confounding. *Journal of Clinical Epidemiology* 2008; **61**: 247–255.

4. Foody JAM, Farrell MH, Krumholz HM. β-blocker therapy in heart failure: Scientific Review. *Journal of the American Medical Association* 2002; **278**: 883–889.

5. Glynn RJ, Knight EL, Levin R, Avorn, J. Paradoxical relations of drug treatment with mortality in older persons. *Epidemiology* 2001; **12**: 682–689.

6. Rosenbaum PR, Rubin DB. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society Series B* 1983; **45**: 212–218.

7. Yanagawa T. Case-control studies: Assessing the effect of a confounding factor. *Biometrika* 1984; **71**: 191–94.

8. Lin DY, Psaty BM, Kronmal RA. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics* 1998; **54**: 948–963.

9. Greenland S. The impact of prior distributions for uncontrolled confounding and response bias. *Journal of the American Statistical Association* 2003; **98**: 47–54.

10. Jackson C, Best NG, Richardson S. Bayesian graphical models for regression on multiple datasets with different variables, *Biostatistics* 2009; **10**: 335–51.

11. Molitor N, Best NB, Jackson C, Richardson S. 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* 2009; **172**: 615–37.

12. Hernán MA, Robins JM. Letter to the editor. *Biometrics* 1999; **55**: 1316–1317.

13. VanderWeele TJ. Sensitivity analysis: Distributional assumptions and confounding assumptions. *Biometrics* 2009; **64**: 645–649.

14. Fewell Z, Smith D, Sterne JAC. The Impact of Residual and Unmeasured Confounding in Epidemiologic Studies: A Simulation Study. *American Journal of Epidemiology* 2007; **166:** 646–656.

15. Agresti A. *Categorical Data Analysis, 2nd Edition.* Wiley:New Jersey, 2002.

16. Polanczyk CA, Rohde LE, Philbin EA, Di Salvo TG. A new casemix adjustment index for hospital mortality among patients with congestive heart failure. *Medical Care* 1998; **36**: 1489–99.

17. Vasan RS, Larson MG, Benjamin EJ, Evans JC, Reiss CK, Levy D. Congestive heart failure in subjects with normal versus reduced left ventricular ejection fraction: Prevalence and mortality in a population-based cohort. *Journal of the American College of Cardiology* 1999; **33**: 1948–1955.

18. Joffe MM. Confounding by indication: The case of calcium channel blocker. *Pharmacoepidemiology and Drug Safety* 2000; **9**: 37–41.

19. Gustafson P, McCandless LC, Levy AR, Richardson SR. Simple Bayesian sensitivity analysis for mismeasured and unobserved confounders. *Biometrics (in press)* 2010.

20. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis, 2nd Edition.* Chapman Hall/CRC:New York, 2004

21. Bernardo JM, Smith AFM. *Bayesian Theory.* Wiley:New York, 1994.

22. 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.

23. Lapane KL, Zierler S, Lasater TM, Stein M, Barbour MM, Hume AL. Is a history of depressive symptoms associated with an increased risk of infertility in women? *Psychosomatic medicine* 1995; **57**: 509–13.

24. Juurlink DN, Mamdani MM, Kopp A, Redelmeier DA. The risk of suicide with selective serotonin reuptake inhibitors in the elderly. *American Journal of Psychiatry* 2006; **163**: 813–821.

Table 1: Adjusted log odds ratios (95% interval estimates) in outcome models predicting mortality.

	Naive Analysis	Bayesian Sensitivity Analysis	
		$U \not\perp\!\!\!\perp \boldsymbol{C} \mid X$	$U \perp\!\!\!\perp \boldsymbol{C} \mid X$
Beta blocker use	-0.32 (-0.48, -0.16)	-0.31 (-0.64, 0.07)	-0.28 (-0.66, 0.15)
Disease indicator variables			
CVD	0.36 (-0.30, 1.03)	0.28 (-0.41, 0.98)	0.25 (-0.41, 0.88)
COPD	0.07 (-0.20, 0.34)	0.05 (-0.33, 0.42)	0.06 (-0.22, 0.33)
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)	1.89 (1.24, 2.53)	1.94 (1.31, 2.67)
MSRD	0.63 (0.43, 0.83)	0.57 (0.19, 0.88)	0.62 (0.42, 0.82)
VENTRAR	0.35 (-0.29, 1.00)	0.25 (-0.34, 0.81)	0.22 (-0.40, 0.88)
MLD	0.53 (0.01, 1.04)	0.42 (-0.08, 0.94)	0.43 (-0.01, 0.83)
CAN	1.06 (0.68, 1.44)	0.91 (0.45, 1.37)	0.97 (0.59, 1.35)
CARS	-0.01 (-0.58, 0.56)	-0.02 (-0.61, 0.51)	-0.01 (-0.51, 0.50)
Patient characteristics			
Female sex	-0.29 (-0.4, -0.17)	-0.30 (-0.42, -0.18)	-0.30 (-0.43, -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, 1.04)
85+	1.21 (0.99, 1.43)	1.23 (1.02, 1.48)	1.25 (1.01, 1.48)
Characteristics of hospitalization			
Transferred admission	0.72 (0.55, 0.88)	0.73 (0.56, 0.90)	0.74 (0.56, 0.91)
Hospital stay (# days)	0.08 (0.05, 0.12)	0.09 (0.05, 0.13)	0.09 (0.05, 0.13)
Heart Failure medications			
Digoxin	0.04 (-0.09, 0.17)	0.03 (-0.09, 0.18)	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, 0.01)
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 (-0.45, 0.28)	-0.08 (-0.45, 0.30)
Statin	-0.16 (-0.38, 0.06)	-0.17 (-0.40, 0.05)	-0.16 (-0.40, 0.07)

Table 2: Description of variables and parameters.

Quantity	Dimension	Description
Data		
Y	1	Binary outcome indicating death within one year of hospital discharge
X	1	Binary treatment indicating being dispensed a beta blocker within 30 days of hospital discharge
$\boldsymbol{C}$	p	Vector of p measured confounding variables
U	1	Binary unmeasured confounding variable
Parameters		
β_0	1	y-intercept in outcome model
β_{XY}	1	Causal log odds ratio for effect of X on Y conditional on $(\boldsymbol{C}, U)$
β_{UY}	1	Log odds ratio for association between U and Y given $(X, \boldsymbol{C})$
$\boldsymbol{\beta_{CY}}$	p	Log odds ratios for association between $\boldsymbol{C}$ and Y given (X, U)
γ_U	1	Main effect for U in equation (4). Note that $\frac{\exp(\gamma_U)}{1+\exp(\gamma_U)}$ is the prevalence of $U = 1$ when $X = 0$ and $C_1, \ldots, C_p = 0$
γ_X	1	Main effect for X in equation (4)
$\boldsymbol{\gamma}_C$	p	Main effects for $\boldsymbol{C}$ in equation (4)
γ_{XU}	p	Log odds ratio for association between X and U given $\boldsymbol{C}$
$\boldsymbol{\gamma_{CU}}$	p	Log odds ratios for association between $\boldsymbol{C}$ and U given X
$\boldsymbol{\gamma_{CX}}$	p	Log odds ratios for association between $\boldsymbol{C}$ and X given U
$\boldsymbol{\gamma_{C\oplus C}}$	$\binom{p}{2}$	Log odds ratios for pairwise conditional associations among components of $\boldsymbol{C}$
Hyperparameters		
σ^2_β	1	Prior variance of β_{XY}, $\boldsymbol{\beta_{CY}}$, β_{UY}
μ_γ	1	Prior mean of γ_{XU}, $\boldsymbol{\gamma_{CU}}$, $\boldsymbol{\gamma_{CX}}$, $\boldsymbol{\gamma_{C\oplus C}}$
σ^2_γ	1	Prior variance of γ_{XU}, $\boldsymbol{\gamma_{CU}}$, $\boldsymbol{\gamma_{CX}}$, $\boldsymbol{\gamma_{C\oplus C}}$
μ_0	1	Prior mean of γ_X, $\boldsymbol{\gamma_C}$, γ_U
σ^2_0	1	Prior variance of γ_X, $\boldsymbol{\gamma_C}$, γ_U

Table 3: Description of the confounding induced by the $q = 9$ disease indicator variables in the association between X and Y.

Section A: Adjusted log odds ratios (standard errors) copied from the Naive Analysis of Table 1 that describe the association between selected components of $\boldsymbol{C}$ and Y given X.

Beta Blocker	CVD	COPD	HYPNAT	MTSTD	MSRD	VENTRAR	MLD	CAN	CARS
-0.32 (0.08)	0.36 (0.3)	0.07 (0.1)	0.10 (0.2)	1.79 (0.3)	0.63 (0.1)	0.35 (0.3)	0.53 (0.3)	1.06 (0.2)	-0.01 (0.3)

Section B: Log odds ratios (standard errors) for the $\binom{10}{2} = 45$ pairwise conditional associations among treatment X and the $q = 9$ binary disease 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
CAN	.	.	.	.	.	.	.	.	.	0.6 (1.0)
CARS	.	.	.	.	.	.	.	.	.	.

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 2.

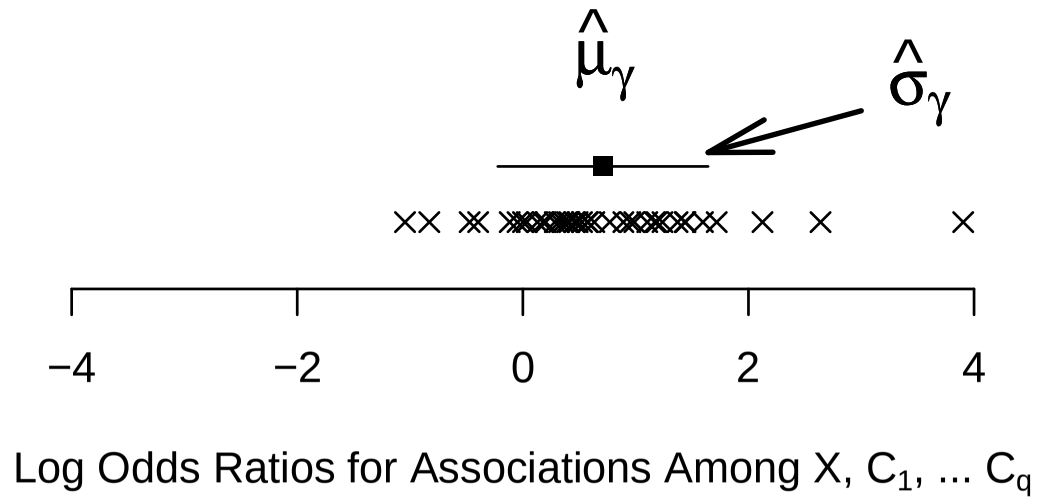

Figure 2: 95% credible intervals for the beta blocker effect β_{XY} when we repeat the Bayesian sensitivity analysis with $\boldsymbol{\gamma_{CU}}$ held fixed at **0** versus **1** versus **2** ... versus **5**. The grey shaded region indicates the width and position of the Naive interval estimate for β_{XY}, which is (-0.48, -0.16), and taken from Table 1.