STATISTICS IN MEDICINE
Statist. Med. 2007; **26**:2331–2347
Published online 22 September 2006 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.2711

Bayesian sensitivity analysis for unmeasured confounding in observational studies

Lawrence C. McCandless[1,*,†], Paul Gustafson[1] and Adrian Levy[2]

[1]*Department of Statistics, University of British Columbia, Vancouver BC, Canada*
[2]*Department of Healthcare and Epidemiology, University of British Columbia, Vancouver BC, Canada*

SUMMARY

We consider Bayesian sensitivity analysis for unmeasured confounding in observational studies where the association between a binary exposure, binary response, measured confounders and a single binary unmeasured confounder can be formulated using logistic regression models. A model for unmeasured confounding is presented along with a family of prior distributions that model beliefs about a possible unknown unmeasured confounder. Simulation from the posterior distribution is accomplished using Markov chain Monte Carlo. Because the model for unmeasured confounding is not identifiable, standard large-sample theory for Bayesian analysis is not applicable. Consequently, the impact of different choices of prior distributions on the coverage probability of credible intervals is unknown. Using simulations, we investigate the coverage probability when averaged with respect to various distributions over the parameter space. The results indicate that credible intervals will have approximately nominal coverage probability, on average, when the prior distribution used for sensitivity analysis approximates the sampling distribution of model parameters in a hypothetical sequence of observational studies. We motivate the method in a study of the effectiveness of beta blocker therapy for treatment of heart failure. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: unmeasured confounding; bias; coverage probability; observational studies; identifiability

1. INTRODUCTION

Consider a retrospective cohort study designed to estimate the effectiveness of beta blocker therapy for treatment of heart failure using linked records of hospital episodes, prescription claims data, and death certificates. An important limitation is that the data do not contain detailed clinical information on patient characteristics that may influence prescribed therapy [1, 2]. Patients who are prescribed

*Correspondence to: Lawrence C. McCandless, Department of Statistics, University of British Columbia, 333-6356 Agricultural Rd., Vancouver BC, Canada V6E 1Z2.
†E-mail: lawrence@stat.ubc.ca

Contract/grant sponsor: Michael Smith Foundation Junior Graduate Studentship Award

Received 18 July 2005
Accepted 3 August 2006

a beta blocker when discharged from hospital may differ systematically from patients without such therapy with respect to clinical variables that are not adequately measured. Consequently, the estimated effect of beta blocker therapy on mortality may be biased from unmeasured confounding even after adjusting for measured covariates [3].

A solution is to conduct a sensitivity analysis for unmeasured confounding using Bayesian techniques [4–8]. The model for the relationship between exposure to beta blocker therapy, measured covariates and the outcome is expanded to include bias parameters which reflect the investigators' beliefs about a binary unmeasured confounder. The resulting model is typically non-identifiable. External information about unmeasured confounding is incorporated into the analysis as prior distributions on bias parameters. The posterior distribution of the exposure effect summarizes uncertainty due to unmeasured confounding in addition to random error.

A difficulty with this approach is that the impact of different choices of prior distributions on the performance of interval estimators for the exposure effect is unknown. The method involves Bayesian analysis using a non-identifiable model. Depending on the prior distribution, the posterior mean may be asymptotically biased and credible intervals will not have nominal large sample coverage probability [9]. Hence without identifiability, standard large-sample theory does not guarantee good performance of interval estimators regardless of the choice of prior distributions. This is not a deficiency of Bayesian methods. Without identifiability, no statistical method will afford interval estimates which have correct frequentist coverage levels. While Bayesian approaches to sensitivity analysis for unmeasured confounding are discussed in several recent works [4–8], there has been no investigation of the frequentist performance of interval estimators while addressing the role of the prior distribution used for sensitivity analysis.

We address this limitation in the context of a real data example estimating the effectiveness of beta blocker therapy for treatment of heart failure using healthcare administrative data. In Section 2, we present a method of Bayesian sensitivity analysis (BSA) for unmeasured confounding for observational studies where the association between a binary response, binary exposure, measured confounders and a single binary unmeasured confounder can be formulated using logistic regression models. We propose a family of prior distributions which model beliefs about a confounder which is both unmeasured and unknown. In Section 3, we apply our method to the heart failure data set. In Section 4, we conduct a simulation study to investigate the relationship between the prior distribution used for sensitivity analysis and the frequentist performance of interval estimators for the exposure effect. Following work by Gustafson [9] on Bayesian inference using non-identifiable models, we estimate the coverage probability of interval estimators when averaged with respect to various distributions over the parameter space. We demonstrate that interval estimators will have desirable frequentist properties when averaged over a hypothetical sequence of observational studies provided that the prior distribution used for sensitivity analysis approximates the sampling distribution of model parameters in the sequence of studies. Section 5 concludes with a summary and discussion.

2. BAYESIAN SENSITIVITY ANALYSIS (BSA) FOR UNMEASURED CONFOUNDING

2.1. A model for unmeasured confounding

Let X, Y and U denote dichotomous random variables taking values 1 or 0 to indicate the presence or absence of an exposure, disease or unmeasured confounder, respectively. Let Z denote a $p \times 1$ vector of measured covariates.

DOI: 10.1002/sim

Following Lin *et al.* [10], we use the factorization $P(Y, U|X, Z) = P(Y|X, U, Z)P(U|X, Z)$ and model the confounding effect of U using logistic regression models:

$$\text{logit}[P(Y = 1|X, U, Z)] = \beta_0 + \beta_1 X + \lambda U + \eta' Z \tag{1}$$

$$\text{logit}[P(U = 1|X, Z)] = \gamma_0 + \gamma_1 X + \xi' Z \tag{2}$$

The variables X and U are assumed to not interact in their effect on Y. Extending equation (1) to include interactions is straightforward, but creates ambiguity about the targets of inference. Greenland [4] discusses these issues and argues that excluding interaction terms can typically be justified in this context because it simplifies modelling and because their inclusion only modestly affects inferences from sensitivity analysis.

As with other models for unmeasured confounding, the model is non-identifiable and indexed by the bias parameters γ_0, γ_1, λ and ξ, which reflect assumptions about the confounding effect of U. The quantity $\exp(\lambda)$ models the odds ratio of the association between U and Y given X and Z, while the quantity $\exp(\gamma_1)$ models the odds ratio of the association between U and X given Z. The parameter ξ models the correlation between U and Z given X. Non-identifiability is apparent from the fact that when $\lambda = 0$ the observed risk of Y within strata of X and Z, given by $P(Y = 1|X, Z) = P(Y = 1|X, U = 1, Z)P(U = 1|X, Z) + P(Y = 1|X, U = 0, Z)P(U = 0|X, Z)$, does not depend on γ_0, γ_1 or ξ. Different parameter values may yield identical likelihood functions for the observed data.

2.2. *Prior specification*

In principal, a prior distribution $f(\beta_0, \beta_1, \lambda, \eta, \gamma_0, \gamma_1, \xi)$ from any standard parametric family can be used for BSA. Because our intent is to model prior beliefs about unmeasured confounding, we consider the case where $f(\beta_0, \beta_1, \lambda, \eta, \gamma_0, \gamma_1, \xi) = f(\lambda, \gamma_0, \gamma_1, \xi)f(\beta_0, \beta_1, \eta)$ and restrict our attention to specifying $f(\lambda, \gamma_0, \gamma_1, \xi)$. We propose a family of priors distributions to model the effect of a confounding variable U which is both unmeasured and unknown.

For the bias parameter λ, we assign

$$f(\lambda) \propto \exp\left\{-\frac{\lambda^2}{2c_1^2}\right\} \tag{3}$$

to model prior beliefs about the association between U and Y. This normal density is symmetric and centred at zero to reflect the fact that because U is unknown, little can be said about whether λ is positive or negative. But this does not preclude prior assertions about the possible magnitude of λ. Depending on the investigation, it may be unlikely that U is a strong predictor of the disease. Consequently, we specify $c_1 > 0$ to limit the association between U and Y. For example, if the investigator felt that λ might be large in magnitude, then she might specify $c_1 = \log(6)/1.96$. This would model the prior belief that the log-odds ratio of the association between U and Y is between $\pm\log(6)$ with probability 95 per cent.

A prior distribution for γ_0 and γ_1 is assigned as a bivariate normal distribution

$$f(\gamma_0, \gamma_1) \propto \exp\left\{-\frac{1}{2}\begin{pmatrix}\gamma_0\\ \gamma_1\end{pmatrix}'\begin{pmatrix}c_2^2 & -c_3^2/2\\ -c_3^2/2 & c_3^2\end{pmatrix}^{-1}\begin{pmatrix}\gamma_0\\ \gamma_1\end{pmatrix}\right\} \tag{4}$$

DOI: 10.1002/sim

where $c_2, c_3 > 0$ and $2c_2 > c_3$ so that the covariance matrix is positive definite. We specify c_3 to limit the possible associations between U and X. Additionally, the prior distribution limits the prevalence of U within exposure groups, denoted $P(U=1|X)$. To see this, note that

$$P(U=1|X,Z) = \frac{1}{1+\exp\{-(\gamma_0+\gamma_1 X+\xi' Z)\}}$$
$$\approx \frac{1}{1+\exp\{-(\gamma_0+\gamma_1 X)\}} + \frac{\exp\{-(\gamma_0+\gamma_1 X)\}}{[1+\exp\{-(\gamma_0+\gamma_1 X)\}]^2}\xi' Z$$

based on a first order Taylor expansion of $P(U=1|X,Z)$ around $E\{Z\}$, where the components of Z have been re-scaled to have mean zero and unit variance. Hence $P(U=1|X) = E\{P(U=1|X,Z)|X\} \approx (1+\exp(-(\gamma_0+\gamma_1 X)))^{-1}$, or equivalently, $\text{logit}[P(U=1|X)] \approx \gamma_0 + \gamma_1 x$. Now observe that based on in equation (4), we have $\text{Var}\{\gamma_0+\gamma_1 X\} = c_2^2$ for $X=0, 1$ because $\text{Var}\{\gamma_0\} = c_2^2$ and $\text{Var}\{\gamma_0+\gamma_1\} = \text{Var}\{\gamma_0\} + \text{Var}\{\gamma_1\} + 2\,\text{Cov}\{\gamma_0, \gamma_1\} = c_2^2 + c_1^2 - 2c_1^2/2 = c_2^2$. Consequently, c_2 limits the magnitude of $\text{logit}[P(U=1|X)]$. If the user specified $c_2 = \log(6)/1.96$, this would model the belief that $\text{logit}[P(U=1|X)]$ lies between $\pm\log(6)$ with probability 95 per cent.

We assign

$$f(\xi) \propto \exp\left\{-\frac{\xi'\xi}{2c_4^2}\right\} \tag{5}$$

where c_4 limits the associations between Z and U. For example, setting $c_4 = \log(6)/1.96$ implies that a unit change of one standard deviation of any component Z changes the log-odds of U by at most $\pm\log(6)$ with probability 95 per cent. If U was thought to be correlated with C, then a larger value of c_4 might be suitable. However, because Z is re-scaled to have unit variance, the interpretation of ξ depends on the variance of Z.

To improve posterior simulation, we may further restrict γ_1 to be greater than zero without affecting the resulting posterior distribution. The reason is that, because U is unmeasured, the labelling $U=1$ or 0 is arbitrary, and we may assume that U is more prevalent in exposed subjects. If U were more prevalent among unexposed subjects, we could define a new unmeasured confounder $U^* = 1-U$ and switch the sign of λ. The resulting variable U^* would have the same confounding effect of U and would be more prevalent in exposed subjects.

2.3. Posterior simulation

Let y, x and u denote vectors of length n of the observed responses, exposures and unmeasured confounders in n study subjects and let z denote an $n \times p$ matrix of measured covariates. Our objective is to study $f(\beta_0, \beta_1, \lambda, \eta, \gamma_0, \gamma_1, \xi|y, x, z)$, the posterior distribution of model parameters given y, x and z and prior assumptions about unmeasured confounding. To accomplish this, we sample from the posterior using Markov chain Monte Carlo (MCMC). To illustrate the main ideas, observe that if u were measured then posterior inferences could be accomplished by fitting the logistic regression models in equations (1) and (2). Therefore, posterior simulation may proceed using two Gibbs sampling steps by treating u as missing data that is integrated out of the joint posterior distribution $f(\beta_0, \beta_1, \lambda, \eta, \gamma_0, \gamma_1, \xi, u|y, x, z)$ [11]. The imputation step samples the latent u from $f(u|\beta_0, \beta_1, \lambda, \eta, \gamma_0, \gamma_1, \xi, y, x, z)$. The posterior step samples from the posterior distributions

DOI: 10.1002/sim

from Bayesian logistic regressions of y on x, u and z and from u on x and z. Details are given in Appendix A.

3. ESTIMATING THE EFFECTIVENESS OF BETA BLOCKER THERAPY FOR TREATMENT OF HEART FAILURE

Returning to the example described in the introduction, we apply BSA to estimate the effect of beta blockers on 1 year all-cause mortality among elderly residents of Vancouver BC, who have suffered an acute exacerbation of heart failure. Hospital discharge summaries, prescription claims records and death certificates were obtained from the British Columbia Linked Health Database [12]. The study sample included all 945 Vancouver residents, aged 65 years or older, who were discharged alive with a primary discharge diagnosis of heart failure (International Classification of Disease ninth revision (ICD9) code 428) between April 1999 and March 2001. Only patients with a new hospitalization for heart failure were included. Letting X denote exposure to beta blockers, we set X equal to one if the subject was dispensed a beta blocker within 30 days of discharge from hospital and zero otherwise. The response variable Y was set equal to one if the subject died within 1 year of discharge from hospital and zero otherwise. Measured covariates included age (categorical with three levels; 65–74, 75–84, $\geqslant$85 years), sex (binary with 1 indicating female and zero otherwise), Polynczyk congestive heart failure casemix index [13] (continuous), hospital length of stay in unit number of 10 day intervals (continuous), and contraindications (binary with 1 indicating presence of asthma or chronic obstructive pulmonary disorder (ICD9 code 491.0–493.9) or diabetes (ICD9 code 250)).

We begin by estimating the effectiveness of beta blocker therapy while adjusting for measured covariates, but ignoring possible unmeasured confounding. We fit a logistic regression model of Y on X stratifying on quintiles of the estimated propensity score [14]. Thus we let Z be a 4×1 vector with components taking values of one or zero to indicate quintile group membership. The propensity scores are estimated using the fitted values from logistic regression of X on the measured covariates with main effects and no interactions. To assess the balancing properties of the propensity scores, we examine the distribution of measured covariates among exposed and unexposed patients within quintile groups. The results are presented in Table I. Differences in the covariate distributions are compared using t-tests and chi-square tests [14]. While some imbalances are observed, none are statistically significant, and this indicates that exposed and unexposed patients are fairly comparable within quintile groups. The analysis yields a log odds ratio of -0.69 with 95 per cent two-tailed credible interval $(-1.06, -0.33)$ for the effect of X on Y, suggesting that beta blocker therapy is associated with a significant reduction in odds of mortality.

Because of the lack of clinical information in healthcare administrative data, there are concerns that the treatment effect estimate may be biased because it ignores possible unmeasured confounders such as indications of disease severity [3]. We apply BSA to assess the sensitivity of the treatment effect estimates to confounding from a possible unknown and unmeasured binary confounder. We let Z from equations (1) and (2) be a 4×1 vector of indicator variables indicating membership in propensity score quintiles. We assign a diffuse independent multivariate normal prior for (β_0, β_1, η), and we consider several prior distributions for the bias parameters $(\lambda, \gamma_0, \gamma_1, \xi)$ based on the family specified in Section 2.2. We assign $c_1 = c_3 = c_4$, and we let c_1, c_2, c_3 and c_4 equal $\log(3)/1.96$ or $\log(6)/1.96$. This corresponds to priors where λ, γ_1 and each component of ξ are normally

DOI: 10.1002/sim

Table I. Number (%) or mean ± standard deviation for measured covariates among exposed and unexposed patients, within quintiles of the propensity score.

	Quintile group #1		Quintile group #2		Quintile group #3		Quintile group #4		Quintile group #5	
	Beta blocker		Beta blocker		Beta blocker		Beta blocker		Beta blocker	
	Yes $n=41$	No $n=148$	Yes $n=40$	No $n=149$	Yes $n=47$	No $n=142$	Yes $n=48$	No $n=141$	Yes $n=58$	No $n=131$
Sex	21 (51)	92 (62)	21 (52)	85 (57)	21 (45)	77 (54)	26 (54)	73 (52)	34 (59)	71 (54)
Age 75–84	19 (46)	65 (44)	14 (35)	53 (36)	19 (40)	64 (45)	20 (42)	54 (38)	25 (43)	61 (47)
Age 84+	12 (29)	57 (39)	11 (28)	53 (36)	16 (34)	47 (33)	20 (42)	47 (33)	19 (33)	45 (34)
Contraindications	12 (29)	37 (25)	11 (28)	42 (28)	12 (26)	32 (23)	13 (27)	35 (25)	18 (31)	39 (30)
Casemix index	5.7 ± 2	5.8 ± 2	5.8 ± 4	5.7 ± 2	5.9 ± 3	5.9 ± 3	5.9 ± 2	5.9 ± 3	5.6 ± 2	6.2 ± 3
Hospital length of stay	0.86 ± 0.9	1.0 ± 1	1.0 ± 0.9	0.93 ± 1	0.78 ± 1	1.0 ± 1	0.86 ± 0.9	1.1 ± 1	0.82 ± 0.8	1.2 ± 2

Table II. Posterior means (95% two-tailed credible intervals) of log-odds ratios of the beta blocker effect β_1 from BSA stratifying on quintiles of the propensity score.

		$c_1 = c_3 = c_4 = \frac{\log(3)}{1.96}$		$c_1 = c_3 = c_4 = \frac{\log(6)}{1.96}$	
	$c_1 = c_2 = c_3 = c_4 = 0$	$c_2 = \frac{\log(3)}{1.96}$	$c_2 = \frac{\log(6)}{1.96}$	$c_2 = \frac{\log(3)}{1.96}$	$c_2 = \frac{\log(6)}{1.96}$
Beta blocker	−0.69 (−1.06, −0.33)	−0.69 (−1.09, −0.31)	−0.69 (−1.09, −0.31)	−0.68 (−1.15, −0.20)	−0.67 (−1.12, −0.21)

DOI: 10.1002/sim

distributed, and lie between $\pm\log(k)$ with probability 95 per cent, for $k=3, 6$. The selected values of c_2 correspond to logit$[P(U=1|X)]$, for $X=0, 1$, lying approximately between logit$[1/(k+1)]$ and logit$[k/(k+1)]$ with probability 95 per cent, for $k=3, 6$. We also study the limiting case where c_1, c_2, c_3, c_4 equal zero. In this case, the prior distributions for $\lambda, \gamma_0, \gamma_1, \xi$ are point masses at 0, and BSA corresponds to logistic regression of Y on X and Z ignoring unmeasured confounding. In each analysis, the components of Z are rescaled to have mean zero and unit variance, and a Markov chain of length 500 000 iterations is obtained with the 10 000 initial iterations discarded. Convergence is assessed using the method of Gelman and Rubin [15].

Table II gives posterior means and 95 per cent two-tailed credible intervals for the log odds ratios of the beta blocker effect β_1 for different c_1, c_2, c_3 and c_4. Posterior density estimates are plotted in Figure 1. The results indicate that the protective effect of beta blocker exposure on mortality is robust to a variety of assumptions about unmeasured confounding. When $c_1=c_3=c_4=0$, BSA is equivalent to logistic regression of Y on X and Z ignoring unmeasured confounding. Larger values of $c_1=c_3=c_4$ yield modest increases in the length of credible intervals. This is depicted by the flattening of the density functions in Figure 1. These results are consistent with intuition because they correspond to greater uncertainty about the magnitude of unmeasured confounding. The parameter c_2 does not appear to greatly affect credible intervals.

For comparison, we contrast BSA with the method of sensitivity analysis for unmeasured confounding proposed by Rosenbaum and Rubin [16]. The method of Rosenbaum and Rubin involves fixing the bias parameters and then maximizing the likelihood of the observed data, which has been integrated over the distribution of unmeasured confounder. For the beta blocker

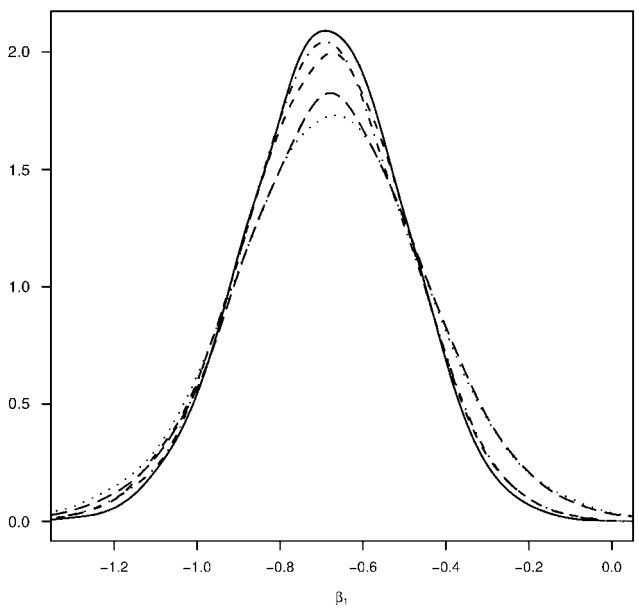

Figure 1. Posterior distribution of the log-odds ratio of the beta blocker effect β_1 from BSA stratifying on quintiles of the propensity score. The solid curve refers to the naive analysis with $c_1=c_2=c_3=c_4=0$ ignoring unmeasured confounding. The dashed curves refer to the various non-zero combinations of c_1, c_2, c_3 and c_4.

DOI: 10.1002/sim

Table III. Point estimates of the beta blocker effect β_1 as a function of the bias parameters λ, γ_1 and γ_1 from sensitivity analysis using the method of Rosenbaum and Rubin [16].

		λ				
γ_1	γ_0	$-\frac{\log(6)}{1.96}$	$-\frac{\log(3)}{1.96}$	0	$\frac{\log(3)}{1.96}$	$\frac{\log(6)}{1.96}$
0	$-\frac{\log(6)}{1.96}$	−0.71	−0.70	−0.69	−0.70	−0.72
	$-\frac{\log(3)}{1.96}$	−0.71	−0.70	−0.69	−0.70	−0.72
	0	−0.72	−0.70	−0.69	−0.70	−0.72
	$\frac{\log(3)}{1.96}$	−0.72	−0.70	−0.69	−0.70	−0.71
	$\frac{\log(6)}{1.96}$	−0.72	−0.70	−0.69	−0.70	−0.71
$\frac{\log(3)}{1.96}$	$-\frac{\log(6)}{1.96}$	−0.60	−0.63	−0.69	−0.77	−0.84
	$-\frac{\log(3)}{1.96}$	−0.59	−0.62	−0.69	−0.78	−0.85
	0	−0.59	−0.62	−0.69	−0.78	−0.84
$\frac{\log(6)}{1.96}$	$-\frac{\log(6)}{1.96}$	−0.53	−0.58	−0.69	−0.82	−0.92
	$-\frac{\log(3)}{1.96}$	−0.51	−0.58	−0.69	−0.83	−0.92
	0	−0.51	−0.58	−0.69	−0.82	−0.90

Standard errors equal 0.19 for all treatment effect estimates.

data, this calculation involves maximizing the likelihood function

$$L(\beta_0, \beta_1, \eta) = \prod_{i=1}^{n} [f(y_i|x_i, U=1, z_i) f(U=1|x_i, z_i) + f(y_i|x_i, U=0, z_i) f(U=0|x_i, z_i)]$$

$$= \prod_{i=1}^{n} \left[\frac{\exp\{y_i(\beta_0 + \beta_1 x_i + \lambda + \eta' z_i)\}}{1 + \exp\{\beta_0 + \beta_1 x_i + \lambda + \eta' z_i\}} \frac{\exp\{\gamma_0 + \gamma_1 x_i + \xi' z_i\}}{1 + \exp\{\gamma_0 + \gamma_1 x_i + \xi' z_i\}} \right.$$

$$\left. + \frac{\exp\{y_i(\beta_0 + \beta_1 x_i + \eta' z_i)\}}{1 + \exp\{\beta_0 + \beta_1 x_i + \eta' z_i\}} \frac{1}{1 + \exp\{\gamma_0 + \gamma_1 x_i + \xi' z_i\}} \right]$$

over the parameters (β_0, β_1, η), for fixed bias parameters $(\lambda, \gamma_0, \gamma_1, \xi)$.

Table IIIpresents the results of the sensitivity analysis using the method of Rosenbaum and Rubin. Each cell of Table III contains a point estimate of β_1 for different values of $(\lambda, \gamma_0, \gamma_1)$. The components of the bias parameters ξ are set equal to zero to ease the presentation of results. For example, consider the bottom right cell in Table III. We have $\lambda = \gamma_1 = \log(6)/1.96$. This corresponds to the instance where U increases the risk of Y and is more common among patients

DOI: 10.1002/sim

with $X=1$. Hence we believe that the naive odds ratio ignoring unmeasured confounding is attenuated towards zero. The sensitivity analysis yields $\hat{\beta}_1=-0.90$ which is driven away from zero. When λ or γ_1 equals zero there is no unmeasured confounding. We do not consider $\gamma_1<0$. As was emphasized in Section 2.2, because U is unmeasured, the labelling of U equal to one or zero is arbitrary. Therefore, we can assume that the event $U=1$ is more common among exposed subjects. In Table III, standard errors for the point estimates are obtained from the inverse of the observed information.

A limitation of the method of sensitivity analysis of Rosenbaum and Rubin is that the results are presented in large tables. Standard errors are reported separately. The user must qualitatively synthesize bias uncertainty with variability due to random sampling. Consequently the results of the sensitivity analysis may be difficult to interpret, particularly when there are many bias parameters. For the heart failure data set there are a total of seven bias parameters (λ, γ_0, γ_1 and ξ_1, ξ_2, ξ_3, ξ_4). In contrast, BSA incorporates prior information about bias parameters and yields a single interval estimate for the exposure effect which summarizes uncertainty from bias and random sampling.

The BSA methodology can alternatively be used in conjunction with direct regression on measured covariates rather than on quintiles of the propensity score. To illustrate, we let Z from equations (1) and (2) equal a 6×1 vector of measured covariates (sex, age 75–84, age 84+, contraindications, casemix index, and hospital length of stay). Table IV gives posterior means and 95 per cent two-tailed credible intervals for the log odds ratios of the beta blocker effect and covariate effects for different c_1, c_2, c_3 and c_4. For the beta blocker effect, the results are similar to those of Table II obtained from the regression on propensity score quintiles. Interestingly, large values of $c_1=c_3=c_4$ yield wider posterior distributions for the other regression coefficients. This is not unexpected because allowing ξ to be non-zero in equation (2) implies that U is an unmeasured confounder for the effect of Z on Y.

While the credible intervals for the exposure effect in Tables II and IV widen with increasing c_1, c_2, c_3 and c_4, all of the interval estimates exclude zero, indicating that the protective effect of beta blocker exposure on mortality is insensitive to a variety of assumptions about unmeasured confounding. In epidemiologic research on therapeutics using healthcare administrative data, uncertainty from random error is often small in relation to uncertainty about bias [2]. But the results of this analysis indicate that in certain settings, large prior uncertainty about possible bias does not necessarily translate to large increases in posterior uncertainty about exposure effects. The combination of large sample size and large exposure effects may yield robustness of effect estimates to pessimistic assumptions about unmeasured confounding. It is emphasized that only one source of bias has been investigated. Other sources of bias, such as selection bias or information bias, can distort the measure of association [2].

4. A SIMULATION STUDY OF THE FREQUENTIST PERFORMANCE OF INTERVAL ESTIMATORS FROM BSA

4.1. Average coverage probability of interval estimators

An issue in interpreting the analysis of the beta blocker data is that we do not know how the choice of prior distribution used for sensitivity analysis affects the performance of interval estimates. For identifiable models, large sample theory guarantees that under standard regularity conditions the

DOI: 10.1002/sim

Table IV. Posterior means (95% two-tailed credible intervals) of log-odds ratios of the beta blocker effect β_1 and covariate effects from BSA stratifying on measured covariates.

		$c_1 = c_3 = c_4 = \frac{\log(3)}{1.96}$		$c_1 = c_3 = c_4 = \frac{\log(6)}{1.96}$	
	$c_1 = c_2 = c_3 = c_4 = 0$	$c_2 = \frac{\log(3)}{1.96}$	$c_2 = \frac{\log(6)}{1.96}$	$c_2 = \frac{\log(3)}{1.96}$	$c_2 = \frac{\log(6)}{1.96}$
Beta blocker	−0.75 (−1.16, −0.36)	−0.76 (−1.17, −0.35)	−0.76 (−1.18, −0.35)	−0.76 (−1.26, −0.28)	−0.76 (−1.25, −0.28)
Sex	−0.30 (−0.61, −0.01)	−0.31 (−0.72, 0.09)	−0.30 (−0.71, 0.09)	−0.31 (−1.01, 0.41)	−0.30 (−1.04, 0.43)
Age 75–84	0.65 (0.21, 1.10)	0.66 (0.15, 1.19)	0.66 (0.14, 1.19)	0.67 (−0.17, 1.49)	0.67 (−0.18, 1.50)
Age 84+	1.02 (0.56, 1.50)	1.02 (0.48, 1.56)	1.02 (0.48, 1.57)	0.98 (0.24, 1.71)	0.99 (0.29, 1.67)
Contraindications	0.31 (−0.05, 0.66)	0.32 (−0.12, 0.77)	0.32 (−0.11, 0.77)	0.31 (−0.27, 0.93)	0.31 (−0.27, 0.91)
Casemix index	0.22 (0.16, 0.29)	0.22 (0.15, 0.30)	0.22 (0.15, 0.30)	0.23 (0.13, 0.33)	0.23 (0.13, 0.33)
Hospital length of stay	−0.01 (−0.12, 0.10)	−0.01 (−0.15, 0.11)	−0.01 (−0.15, 0.11)	−0.03 (−0.20, 0.13)	−0.03 (−0.21, 0.13)

DOI: 10.1002/sim

posterior mean is a consistent estimator of model parameters and interval estimators have nominal large sample coverage probability [11]. But BSA involves Bayesian analysis using a non-identifiable model. Consequently, depending on the prior used for sensitivity analysis, the posterior mean may be asymptotically biased and interval estimators will not have nominal large sample coverage probability [9].

When coverage probability varies over the parameter space, one approach to characterizing the performance of interval estimators is to study the average coverage probability [17, 18]. Specifically, let $I(y, x, z)$ denote a $1-\alpha$ credible interval for the exposure effect β_1 constructed using the prior distribution $f(\theta)$. The coverage probability of $I(y, x, z)$, for fixed $\theta=(\beta_0, \beta_1, \lambda, \eta, \gamma_0, \gamma_1, \xi)$, is

$$C_f(\theta)=\int 1\{\beta_1 \in I(y, x, z)\} f(y, x, z|\theta)\, \mathrm{d}(y, x, z)$$

where $1\{\beta_1 \in I(y, x, z)\}$ is the indicator function which takes value one if $\beta_1 \in I(y, x, z)$ and zero otherwise. The average coverage probability of $I(y, x, z)$, with respect to some density $g(\theta)$, is given by

$$\Omega_{g,f}=\int C_f(\theta) g(\theta)\, \mathrm{d}\theta$$

It is well known that if the density $g(\theta)$ is equal to the prior density $f(\theta)$ used to construct $I(y, x, z)$, then $\Omega_{g,f}=1-\alpha$ [17, 18]. Moreover, this property holds irrespective of sample size and whether the model is identifiable or not. Therefore, if the prior distribution used for BSA equals the distribution $g(\theta)$ used to calculate $\Omega_{g,f}$, then $1-\alpha$ credible intervals will have the appealing property that the long run average coverage under repeated sampling of data and parameters from $f(y, x, z|\theta)g(\theta)$ is equal to $1-\alpha$. This reasoning can be used to describe the frequentist properties of BSA credible intervals in a hypothetical sequence of observational studies. If the data in a sequence of observational studies are obtained by first sampling θ from $g(\theta)$ and then sampling data given θ, then BSA credible intervals will have nominal coverage probability, on average, provided that $f(\theta)=g(\theta)$. Thus while BSA involves Bayesian analysis using a non-identifiable model, interval estimators may be good frequentist properties depending on the prior distribution used for sensitivity analysis.

Typically, when analysing observational data, the model $f(y, x, z|\theta)$ is assumed to be known, but parameters themselves and the distribution $g(\theta)$ used to sample parameters are unknown. Therefore, it would appear that there is no guarantee of choosing a prior distribution $f(\theta)$ such that $f(\theta)=g(\theta)$. However, the investigator may have general *a priori* knowledge of $g(\theta)$, and may choose $f(\theta)$ such that $f(\theta)\approx g(\theta)$. Thus to characterize the average coverage probability of interval estimators from BSA, the important issue is to determine how badly the relation $\Omega_{g,f}=1-\alpha$ breaks down as $g(\theta)$ departs from $f(\theta)$.

4.2. Simulation study

We use simulated data sets to study $\Omega_{g,f}$ as a function of $f(\theta)$ and $g(\theta)$. To clarify, $g(\theta)$ is the density used to sample parameters, while $f(\theta)$ is the prior distribution used to analyse the simulated data sets. We assume that $f(\theta)$ and $g(\theta)$ are members of the family of densities described in Section 2.2. This implies that we can write $f(\theta)=f(\beta_0, \beta_1, \lambda, \eta, \gamma_0, \gamma_1, \xi)=f(\lambda)f(\gamma_0, \gamma_1)f(\beta_0, \beta_1, \eta, \xi)$, and $g(\theta)=g(\beta_0, \beta_1, \lambda, \eta, \gamma_0, \gamma_1, \xi)=g(\lambda)g(\gamma_0, \gamma_1)g(\beta_0, \beta_1, \eta, \xi)$. We study the case where Z is

DOI: 10.1002/sim

comprised of two continuous covariates and

$$f(\beta_0, \beta_1, \eta, \xi) = g(\beta_0, \beta_1, \eta, \xi) \propto \exp\left[-\left\{ 2\left(\frac{\log(15)}{1.96} \right)^2 \right\}^{-1} (\beta_0^2 + \beta_1^2 + \eta'\eta + \xi'\xi) \right]$$

For both $f(\theta)$ and $g(\theta)$, we assign $c_1 = c_3 = c_4$, and we let c_1, c_2, c_3 and c_4 equal $\log(3)/1.96$ or $\log(6)/1.96$. We also study the limiting case where c_1, c_2, c_3 and c_4 equal zero.

For fixed $f(\theta)$ and $g(\theta)$, we estimate $\Omega_{g,f}$ using the following algorithm of Rubin and Schenker [18]:

1. For $t = 1, 2, \ldots, 50$:
   (a) Sample $\theta^{(t)}$ from $g(\theta)$.
   (b) Given $\theta^{(t)}$, sample $(y, x, z)^{(t)}$, with $n = 1000$, according to the decomposition

$$f(y, x, z|\theta) = f(y|x, u, z, \beta_0, \beta_1, \lambda, \eta) f(u|x, z, \gamma_0, \gamma_1, \xi) f(x) f(z) \tag{6}$$

   based on the models in equations (1) and (2), where $f(x) \sim$ Bernoulli(0.5) and $f(z) \sim$ independent uniform$(-2, 2)$.
   (c) Construct $I(y, x, z)^{(t)}$, an 80 per cent two-tailed credible interval for β_1 with respect to $f(\theta)$, using a single Markov chain of length 10 000 with 500 initial iterations discarded.
   (d) Calculate

$$Q(y, x, z)^{(t)} = \int 1\{\beta_1 \in I(y, x, z)^{(t)}\} \frac{f(y, x, z|\theta) g(\theta)}{\int f(y, x, z|\theta') g(\theta')\, \mathrm{d}\theta'}\, \mathrm{d}\theta$$

   the posterior probability of β_1 falling $I(y, x, z)^{(t)}$, with respect to the prior $g(\theta)$, using a single Markov chain of length 10 000 with 500 initial iterations discarded.

2. Calculate $\hat{\Omega}_{g,f} = (1/50) \sum_{t=1}^{50} Q(y, x, z)^{(t)}$, which is an unbiased estimator of $\Omega_{g,f}$ [18].

The sample size $n = 250$ is also considered and the results are discussed qualitatively.

Table V presents $\hat{\Omega}_{g,f}$ as a function of $f(\theta)$ and $g(\theta)$. Also included is the average length of the simulated interval estimates. As expected, the average coverage probabilities along diagonal elements of the table are within simulation error of 80 per cent, reflecting proper coverage of credible intervals because of equality between $f(\theta)$ and $g(\theta)$. In these instances, the distribution used to sample model parameters for the simulation is also used as a prior distribution for BSA. The consequence is that credible intervals have nominal coverage probability, on average, despite the fact that the model is non-identifiable. The off-diagonal elements of the table describe $\Omega_{g,f}$ when $f(\theta) \neq g(\theta)$, meaning that the investigator has incorrectly specified the true sampling distribution of bias parameters $\lambda, \gamma_0, \gamma_1$ and ξ in the sequence of observational studies. Table V demonstrates that $\Omega_{g,f} \approx 80$ per cent when $f(\theta) \approx g(\theta)$ with $\Omega_{g,f}$ being either greater than or less than 80 per cent, depending on the relationship between $f(\theta)$ and $g(\theta)$. For example, consider the estimate $\hat{\Omega}_{g,f} = 87.4$ per cent in the top right corner of the table. Here we have $c_1 = c_2 = c_3 = c_4 = \log(6)/1.96$ for the density $f(\theta)$ and $c_1 = c_2 = c_3 = c_4 = 0$ for the density $g(\theta)$. This means that we have a sequence of data sets without bias from unmeasured confounding. But the analyst falsely believes that there is a great deal of possible unmeasured confounding in each data set, and she chooses a prior distribution for BSA with large values of c_1, c_2, c_3 and c_4. The result is that credible intervals cover β_1 too often, 87.4 per cent of the time.

DOI: 10.1002/sim

Table V. Estimated average coverage probabilities $\hat{\Omega}_{g,f}$ and the average lengths of 80% two-tailed credible interval as a function of $f(\theta)$ and $g(\theta)$, when $n = 1000$.

			$f(\theta)$				
				$c_1 = c_3 = c_4 = \frac{\log(3)}{1.96}$		$c_1 = c_3 = c_4 = \frac{\log(6)}{1.96}$	
			$c_1 = c_3 = c_4 = 0$	$c_2 = \frac{\log(3)}{1.96}$	$c_2 = \frac{\log(6)}{1.96}$	$c_2 = \frac{\log(3)}{1.96}$	$c_2 = \frac{\log(6)}{1.96}$
$g(\theta)$	$c_1 = c_2 = c_3 = c_4 = 0$	Coverage (%)	80.1	82.7	82.4	87.7	87.4
		Length	0.47	0.51	0.51	0.61	0.61
	$c_1 = c_3 = c_4 = \frac{\log(3)}{1.96}$, $c_2 = \frac{\log(3)}{1.96}$	Coverage (%)	76.6	80.2	79.7	85.6	85.6
		Length	0.45	0.53	0.53	0.64	0.61
	$c_2 = \frac{\log(6)}{1.96}$	Coverage (%)	77.5	80.0	80.1	86.2	85.5
		Length	0.48	0.53	0.53	0.62	0.60
	$c_1 = c_3 = c_4 = \frac{\log(6)}{1.96}$, $c_2 = \frac{\log(3)}{1.96}$	Coverage (%)	69.3	72.6	72.4	80.4	79.3
		Length	0.48	0.52	0.50	0.61	0.60
	$c_2 = \frac{\log(6)}{1.96}$	Coverage (%)	70.5	73.0	73.5	80.6	79.4
		Length	0.49	0.50	0.50	0.60	0.59

Simulation standard errors for $\hat{\Omega}_{g,f}$ range from 0.1 to 0.6%. Simulation standard errors for average interval lengths range from 0.01 to 0.03.

There are similar findings when the simulation is repeated with sample size $n = 250$ (results not shown and are available from the authors upon request). Credible intervals are longer on average compared to when $n = 1000$. Estimates of $\Omega_{g,f}$ are within simulation error of 80 per cent when $f(\theta) = g(\theta)$. When $f(\theta) \neq g(\theta)$, $\hat{\Omega}_{g,f}$ are closer to 80 per cent relative to the case when $n = 1000$. This indicates that misspecifying $f(\theta)$ more adversely affects the performance of interval estimators when the sample size is large.

The results of the simulation study suggest that BSA interval estimators will have good frequentist performance provided that the investigator uses a prior distribution $f(\theta)$ which approximates $g(\theta)$, the true sampling distribution of model parameters in a hypothetical sequence of studies. To illustrate the value of this result for the practitioner in light of their own data set and analysis, we point out that the usual approach to analysing observational data is to assume that unmeasured confounding is absent. This corresponds to choosing $f(\theta)$ where $c_1 = c_2 = c_3 = c_4 = 0$. Table V shows that this method will perform poorly when there is large possible unmeasured confounding. Along the bottom row, we have $g(\theta)$ with $c_1 = c_2 = c_3 = c_4 = \log(6)/1.96$, which corresponds to sequences of observational studies with large unmeasured confounding on average. When the practitioner analyses the data sets and ignores unmeasured confounding, this yields $\hat{\Omega}_{g,f}$ equal to 70.5 per cent rather than the desired value of 80 per cent. In contrast, choosing $f(\theta)$ with

DOI: 10.1002/sim

non-zero values for c_1, c_2, c_3 and c_4 improves the average coverage probability. This indicates that the performance of interval estimation of exposure effects can be improved by acknowledging uncertainty about unmeasured confounding, even if it is only approximate, rather than assuming that bias is absent.

5. DISCUSSION

We propose BSA for observational studies where the associations between a binary exposure, binary outcome, measured covariates and a single binary unmeasured confounder can be modelled using logistic regression. The proposed family of prior distributions model beliefs about unknown and unmeasured confounding, and are applicable in settings where there is a possible unknown risk factor for the outcome which also predicts exposure. Other families of prior distributions can be used through suitable modification of the acceptance ratios in the Metropolis updating steps used in posterior simulation. For example, non-zero prior means for the bias parameters could be specified.

Because the model is not identifiable the usual asymptotic properties for posterior distributions do not apply, complicating an evaluation of the performance of interval estimators. The posterior distribution is sensitive to the choice of prior distribution used for data analysis, and this sensitivity increases with increasing sample size [9]. This seems unappealing because inferences about model parameters are driven by subjective prior information. However, the method accurately characterizes uncertainty in observational studies because uncertainty about bias does not diminish with increasing sample size [2, 5].

To illustrate the impact of the choice of prior distribution used for sensitivity analysis on the performance of interval estimators for the treatment effect, we used simulations to study coverage probability when averaged over a hypothetical sequence of observational studies. The results indicate that credible intervals will have approximately nominal coverage probability, on average, provided that prior distribution used for sensitivity analysis approximates the sampling distribution of model parameters in the sequence of studies. Thus interval estimators may have good frequentist properties even though the model used for BSA is not identifiable. While average coverage probability is an unusual measure of frequentist performance, its relevance has been advocated elsewhere [19] because it characterizes performance of interval estimators in the setting in which they are actually used, namely to analyse data from different investigations, rather than data under hypothetical repeated sampling given fixed parameters. Using average coverage probability to measure performance is natural in meta-analysis where the sampling distribution of model parameters is usually modelled in the analysis.

The usual analysis of observational data assumes that unmeasured confounding is absent. This approach is appealing because it involves identifiable models and yields interval estimators with standard asymptotic properties. But the simulations in Section 4 indicate that in a sequence of observational studies with large unmeasured confounding, on average, credible intervals which ignore unmeasured confounding will not have nominal average coverage probability. They will also not have nominal large sample coverage probability conditional on fixed non-zero values of the bias parameter. Therefore, it appears that performance may be improved by acknowledging uncertainty about unmeasured confounding, even approximately, rather than assuming it is absent.

The proposed method has limitations. While the MCMC algorithm can be run using the statistical package R [20], we observed that sampler convergence deteriorates with increasing sample size.

DOI: 10.1002/sim

Using a diffuse prior distribution for bias parameters (e.g. values of c_1, c_2, c_3 and c_4 greater than $\log(6)/1.96$) aggravated this problem. Other investigations of Bayesian methods using non-identifiable models have reported similar findings [9, 21]. Thus more involved MCMC algorithms must be developed in some situations. Another limitation of sensitivity analysis is the difficulty of assessing the appropriateness of the model for unmeasured confounding. Because U is unobserved, the data cannot be used to verify modelling assumptions about the association between U and measured variables. It would be informative to study the performance of BSA under model misspecification. One relevant setting is the case when U is used to approximate a vector of unmeasured confounders. This approach to sensitivity analysis is discussed by Greenland [5].

APPENDIX A

We outline a method of simulating from the posterior distribution $f(\beta_0, \beta_1, \lambda, \eta, \gamma_0, \gamma_1, \xi | y, x, z)$ using Gibbs sampling with Metropolis updating. The imputation step (I-step) samples the latent u from $f(u|\beta_0, \beta_1, \lambda, \eta, \gamma_0, \gamma_1, \xi, y, x, z)$. To accomplish this, write the joint distribution of y, u and model parameters given x and z as

$$\begin{aligned} f(\beta_0, \beta_1, \lambda, \eta, \gamma_0, \gamma_1, \xi, u, y|x, z) &= f(y|\beta_0, \beta_1, \lambda, \eta, u, x, z) f(u|\gamma_0, \gamma_1, \xi, x, z) \\ &\quad \times f(\beta_0, \beta_1, \lambda, \eta, \gamma_0, \gamma_1, \xi) \\ &\propto \prod_{i=1}^{n} \left[\frac{\exp\{y_i(\beta_0 + \beta_1 x_i + \lambda u_i + \eta' z_i)\}}{1 + \exp\{\beta_0 + \beta_1 x_i + \lambda u_i + \eta' z_i\}} \right. \\ &\quad \left. \times \frac{\exp\{u_i(\gamma_0 + \gamma_1 x_i + \xi' z_i)\}}{1 + \exp\{\gamma_0 + \gamma_1 x_i + \xi' z_i\}} \right] \\ &\quad \times f(\beta_0, \beta_1, \lambda, \eta, \gamma_0, \gamma_1, \xi) \end{aligned}$$

based on the likelihood functions for logistic regression. By conditioning on $\beta_0, \beta_1, \lambda, \eta, \gamma_0, \gamma_1, \xi, y$, we may write

$$f(u|\beta_0, \beta_1, \lambda, \eta, \gamma_0, \gamma_1, \xi, y, x, z) \propto \prod_{i=1}^{n} \frac{\exp\{u_i(y_i\lambda + \gamma_0 + \gamma_1 x_i + \xi' z_i)\}}{1 + \exp\{\beta_0 + \beta_1 x_i + \lambda u_i + \eta' z_i + \xi' z_i\}}$$

and therefore

$$f(u_i|\beta_0, \beta_1, \lambda, \eta, \gamma_0, \gamma_1, \xi, y, x, z) \propto \frac{\exp\{u_i(y_i\lambda + \gamma_0 + \gamma_1 x_i + \xi' z_i)\}}{1 + \exp\{\beta_0 + \beta_1 x_i + \lambda u_i + \eta' z_i\}}$$

where u_i denotes U for the ith subject. Since u_i is dichotomous, this expression can be normalized and written in closed form. Sampling from $f(u|\beta_0, \beta_1, \lambda, \eta, \gamma_0, \gamma_1, \xi, y, x, z)$, amounts to sampling an $n \times 1$ vector of Bernoulli random variables with success probabilities $f(u_i|\beta_0, \beta_1, \lambda, \eta, \gamma_0, \gamma_1, \xi, y, x, z)$, for $i = 1, \ldots, n$.

The posterior step (P-Step) samples from the posterior distributions in the logistic regression models of equations (1) and (2) using the sampled u. The first model regresses y on x, u and z,

DOI: 10.1002/sim

while the second model regresses u on x and z. Bayesian logistic regression can be accomplished using the independence sampler Metropolis Hastings algorithm with a proposal distribution given by the asymptotic approximation to the distribution of the maximum likelihood estimator. Further details of Bayesian logistic regression are given in Reference [11].

To summarize, simulation from the posterior distribution $f(\beta_0, \beta_1, \lambda, \eta, \gamma_0, \gamma_1, \xi | y, x, z)$ proceeds as follows:

1. Obtain reasonably overdispersed starting values for $(\beta_0^{(0)}, \beta_1^{(0)}, \lambda^{(0)}, \eta^{(0)}, \gamma_0^{(0)}, \gamma_1^{(0)}, \xi^{(0)})$ and $u^{(0)}$ by simulating $u^{(0)}$ as an $n \times 1$ vector of Bernoulli random variables with success probabilities equal to $\frac{1}{2}$ and then sampling $(\beta_0^{(0)}, \beta_1^{(0)}, \lambda^{(0)}, \eta^{(0)}, \gamma_0^{(0)}, \gamma_1^{(0)}, \xi^{(0)})$ from the asymptotic distribution of the maximum likelihood estimator obtained via regressions of y on x, $u^{(0)}$, z, and $u^{(0)}$ on x and z.
2. For $t = 1, 2, \ldots$

   I-Step: Sample $u^{(t)}$ from $f(u | \beta_0^{(t-1)}, \beta_1^{(t-1)}, \lambda^{(t-1)}, \eta^{(t-1)}, \gamma_0^{(t-1)}, \gamma_1^{(t-1)}, \xi^{(t-1)}, y, x, z)$.

   P-step: Sample $\beta_0^{(t)}, \beta_1^{(t)}, \lambda^{(t)}, \eta^{(t)}$ using a Metropolis Hastings step with target density $f(\beta_0, \beta_1, \lambda, \eta | u^{(t)}, y, x, z)$ and proposal distribution obtained by logistic regression of y on x, z and $u^{(t)}$.

   Sample $(\gamma_0^{(t)}, \gamma_1^{(t)}, \xi^{(t)})$ using a Metropolis Hastings step with target density $f(\gamma_0, \gamma_1, \xi | u^{(t)}, y, x, z)$ and proposal density obtained by logistic regression of $u^{(t)}$ on x, z.
3. Discard $u^{(0)}, u^{(1)}, \ldots$

After discarding a suitable number of initial iterations, the sequence $(\beta_0^{(t)}, \beta_1^{(t)}, \lambda^{(t)}, \eta^{(t)}, \gamma_0^{(t)}, \gamma_1^{(t)}, \xi^{(t)})$ for $t = 1, 2, \ldots$, comprises a sample from the required posterior distribution. Convergence of the Markov chain can be assessed using standard tools [11].

ACKNOWLEDGEMENTS

Funding for this research was provided through a Michael Smith Foundation Junior Graduate Studentship Award.

REFERENCES

1. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine* 1997; **15**:757–763.
2. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *Journal of Clinical Epidemiology* 2005; **58**:323–337.
3. Levy AR, Tamblyn RM, Abrahamowicz M, McLeod PJ, Fitchett D. Use of time-dependent measures to estimate benefits of beta-blockers after myocardial infarction. *Pharmacoepidemiology and Drug Safety* 2004; **13**:623–631.
4. Greenland S. The impact of prior distributions for uncontrolled confounding and response bias: a case study of the relation of wire codes and magnetic fields to childhood leukemia. *Journal of the American Statistical Association* 2003; **98**:47–54.
5. Greenland S. Multiple bias modelling for analysis of observational data (with Discussion). *Journal of the Royal Statistical Society, Series A* 2005; **168**:267–306.
6. Steenland K, Greenland S. Monte-Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. *American Journal of Epidemiology* 2004; **160**:384–392.
7. Robins JM, Rotnitzky A, Scharfstein DO. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In *Statistical Models in Epidemiology*, Halloran ME, Berry D (eds). Springer: New York, 1999; 1–92.

DOI: 10.1002/sim

8. Chib S. On inferring effects of binary treatments with unobserved confounders. In *Bayesian Statistics*, Bernardo J, DeGroot M, Lindley D, Smith A (eds), vol. 7. Elsevier: New York, 2003; 65–84.
9. Gustafson P. On model expansion, model contraction, identifiability, and prior information: two illustrative scenarios involving mismeasured variables. *Statistical Science* 2005; **20**:111–140.
10. Lin DY, Psaty BM, Kronmal RA. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics* 1998; **54**:948–963.
11. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis* (2nd edn). Chapman & Hall/CRC Press: New York/Boca Raton, FL, 2004.
12. Chamberlayne R, Green B, Barer ML, Hertzman C, Lawrence WJ, Sheps SB. Creating a population-based linked health database: a new resource for health services research. *Canadian Journal of Public Health* 1998; **89**:270–273.
13. 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**:1489–1499.
14. Austin PC, Mamdani MM. A comparison of propensity score methods: a case-study estimating the effectiveness of post-AMI statin use. *Statistics in Medicine* 2006; **25**:2084–2106.
15. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Statistical Science* 1992; **7**:457–511.
16. Rosenbaum PR, Rubin DB. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society* 1983; **45**:212–218.
17. Rubin DB. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics* 1984; **12**:1151–1172.
18. Rubin DB, Schenker N. Efficiently simulating the coverage properties of interval estimates. *Applied Statistics* 1986; **2**:159–167.
19. Berger JO, Bayarri MJ. The interplay of Bayesian and frequentist analysis. *Statistical Science* 2004; **9**:58–80.
20. 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
21. Gelfand AE, Sahu SK. Identifiability, improper priors, and Gibbs sampling for generalized linear models. *Journal of the American Statistical Association* 1999; **94**:247–253.

DOI: 10.1002/sim