Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–17

DOI: 10.1177/0962280217729844
journals.sagepub.com/home/smm
SAGE

Bayesian sensitivity analysis for unmeasured confounding in causal mediation analysis

Lawrence C McCandless and Julian M Somers

Abstract

Causal mediation analysis techniques enable investigators to examine whether the effect of the exposure on an outcome is mediated by some intermediate variable. Motivated by a data example from epidemiology, we consider estimation of natural direct and indirect effects on a survival outcome. An important concern is bias from confounders that may be unmeasured. Estimating natural direct and indirect effects requires an elaborate series of assumptions in order to identify the target quantities. The analyst must carefully measure and adjust for important predictors of the exposure, mediator and outcome. Omitting important confounders may bias the results in a way that is difficult to predict. In recent years, several methods have been proposed to explore sensitivity to unmeasured confounding in mediation analysis. However, many of these methods limit complexity by relying on a handful of sensitivity parameters that are difficult to interpret, or alternatively, by assuming that specific patterns of unmeasured confounding are absent. Instead, we propose a simple Bayesian sensitivity analysis technique that is indexed by four bias parameters. Our method has the unique advantage that it is able to simultaneously assess unmeasured confounding in the mediator–outcome, exposure–outcome and exposure–mediator relationships. It is a natural Bayesian extension of the sensitivity analysis methodologies of VanderWeele, which have been widely used in the epidemiology literature. We present simulation findings, and additionally, we illustrate the method in an epidemiological study of mortality rates in criminal offenders from British Columbia.

Keywords

Bayesian analysis, sensitivity analysis, causal inference, Markov chain Monte Carlo, unmeasured confounding

1 Introduction

In epidemiology research, investigators often seek to examine the extent to which the effect of the exposure on an outcome is mediated by some intermediate variable.[1–3] Causal mediation analysis techniques enable estimation of the total effect of the exposure on the outcome, the effect of the exposure that acts through a given set of intermediate variables (indirect effect) and the effect of the exposure that remains unexplained by those same intermediate variables (direct effect).

Motivated by an epidemiologic study of mortality rates in criminal offenders, we consider estimation of natural direct and indirect effects on a survival outcome in proportional hazards regression.[4,5] To illustrate the setting, we adopt the potential outcomes framework.[2] Let X denote an exposure variable, and M_X denote the value of the mediator M that would have been observed if the exposure had been set to X. Additionally, let $h_{X,M}(t|\boldsymbol{C})$ denote the value of the hazard function that would have been observed if the exposure and mediator had been set to X and M, respectively. Following VanderWeele,[5] the total effect hazard ratio, conditional on $\boldsymbol{C}$, denoted $\frac{h_1(t|\boldsymbol{C})}{h_0(t|\boldsymbol{C})} = \frac{h_{1,M_1}(t|\boldsymbol{C})}{h_{0,M_0}(t|\boldsymbol{C})}$ where $h_X(t|\boldsymbol{C}) = h_{X,M_X}(t|\boldsymbol{C})$, can be decomposed as the product of the natural direct

Faculty of Health Sciences, Simon Fraser University, Burnaby, Canada

Corresponding author:
Lawrence C McCandless, Faculty of Health Sciences, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada.
Email: lmccandl@sfu.ca

and natural indirect effect hazard ratios

$$\frac{h_1(t|\boldsymbol{C})}{h_0(t|\boldsymbol{C})} = \frac{h_{1,M_1}(t|\boldsymbol{C})}{h_{0,M_0}(t|\boldsymbol{C})} = \frac{h_{1,M_1}(t|\boldsymbol{C})}{h_{1,M_0}(t|\boldsymbol{C})} \times \frac{h_{1,M_0}(t|\boldsymbol{C})}{h_{0,M_0}(t|\boldsymbol{C})} \tag{1}$$

There is also a corresponding decomposition of the total effect that multiplies and divides equation (1) by $h_{0,M_1}(t|\boldsymbol{C})$.[5]

An important concern in mediation analysis is understanding bias from confounding variables that may be unmeasured. The estimation of natural direct and indirect effects requires that the analyst make an elaborate series of assumptions to identify the target quantities. For survival outcomes, we require a total of four unmeasured confounding assumptions.[5] Conditional on covariates, we require (1) no confounding for the exposure–outcome relationship; (2) no confounding for the mediator–outcome relationship; (3) no confounding for the exposure–mediator relationship; and finally, (4) no mediator–outcome confounder that is itself affected by the exposure.[6] Consequently, the analyst must carefully record information about important predictors of the exposure, mediator and outcome so that they can be controlled for in the analysis. Omitting important confounders will bias the results of the mediation analysis in a way that is difficult to predict.

Moreover, the problem of unmeasured confounding is a particularly serious in epidemiologic studies using large databases of electronic health records.[7,8] These data benefit from having large representative samples from entire populations. However, because they are not designed for research purposes they are often missing information on important clinical variables that are predictive of the exposure, outcome and mediating variables. In large sample sizes, the conventional measures of random error uncertainty (e.g. p-values and 95% confidence intervals) diminish to zero size. However, uncertainty from unmeasured confounding does not diminish with increasing sample size. p-Values and 95% confidence intervals can seriously under report the uncertainty about target parameters.[8]

Sensitivity analysis is an important tool for examining the plausibility of assumptions. Several methods have been proposed to explore sensitivity to unmeasured confounding in causal mediation analysis, including the works of VanderWeele,[9,10] Kim et al.,[11,12] Imai et al.[13] and others (e.g. Albert and Wang[14] and Hafeman[15]). VanderWeele[9] describes a regression-based approach that has been widely used in epidemiology, and that extends the Baron and Kenny[16] method from the field of psychology. Imai et al.[13] describe a method using a single parameter for the correlation in the error terms of a linear structural equation model. Kim et al.[11] use a Bayesian non-parametric approach that involves direct modelling of the unobservable potential outcomes and mediators. However, no existing methods are able to assess the effect of unmeasured confounding on the individual relationships between the exposure, mediator and outcome (i.e. assumptions 1–4 above). For example, VanderWeele[9,10] describes an approach to assess unmeasured mediator–outcome confounding while simultaneously requiring that exposure–outcome and exposure–mediator confounding is absent. But in practice the investigator often has a specific omitted variable in mind, and it is necessary to examine the different assumptions simultaneously. Such sensitivity analyses are inherently complex. The results are often presented using high dimensional tables where sensitivity parameters are presented over a grid of values (e.g. Chiolero et al.[17]). A further issue is that these methods give no formal quantification of uncertainty about the sensitivity parameter inputs.[18]

An alternative strategy is to approach the problem from a Bayesian perspective. McCandless et al.[19] proposed a Bayesian sensitivity analysis (BSA) for unmeasured confounding in observational studies. The idea is to incorporate uncertainty about unmeasured confounding by using prior distributions for sensitivity parameters.[20] The posterior distribution for the causal effect estimate incorporates uncertainty from bias (systematic error) in addition to uncertainty from random sampling (random error). An advantage of BSA is that it gives dimension reduction when there are multiple sensitivity parameters inputs because it averages over uncertainty in the prior. Furthermore, posterior credible intervals will often have better frequentist coverage probability compared to interval estimates that ignore unmeasured confounding.[18]

In this manuscript, we build on the work of McCandless et al.[19] and propose a novel Bayesian method to explore sensitivity to unmeasured confounding in causal mediation analysis. The approach is a natural Bayesian extension of the methodology of VanderWeele.[9,10] It has the unique advantage that it is able to simultaneously assess unmeasured mediator–outcome, exposure–outcome and exposure–mediator confounding. We illustrate the method in an epidemiologic study of mortality in 66,358 criminal offenders from British Columbia, Canada.[21]

2 Motivating example of causal mediation analysis

2.1 Description of the data

McCandless et al.[21] examined patterns of mortality in an epidemiological study of criminal offenders. The analysis was part of a broader initiative between the provincial Government of British Columbia and Simon Fraser University to examine health services use, public safety and incarceration among people with mental illness.[22]

In the present investigation, our objective is to determine whether the higher rates of mortality among offenders with mental illness are mediated by higher rates of criminal justice encounters (e.g. prison sentences). In Canada, and internationally, there is a well known 'revolving door syndrome' where low-level criminals with mental illness and drug addictions become trapped in a cycle of incarceration and police encounters that lead to poor health outcomes and death.[22]

Table 1 gives a description of the data set. The data were drawn from 66,358 adult criminal offenders who were followed longitudinally from the date of their first criminal sentence, after 1 April 2001, until death or censoring on 31 March 2010 (mean follow-up time 4.5 years). To set the stage for our proposed sensitivity analyses, we replicate the causal mediation analysis of McCandless et al.[21] using Bayesian methods. For each individual in the data set, the outcome variable T is the length of time, measured in days, from the start of follow-up until either death or censoring. We let δ be a zero-one indicator variable for death ($\delta = 1$) or censoring ($\delta = 0$). The exposure variable X is having a Dual-Diagnosis (DDx) of both a mental disorder and substance use disorder based on electronic medical records for physician services prior to the start of follow-up. The mediating variable M is taken as continuous, and it is the individual-level rate of criminal sentencing, log transformed. For each individual in the data set, the quantity M is a derived variable that was taken from the larger anonymized data set of McCandless et al.[21] For example, if $M = 0.75$ for a particular individual, then the sentencing rate for that individual, during the course of follow-up, was $\exp(0.75) = 2.12$ sentences per year. Finally, we let $\boldsymbol{C}$ denote a vector of five binary participant variables for covariate adjustment, which are listed in Table 1.

In Table 1, a total of 1663 (2.5%) participants died over the course of follow-up, and the rate of mortality was nearly three time higher in the exposed group than in the unexposed group. The mean of the continuous mediating variable M (sentencing rate, log transformed) was -0.57, which corresponds to $\exp(-0.57) = 0.56$ sentences per year, and exposed participants were sentenced at higher rates than unexposed participants ($\exp(-0.30) = 0.75$

Table 1. Characteristics of 66,358 criminal offenders who were sentenced through British Columbia provincial courts between 1 January 2001 and 31 March 2010.

	Number (%) or Mean (SD)		
Variable	Exposed to DDx ($n = 6,597$)	Unexposed to DDx ($n = 59,761$)	Total ($n = 66,358$)
Outcome variable			
Died	412 (6.2%)	1,251 (2.1%)	1,663 (2.5%)
Mediating variable			
Sentencing rate (log sentences/year)	−0.30 (1.11)	−0.61 (1.08)	−0.57 (1.09)
Covariates			
Male gender	4497 (68%)	50,099 (84%)	54,596 (82%)
Age			
<25	1552 (24%)	19,731 (33%)	21,283 (32%)
25 − 40	2781 (42%)	22,417 (38%)	25,198 (38%)
>40	2264 (34%)	17,613 (29%)	19,877 (30%)
Education			
Grade 9 or less	939 (14%)	7,477 (13%)	8,416 (13%)
Higher than grade 9	5658 (86%)	52,284 (87%)	57,942 (87%)
Ethnicity			
Caucasian	5110 (77%)	40,645 (68%)	45,755 (69%)
Other	1487 (23%)	19,116 (32%)	20,603 (31%)
Years of follow-up	4.5 (2.6)	4.5 (2.6)	4.5 (2.6)

DDx: dual-medical diagnosis of both a mental disorder and substance use disorder; SD: standard deviation.

sentences per year versus $\exp(-0.61) = 0.54$ sentences per year). Exposed participants were more likely to be older, Caucasian and female.

2.2 Bayesian regression analysis of mortality

We begin by presenting a Bayesian survival analysis of mortality rates ignoring the problem of unmeasured confounding. Following McCandless et al.,[21] we assume a Weibull proportional hazards model for mortality that follows the equation

$$T|X, M, \boldsymbol{C}, \delta \sim \text{Weibull}\big(\exp(\beta_0 + \beta_X X + \beta_M M + \beta_c^{\mathrm{T}} \boldsymbol{C}), \nu\big) \tag{2}$$

where $\exp(\beta_0 + \beta_X X + \beta_M M + \beta_c^{\mathrm{T}} \boldsymbol{C})$ is the shape parameter and ν is the scale parameter. We omit the location parameter from the three-parameter Weibull distribution because the survival time T, measured in days, is greater than zero. Note that the right hand size of equation (2) does not depend on the censoring indicator variable δ because we make the usual assumption that the censoring is non-informative conditional on $(X, M, \boldsymbol{C})$. Additionally, equation (2) omits an interaction term between the exposure and the mediator. The analysis of McCandless et al.[21] found only negligible evidence of an interaction. See VanderWeele[3,5] for detailed discussion of the importance of exposure–mediator interactions in causal mediation analysis with survival outcomes.

The Weibull regression model is straightforward to implement in a Bayesian analysis using STAN. [23] We assign default non-informative priors to the regression coefficients and scale parameter, given by $P(\beta_0), P(\beta_X), P(\beta_M), P(\beta_{\boldsymbol{C}}), P(\nu) \propto 1$ with $\nu > 0$,[24] and we sample from the posterior distribution using gradient-based Markov chain Monte Carlo. To assess sampler convergence we used the scale reduction factor (SRF) using five independent chains with overdispersed starting values, each of length 1000 iterations after discarding initial warm-up.[25] The SRF is based on a comparison of the within-chain and between-chain variances, and convergence to the target distribution is achieved when the SRF is close to 1.0.

The results are given in Table 2 under the column heading NAIVE. We call the analysis 'NAIVE' because it naively ignores unmeasured confounding. Table 2 demonstrates that offenders who are exposed to DDx (i.e. who have a Dual Diagnosis of both a mental disorder and substance use disorder) have a two-fold higher rate of

Table 2. Bayesian Weibull proportional hazards model for mortality.

		Log hazard ratio (95% CI)	
Variable	Parameter	NAIVE	BSA
Exposure variable			
Exposed to DDx	β_X	0.68 (0.56, 0.80)	0.70 (0.48, 0.91)
Mediating variable			
Sentencing rate (log sentences/year)	β_M	0.91 (0.86, 0.96)	0.92 (0.75, 1.09)
Covariates			
Male gender	β_{C_1}	−0.03 (−0.15, 0.09)	−0.04 (−0.17, 0.09)
Age			
<25	β_{C_2}	−0.68 (−0.82, −0.54)	−0.68 (−0.84, −0.53)
25–40[a]	—	0.0	0.0
>40	β_{C_3}	1.17 (1.07, 1.27)	1.18 (1.07, 1.29)
Education			
Grade 9 or less	β_{C_4}	0.23 (0.10, 0.36)	0.23 (0.11, 0.36)
Higher than Grade 9[a]	—	—	—
Ethnicity			
Caucasian[a]	—	—	—
Other	β_{C_5}	−0.13 (−0.25, −0.01)	−0.13 (−0.24, −0.01)

Note: Posterior means (95% posterior credible intervals) for the log hazard ratio parameters β_X, β_M and $\beta_{\boldsymbol{C}} = (\beta_{C_1}, \ldots, \beta_{C_5})$. DDx: dual diagnosis of both a mental disorder and substance use disorder; BSA: Bayesian sensitivity analysis.
[a]Reference group.

mortality with hazard ratio $\exp(0.68) = 1.97$. Additionally, higher rates of criminal sentencing are associated with greater mortality. In particular, a two-fold increase in the rate of criminal sentencing corresponds to a nearly two-fold higher rate of mortality, with hazard ratio $\exp(0.91 \times \log(2)) = 1.88$.

2.3 Bayesian regression analysis of rate of criminal sentencing

Next, we examine the mediating variable of criminal sentencing. Building on McCandless et al.[21] and the causal mediation analysis approach of VanderWeele,[5] we propose a Bayesian linear model for the mediator M given by

$$M|X, \boldsymbol{C} \sim \text{Normal}(\alpha_0 + \alpha_X X + \alpha_c^{\mathrm{T}} \boldsymbol{C}, \sigma^2) \tag{3}$$

Again, we assign non-informative priors to the model parameters $P(\alpha_0), P(\alpha_X), P(\alpha_c), P(\sigma) \propto 1$ with $\sigma > 0$. We fit the model using STAN, with five independent chains of length 1000 after discarding warm-up, and the SRF was 1.0 indicating sampler convergence.

Table 3 gives the results of fitting the NAIVE linear model for the mediator ignoring unmeasured confounding. Because M is the log transformed individual-level rate of criminal sentencing, this means that the regression coefficients in Table 3 can be interpreted as log rate ratios. The results show that exposed participants experience higher rates of criminal sentencing. Specifically, the DDx exposure is associated with an $\exp(0.35) = 1.42$ higher rate of sentencing. This confirms the well-known finding that offenders with mental illness and addictions are more likely to be arrested, incarcerated and given criminal sentences. Table 3 also identifies other important determinants of criminal sentencing, which include being young, male and lower formal education.

2.4 Bayesian estimation of natural direct and indirect effects ignoring unmeasured confounding

We consider estimation of natural direct and indirect effects on a survival outcome in proportional hazards regression.[5] In the offender data, the natural direct effect, defined as $\frac{h_{1,M_0}(t|\boldsymbol{C})}{h_{0,M_0}(t|\boldsymbol{C})}$ in equation (1), is the effect of the exposure on mortality assuming that the individual-level criminal sentencing rate is set to the value it would have been for each subject if they had not been exposed. This effect could be the result of an intervention, such as Drug Treatment Court,[22] which limits the effect of the exposure on the frequency of criminal sentencing. In contrast,

Table 3. Bayesian linear regression model for the rate of criminal sentencing.

		Log rate ratio (95% CI)	
Variable	Parameter	NAIVE	BSA
Exposure variable			
Exposed to DDx	α_X	0.35 (0.32, 0.38)	0.35 (0.20, 0.50)
Covariates			
Male gender	α_{C_1}	0.08 (0.06, 0.10)	0.08 (0.06, 0.10)
Age			
<25	α_{C_2}	0.20 (0.18, 0.22)	0.20 (0.18, 0.22)
25–40[a]	—	—	—
>40	α_{C_3}	−0.13 (−0.15, −0.11)	−0.13 (−0.15, −0.11)
Education			
Grade 9 or less	α_{C_4}	0.11 (0.08, 0.13)	0.11 (0.08, 0.13)
Higher than Grade 9[a]	—	—	—
Ethnicity			
Caucasian[a]	—	—	—
Other	α_{C_5}	0.05 (0.04, 0.07)	0.05 (0.04, 0.07)

Posterior means (95% posterior credible intervals) for the log rate-ratios parameters α_X, and $\alpha_{\boldsymbol{C}} = (\alpha_{C_1}, \ldots, \alpha_{C_5})$. DDx, dual diagnosis of both a mental disorder and substance use disorder; BSA: Bayesian sensitivity analysis.
[a]Reference group.

the natural indirect effect isolates the effect of sentencing on mortality. It is defined as $\frac{h_{1,M_1}(t|\boldsymbol{C})}{h_{1,M_0}(t|\boldsymbol{C})}$, and it examines the change in mortality if we were to intervene and change the mediator to the value it would have been if each individual were exposed. Note that the controlled direct effect is less meaningful in the offender data because it is not possible to set the rate of criminal sentencing rate to a specific value (e.g. such as $M = 0$). We refer the reader to VanderWeele[3] for further discussion of causal mediation analysis.

For Bayesian estimation of natural direct and indirect effects, we extend the methodology of VanderWeele[5] for time-to-event outcomes. VanderWeele[5] proves that if the parametric models in equation (2) and (3) are correctly specified, and additionally, the cumulative incidence of the outcome is low, then the natural direct and indirect effects on the log hazard ratio scale can be written as

$$\log\left[\frac{h_{1,M_0}(t|\boldsymbol{C})}{h_{0,M_0}(t|\boldsymbol{C})}\right] \approx \beta_X \tag{4}$$

$$\log\left[\frac{h_{1,M_1}(t|\boldsymbol{C})}{h_{1,M_0}(t|\boldsymbol{C})}\right] \approx \alpha_X\beta_M \tag{5}$$

The total effect is given by $\log\left[\frac{h_1(t|\boldsymbol{C})}{h_0(t|\boldsymbol{C})}\right] \approx \beta_X + \alpha_X\beta_M$, and the proportion mediated is given by $\frac{\alpha_X\beta_M}{\beta_X+\alpha_X\beta_M}$. VanderWeele[5] also provides formulas for natural direct and indirect effects when there is an exposure–mediator interaction.

We present Bayesian estimation of the natural direct and indirect effects. Our approach is to fit the models for the mediator and the outcome and then compute the posterior distribution of the quantities in equations (4) and (5). This corresponds to a Bayesian version of the product of coefficients method, as described by Yuan and MacKinnon[26] and Kim et al.[11]

The results of this NAIVE Bayesian mediation analysis ignoring unmeasured confounding are given in Table 4. For comparison, we include results from two other methods: First, the mediation analysis method for survival outcomes of VanderWeele,[5] which is identical to NAIVE except that the point estimates are obtained by maximum likelihood estimation of equations (2) and (3) with bootstrapped standard errors. Second, we apply the method of Lange et al.,[27] which estimates natural direct and indirect effects using a Weibull proportional hazards marginal structural model with stabilized inverse probability weights and bootstrapped standard errors.

In Table 4, there is a large indirect effect of the exposure on mortality that is mediated by criminal sentencing. NAIVE and the method of VanderWeele[5] give nearly identical results because both methods use the same models. However, the results of Lange et al.[27] differ because it uses a marginal structural that includes $\boldsymbol{C}$, but not M, in the model for survival time. In contrast, NAIVE and the method of VanderWeele adjust for M directly in the linear predictor in equation (3). It is well known that the hazard ratio is not collapsible, and that marginal and conditional hazard ratios may differ slightly.[28]

A crucial issue in Table 4 is understanding the role of unmeasured confounding. The regression models in equations (2) and (3) control for $\boldsymbol{C}$, which includes four factors: age, gender, education and ethnicity. However, as argued by McCandless et al.,[21] there are important variables that predict criminal behaviour and mortality that we were unable to measure and adjust for in the multivariable models, including social achievement, family and peer relationships and personality traits. These missing variables may confound the mediator–outcome, exposure–outcome and exposure–mediator relationships. In order to quantify impact of unmeasured confounding on results of the mediation analysis, we require a model for sensitivity analysis.

Table 4. Mediation analysis results in the criminal offender data.

	Log hazard ratio (95% CI)			
Method	Natural direct effect	Natural indirect effect	Total effect	Proportion mediated
NAIVE	0.69 (0.57, 0.80)	0.32 (0.29, 0.35)	1.01 (0.88, 1.12)	32% (27%, 36%)
VanderWeele[5]	0.69 (0.57, 0.81)	0.32 (0.29, 0.35)	1.01 (0.89, 1.13)	32% (27%, 36%)
Lange et al.[27]	0.79 (0.68, 0.90)	0.26 (0.23, 0.29)	1.05 (0.94, 1.16)	25% (21%, 28%)
BSA	0.70 (0.48, 0.92)	0.32 (0.18, 0.47)	1.02 (0.78, 1.25)	32% (18%, 45%)

Note: Log hazard ratios (95% interval estimates) for causal mediation parameters using BSA and three other methods that ignore unmeasured confounding: NAIVE, the method of VanderWeele,[5] and the method of Lange et al.[27] BSA: Bayesian sensitivity analysis.

3 BSA for unmeasured confounding in causal mediation analysis

3.1 A model for unmeasured confounding in causal mediation analysis

We extend the sensitivity analysis methodology of VanderWeele[9,10] using a Bayesian approach that builds on the works of McCandless et al.[19] and Lin et al.[29] Our strategy is to introduce a latent binary variable U that takes values 1 or 0 to indicate the presence or absence of an unmeasured confounder. Figure 1 gives a causal diagram for the hypothesized relationship between variables. In Figure 1, there is arrow between U and each of the measured variables T, M and X. Thus, a unique feature of our method is that U is simultaneously a confounder for the mediator–outcome, exposure–outcome and exposure–mediator relationships.

We model the data parametrically and factorize the joint distribution of $P(T, M, U|X, \boldsymbol{C}, \delta) = P(T|M, U, X, \boldsymbol{C}, \delta)$ $P(M|U, X, \boldsymbol{C})P(U|X, \boldsymbol{C})$ as

$$T|X, M, \boldsymbol{C}, U, \delta \sim \text{Weibull}(\exp(\beta_0 + \beta_X X + \beta_M M + \beta_c^{\mathrm{T}}\boldsymbol{C} + \beta_U U), \nu) \tag{6}$$

$$M|X, \boldsymbol{C}, U \sim \text{Normal}(\alpha_0 + \alpha_X X + \alpha_c^{\mathrm{T}}\boldsymbol{C} + \alpha_U U, \sigma^2) \tag{7}$$

$$U|X, \boldsymbol{C} \sim \text{Bernoulli}(\text{expit}(\gamma_0 + \gamma_X X)) \tag{8}$$

where equation (6) assumes that the censoring indicator δ is not informative about T, conditional on $(X, M, \boldsymbol{C})$, as in equation (2).

This model is indexed by four bias parameters $\beta_U, \alpha_U, \gamma_0$ and γ_X that can be used as the basis of a sensitivity analysis for unmeasured confounding. The parameter β_U is the log hazard ratio for the association between U and survival time T conditional on X, M and $\boldsymbol{C}$. The quantity α_U governs the association between U and the mediator M, given X and $\boldsymbol{C}$. Equation (8) models the prevalence of the unmeasured confounder within levels of X and $\boldsymbol{C}$. In particular, γ_X is the log odds ratio for the association between the unmeasured confounder and X given $\boldsymbol{C}$.

The models in equations (6) to (8) make a series of simplifying assumptions for ease of use. Equation (6) omits interaction terms of any kind. There is no exposure–mediator interaction, and additionally there is no interaction between U and either X, M or C. The rationale for excluding interactions with U is described in a paper by Greenland.[30] Because U is unmeasured, this means that assumptions about interactions are untestable from the data, and small departures from homogeneity of the effects have little impact on the results. Furthermore, equation (8) assumes that U is conditionally independent of $\boldsymbol{C}$ given X. See Hernán and Robins,[31] and VanderWeele[32] for discussion of this assumption.

3.2 Prior distributions for model parameters

We assign uniform prior distributions to the four bias parameters as follows

$$\begin{aligned}
\beta_U &\sim \text{Uniform}(\mu_{\beta_U} - \Delta_{\beta_U}, \mu_{\beta_U} + \Delta_{\beta_U}) \\
\alpha_U &\sim \text{Uniform}(\mu_{\alpha_U} - \Delta_{\alpha_U}, \mu_{\alpha_U} + \Delta_{\alpha_U}) \\
\gamma_0 &\sim \text{Uniform}(\mu_{\gamma_0} - \Delta_{\gamma_0}, \mu_{\gamma_0} + \Delta_{\gamma_0}) \\
\gamma_X &\sim \text{Uniform}(\mu_{\gamma_X} - \Delta_{\gamma_X}, \mu_{\gamma_X} + \Delta_{\gamma_X})
\end{aligned} \tag{9}$$

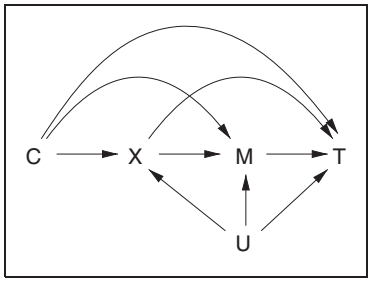

Figure 1. Causal diagram in which the unmeasured confounder U confounds the mediator–outcome, exposure–outcome and exposure–mediator relationship.

These priors approximate a uniform grid of values, which are similar to those encountered in sensitivity analyses where the results are presented as a high-dimensional table (see Table 3 of Chiolero et al.[17] for an example). Each of the four prior distributions are parametrized by a mean parameter and width parameter (i.e. μ_{β_U} and Δ_{β_U}) that must be specified by the investigator to represent the size and direction of unmeasured confounding in the particular context. In Section 4, we give an example of how to choose the prior distributions. For the remaining model parameters, we assign default non-informative priors that are similar to those used in Section 2. We assign $P(\beta_0), P(\beta_X), P(\beta_M), P(\beta_{\boldsymbol{C}}), P(\nu) \propto 1$ where $\nu > 0$ for the parameters in the mortality model, and we assign $P(\alpha_0), P(\alpha_X), P(\alpha_{\boldsymbol{C}}), P(\sigma) \propto 1$ where $\sigma > 0$ for the model for criminal sentencing.

3.3 Sampling from the posterior distribution

We sample from the posterior distribution for model parameters by using the marginal likelihood function that integrates over U. For an independent sample of size n, denote the observed data as $data = \{T_i, M_i, X_i, \boldsymbol{C}_i, \delta_i | i \in 1 \ldots n\}$, where T_i is the time-to-death if $\delta_i = 1$ or the time-to-censoring if $\delta_i = 0$. To simplify notation, let $\beta = (\beta_0, \beta_X, \beta_{\boldsymbol{C}}, \beta_U)$, $\alpha = (\alpha_0, \alpha_X, \alpha_{\boldsymbol{C}}, \alpha_U)$, and $\gamma = (\gamma_0, \gamma_X)$. The marginal likelihood function of the data is

$$L(\beta, \nu, \alpha, \sigma, \gamma) = \prod_{i=1}^{n} \left\{ \sum_{u=0}^{1} P(T_i | M_i, X_i, \boldsymbol{C}_i, U = u, \delta_i) P(M_i | X_i, \boldsymbol{C}_i, U = u) \times P(U = u | X_i, \boldsymbol{C}_i) \right\}$$

where

$$\begin{aligned} P(T_i | M_i, X_i, \boldsymbol{C}_i, U, \delta_i) &\propto \{\exp(\beta_0 + \beta_X X_i + \beta_M M_i + \beta_{\boldsymbol{C}} \boldsymbol{C}_i + \beta_U U) \nu T_i^{\nu-1}\}^{\delta_i} \\ &\quad \times \exp\{-\exp(\beta_0 + \beta_X X_i + \beta_M M_i + \beta_{\boldsymbol{C}} \boldsymbol{C}_i + \beta_U U) T_i^{\nu}\} \\ P(M_i | X_i, \boldsymbol{C}_i, U) &\propto \exp\{-(2\sigma)^{-2} (M_i - \alpha_0 + \alpha_X X_i + \alpha \boldsymbol{C}_i + \alpha_U U))^2\} \\ P(U | X_i, \boldsymbol{C}_i) &\propto \{\text{expit}(\gamma_0 + \gamma_X X_i)\}^U \times \{1 - \text{expit}(\gamma_0 + \gamma_X X_i)\}^{1-U} \end{aligned}$$

and $\text{expit}(a) = (1 + \exp(-a))^{-1}$. The posterior distribution is therefore

$$P(\beta, \nu, \alpha, \sigma, \gamma | data) \propto L(\beta, \nu, \alpha, \gamma, \sigma) P(\beta_U) P(\alpha_U) P(\gamma_0) P(\gamma_X) \tag{10}$$

where equation (10) omits the prior distributions on the remaining parameters $\beta_0, \beta_X, \beta_{\boldsymbol{C}}, \nu, \alpha_0, \alpha_X, \alpha_{\boldsymbol{C}}, \sigma$, which are proportional to 1 and described in Section 3.2.

Sampling from the posterior distribution is difficult using conventional MCMC techniques. The reason is because the model for unmeasured confounding is non-identifiable. Distinct points in the parameter space yield the same probability distribution of the observable data. For example when the bias parameters β_U and α_U are equal to zero, then the likelihood function does not depend on γ_0 and γ_X. Consequently, in large datasets, the posterior distribution will exhibit ridges with extreme correlations between parameters. MCMC techniques have difficulty exploring the parameter space.

We sample from the posterior distribution using a novel computation procedure called Monte Carlo sensitivity analysis (MCSA).[20] To conduct an MCSA, we repeatedly sample bias parameters directly from the prior distributions in equation (9), and then use the sampled values to obtain point estimates for the remaining parameters that are corrected for unmeasured confounding. At each Monte Carlo iteration, random sampling error is incorporated using standard asymptotic approximations to the standard error of the point estimates.[20] The result is a sample of bias-corrected estimates from which we can calculate summary statistics such as the mean, percentiles and interval estimates.

We implement MCSA using the following iterative procedure:

(1) For $t \in 1, \ldots, M$, where M is the number of MCSA simulations:
   (a) Sample bias parameters $(\beta_U^{(t)}, \alpha_U^{(t)}, \gamma_0^{(t)}, \gamma_X^{(t)})$ from the prior distribution in equation (9).
   (b) Compute the bias-corrected estimate $(\hat{\beta}_0^{(t)}, \hat{\beta}_X^{(t)}, \hat{\beta}_M^{(t)}, \hat{\beta}_{\boldsymbol{C}}^{(t)}, \hat{\nu}^{(t)}, \hat{\alpha}_0^{(t)}, \hat{\alpha}_{\boldsymbol{C}}^{(t)}, \hat{\sigma}^{(t)})$ using the STAN function optimizing() in the rstan package, which maximizes the joint log posterior density in equation (10), conditional on $(\beta_U^{(t)}, \alpha_U^{(t)}, \gamma_0^{(t)}, \gamma_X^{(t)})$.

(c) Sample $(\beta_0^{(t)}, \beta_X^{(t)}, \beta_M^{(t)}, \boldsymbol{\beta_C^{(t)}}, \nu^{(t)}, \alpha_0^{(t)}, \boldsymbol{\alpha_C^{(t)}}, \sigma^{(t)})$ from a multivariate normal distribution with mean equal to $(\hat{\beta}_0^{(t)}, \hat{\beta}_X^{(t)}, \hat{\beta}_M^{(t)}, \hat{\boldsymbol{\beta}}_{\boldsymbol{C}}^{(\boldsymbol{t})}, \hat{\nu}^{(t)}, \hat{\alpha}_0^{(t)}, \hat{\boldsymbol{\alpha}}_{\boldsymbol{c}}^{(\boldsymbol{t})}, \hat{\sigma}^{(t)})$ and covariance matrix given by the negative inverse of the estimated Hessian at the solution found in step (b).

(2) Compute the median, 2.5th and 97.5th percentiles of $\{(\beta_0^{(t)}, \beta_X^{(t)}, \beta_M^{(t)}, \boldsymbol{\beta_C^{(t)}}, \nu^{(t)}, \alpha_0^{(t)}, \boldsymbol{\alpha_C^{(t)}}, \sigma^{(t)}) | t \in 1 \ldots M\}$ to obtain BSA point and interval estimates for model parameters.

MCSA is a perfect sampler in the sense that it samples from the desired distribution from the start, and there is no Markov chain simulation involved.[20] Consequently, it is not necessary to assess sampler convergence. MCSA has the advantage that it is much more computationally efficient than direct MCMC simulation from the posterior distribution in equation (10) using STAN. For example in the analysis of the criminal offender data set with sample size $n = 66{,}358$, we obtained an MCSA sample of size 10 from the posterior in about 8 min on a new desktop computer. In contrast, direct MCMC simulation using four parallel chains of length 1000 with overdispersed starting values took 66 min to produce a sample of size 2000 (discarding burn-in), which had an effective sample size of only two. The SRF was 38, which means that the between-chain variance was 38 times greater than the within-chain variance, whereas an SRF < 1.2 is generally indicative of achieving sampler convergence.[25]

4 Analysis of the criminal offender data

We apply BSA to the criminal offender data in order to explore sensitivity of the results to bias from unmeasured confounding. As described in Section 2.4, we are concerned that there may be important variables that predict criminal behaviour and mortality that we were unable to measure in the administrative data. For example we were unable to measure and adjust for social achievement, family and peer relationships and personality traits. Frequent offenders tend to be more impulsive with poorer self-control, and these traits are independent risk factors for higher mortality.[33] The mediator–outcome, exposure–outcome and exposure–mediator relationships are all susceptible to bias from these unmeasured factors.

We assign prior distributions to the bias parameter in equation (9) to describe a broad range of possible bias. We set $\mu_{\beta_U} = \mu_{\alpha_U} = \mu_{\gamma_0} = \mu_{\gamma_X} = 0$ and $\Delta_{\beta_U} = \Delta_{\alpha_U} = \Delta_{\gamma_0} = \Delta_{\gamma_X} = \Delta$, for different values of Δ between zero and one. This choice ensures that the prior distributions for bias are symmetric around zero, which means that we make no presumptions about the particular direction of unmeasured confounding. In contrast, the quantity Δ controls the magnitude of unmeasured confounding. When $\Delta = 1$, then the prior distributions imply that following three quantities are bounded between $\exp(-1) = 0.37$ and $\exp(+1) = 2.71$: The hazard ratio $\exp(\beta_U)$ for the U–T association, the rate ratio $\exp(\alpha_U)$ for the U–M association and finally, the odds ratio $\exp(\gamma_X)$ for the X–U association. Additionally, the prior on γ_0 restricts the prevalence of the unmeasured confounder among participants with $X = 0$ to be bounded between $\text{expit}(-1) = 27$ % and $\text{expit}(+1) = 73$ %. Conversely, when $\Delta = 0$ then equation (9) forces $\beta_U = \alpha_U = \gamma_0 = \gamma_X = 0$, which corresponds to the NAIVE analyses presented in Tables 2 to 4. Thus, letting Δ range between zero and one allows an evaluation of continuous departures from the assumption of no unmeasured confounding in the mediator–outcome, exposure–outcome and exposure–mediator relationships.

The results of BSA are given in Tables 2 to 4. In Table 2, BSA posterior credible intervals for β_M and β_X are much wider than NAIVE. This occurs because U is a confounder for the exposure–outcome and the mediator–outcome associations. The point estimates of β_M and β_X for BSA versus NAIVE are nearly identical because the prior distributions for the bias parameters are symmetric at zero. We see similar results in Table 3. U is a confounder for the X–M association, and consequently, the BSA 95% posterior interval for α_X is expanded compared to the NAIVE interval. A further observation is that BSA and NAIVE give almost identical inferences for each of the regression coefficient $\beta_{C_1}, \ldots, \beta_{C_5}$ (and indeed also for $\alpha_{C_1}, \ldots, \alpha_{C_5}$). This occurs because equation (8) assumes that U is conditionally independent of $\boldsymbol{C}$ given X. This corresponds to the absence of an arrow from U to $\boldsymbol{C}$ in Figure 1, and consequently, U is not a confounder for the $\boldsymbol{C} - T$ or $\boldsymbol{C} - M$ relationships.

The bottom row of Table 4 presents the results of BSA for the causal mediation parameters. For BSA, we see large increases in the width of the 95% intervals for the natural direct and indirect effects, and additionally, the total effect. Consequently, there is a commensurate increase in uncertainty in the estimated proportion mediated. Crucially, if we consider 95% credible interval overlap with zero in order to identify non-zero natural direct and indirect effects, then Table 4 clearly demonstrates that criminal sentencing plays a mediating role in the

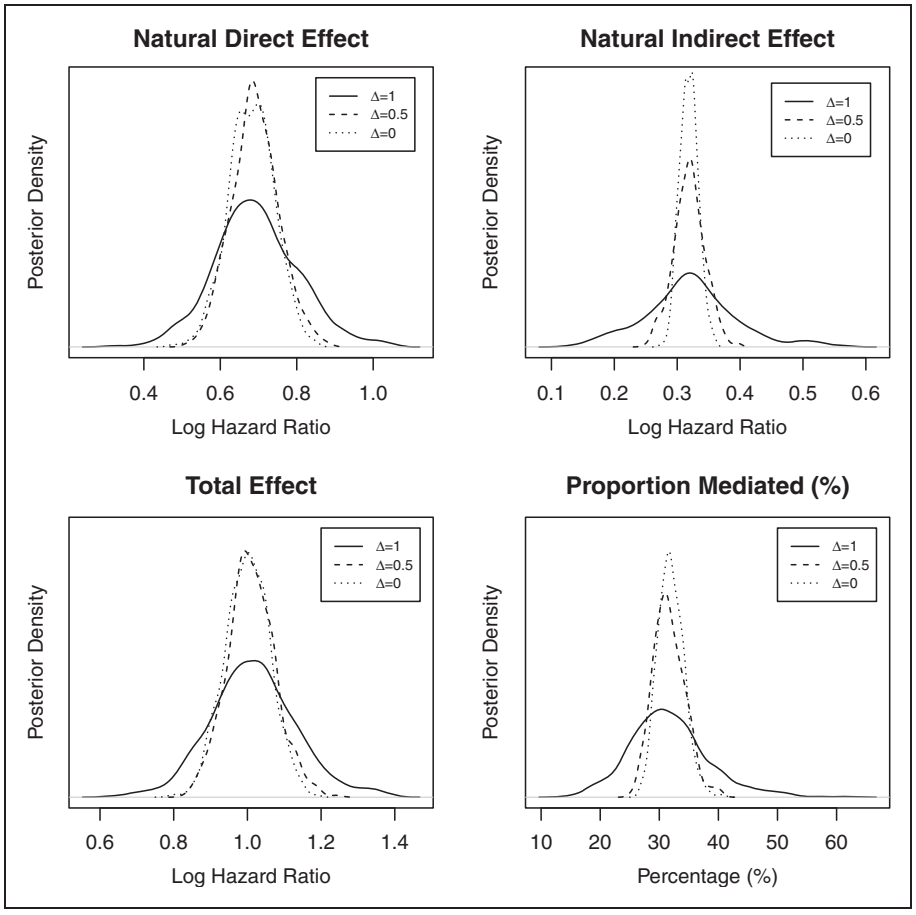

Figure 2. Posterior densities of the causal mediation parameters from Bayesian sensitivity analysis (BSA) applied to the offender data using different priors for the bias parameters. For the prior distributions in equation (9), we set $\mu = 0$ and we used three different values of $\Delta = 0, 0.5$ or 1. When $\Delta = 0$ then there is no unmeasured confounding, and BSA is identical to the NAIVE analysis.

relationship between DDx and mortality. This conclusion is robust to large departures in the assumption of no unmeasured confounding in the mediator–outcome, exposure–outcome and exposure–mediator relationships.

To better illustrate the influence of the prior distribution for the bias parameters on the results, Figure 2 presents posterior densities for the causal mediation parameters using different priors. In each panel, we set $\mu_{\beta_U} = \mu_{\alpha_U} = \mu_{\gamma_0} = \mu_{\gamma_X} = 0$ in the prior distributions in equations (9), and we used different values of Δ that range from zero to one. When $\Delta = 0$, then the BSA assumes there is no unmeasured confounding, and we obtain NAIVE analysis results. When Δ is non-zero, the posterior distributions flatten out. Figure 2 illustrates the surprising result that the natural indirect effect estimate is particularly sensitive to unmeasured confounding. This is because it is calculated from the product of two coefficients, given in equation (5), each of which is susceptible to bias.

5 Simulation studies

In the offender data, the BSA 95% credible intervals are much wider than the corresponding intervals produced by other mediation analysis techniques that ignore unmeasured confounding. Furthermore, Figure 2 reveals that the posterior distributions are very sensitivity to the prior distribution for the bias parameters. It is natural to ask what would happen if the investigator used a poor choice of prior. In a series of papers, McCandless et al.[19] and Gustafson and Greenland[18] demonstrated that if the prior distribution for the bias parameter approximates the true parameter sampling distribution, then BSA 95% interval estimates will have 95% coverage probability, on average, across the parameter space.

To investigate the influence of the prior distribution on the performance of BSA credible intervals for causal mediation parameters, we present simulation results using large data sets that are biased from unmeasured

confounding. Building on the work of Gustafson and Greenland,[18] we randomly sample the true values of bias parameters, denoted as $\beta_U^*, \alpha_U^*, \gamma_0^*, \gamma_X^*$, from various parameter sampling distributions, and then generate simulated data conditional on the bias parameters. We consider data sets of size $n = 10{,}000$ or 50,000, which are generated as follows:

(1) Sample the true bias parameters, as $\beta_U^*, \alpha_U^*, \gamma_0^*, \gamma_X^* \sim \text{Uniform}(-q, q)$, where q is a fixed value chosen to lie between zero and one.
(2) For $i = 1, \ldots, n$:
   (a) Generate the exposure and 5×1 vector of binary confounders, denoted $(X_i, \boldsymbol{C}_i)$, by sampling with replacement from the criminal offender data set.
   (b) Generate the binary unmeasured confounder U_i, conditional on $(X_i, \boldsymbol{C}_i)$, using the logistic regression model

$$U_i | X_i, \boldsymbol{C}_i \sim \text{Bernoulli}(\text{expit}(\gamma_0^* + \gamma_X^* X_i))$$

   (c) Generate the mediator M_i, given $(X_i, \boldsymbol{C}_i, U_i)$, using the linear regression model

$$\begin{aligned} M_i | X_i, \boldsymbol{C}_i, U_i \sim \text{Normal}(&-0.73 + 0.35 X_i + 0.08 C_{1i} + 0.20 C_{2i} - 0.13 C_{3i} \\ &+ 0.11 C_{4i} + 0.05 C_{5i} + \alpha_U^* U_i, 0.66^2) \end{aligned}$$

   where the y-intercept, regression coefficients and variance are taken from Table 3 based on the linear model from equation (3).
   (d) Generate the time-to-death T_i, given $(X_i, M_i, \boldsymbol{C}_i, U_i)$, using the Weibull regression model

$$\begin{aligned} T_i | X_i, M_i, \boldsymbol{C}_i, U_i \sim \text{Weibull}(\exp(&-13 + 0.69 X_i + 0.91 M_i - 0.03 C_{1i} \\ &- 0.67 C_{2i} + 1.17 C_{3i} + 0.22 C_{4i} + -0.13 C_{5i} + \beta_U^* U_i), 1.5), \end{aligned}$$

   where the y-intercept, regression coefficients and shape parameter $\nu = 1.5$ were taken from the Weibull regression results in Table 2.
   (e) Generate a censoring time Q_i, which follows a Gamma distribution with mean and standard deviation equal to 1100 and 760 days, corresponding to the censoring times in the offender data. Then set $\delta_i = 1$ if $T_i \leq Q_i$ and otherwise $\delta_i = 0$ and $T_i = Q_i$.

We analyze the simulated data, given by $\{(T_i, \delta_i, X_i, M_i, \boldsymbol{C_i}) | i \in 1 \ldots n\}$, while discarding the unmeasured confounder $U_1, \ldots, U_n$. We apply BSA to the simulated data where we set $\mu_{\beta_U} = \mu_{\alpha_U} = \mu_{\gamma_0} = \mu_{\gamma_X} = 0$ and $\Delta_{\beta_U} = \Delta_{\alpha_U} = \Delta_{\gamma_0} = \Delta_{\gamma_X} = \Delta = 1$, and additionally, we analyze the data sets using the three methods that ignore unmeasured confounding: NAIVE, and the methods of VanderWeele[5] and Lange et al.[27]

The results are given in Figures 3 and 4 for the case of five simulated datasets with sample size $n = 10{,}000$ or 50,000, respectively. The key finding is that in large data sets, the 95% intervals calculated using methods that ignore unmeasured confounding (e.g. NAIVE) tend to 'miss' the true parameter values because they do not adjust for U. In contrast, the BSA intervals are much wider and more likely to cover the truth because they assign a prior distribution for the bias parameter that exactly matches the true sampling distribution of bias parameters that was used to generate the data.

To examine the performance of BSA when the investigator uses the wrong prior, we repeat the simulations by generating 100 data sets for different scenarios where the true bias parameters are sampled from several different parameter generating distributions. The results are given in Table 5 for data sets of size $n = 10{,}000$ and Table 6 for $n = 50{,}000$. Simulations standard errors for the coverage probability estimates were equal to $\sqrt{0.95 \times 0.05/100} = 2.2$ %. In the top of rows, the true bias parameters $\beta_U^*, \alpha_U^*, \gamma_0^*, \gamma_X^*$ are sampled from a Uniform (−1, 1) distribution, just like in Figures 3 and 4. Thus, BSA uses prior distributions for bias parameters that perfectly match the true parameter generating distribution, and consequently, the 95% credible intervals have approximately 95% coverage probability. In contrast, the 95% intervals calculated using NAIVE, VanderWeele[5] and Lange et al.[27] have worse coverage that is particularly severe when $n = 50{,}000$ because the interval estimates are too narrow and miss the true causal parameters altogether. The method of Lange et al.[27] exhibits the worse coverage. However, as discussed in Section 2.4, it uses a marginal structural model to estimate natural effects, and it is well know that marginal and conditional hazard ratios may differ slightly.[28]

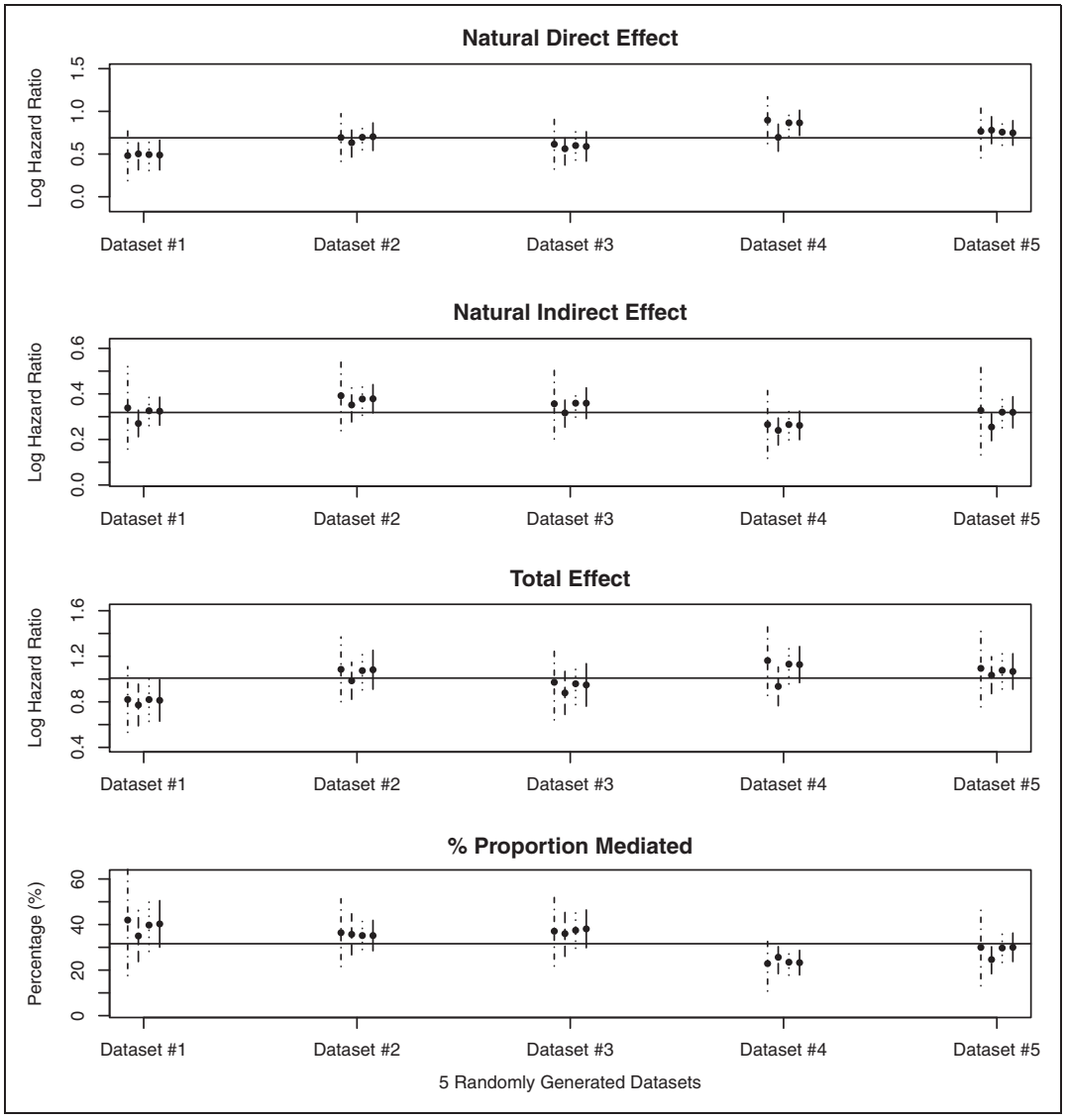

Figure 3. 95% intervals for causal mediation parameters calculated using NAIVE (—), VanderWeele[5] $(\cdots)$, Lange et al.[27] $(- - -)$, or Bayesian sensitivity analysis $(- \cdot -\cdot)$ applied to five simulated data sets (sample size $n = 10{,}000$). In each data set, there is bias from unmeasured confounding that is generated by randomly sampling the true bias parameters $\beta_U^*, \alpha_U^*, \gamma_0^*, \gamma_X^* \sim \text{Uniform}(-1, 1)$. Horizontal lines indicate the true causal mediation parameters.

Tables 5 and 6 reveal what happens when BSA uses the wrong prior for the bias parameters, which does not reflect the true parameter generating distribution. In the bottom rows, the true bias parameters $\beta_U^*, \alpha_U^*, \gamma_0^*, \gamma_X^*$ are fixed at exactly zero, which corresponds to no unmeasured confounding. In this scenario, the 95% interval estimates from BSA are far too wide, and always have 100% coverage probability. In contrast, the NAIVE and VanderWeele[5] intervals have the correct 95% coverage probability. Tables 5 and 6 also illustrate a range of scenarios where NAIVE and BSA both use incorrect assumptions about bias (e.g. a Uniform $(-0.5, 0.5)$). We see that provided BSA uses a prior distribution that correctly approximates the true parameter generating distribution, then BSA 95% intervals will have 95% average coverage probability.

In the simulation studies, large sample sizes were used. Although our focus is on the epidemiological studies using large databases of electronic health records, many applications use smaller data sets and it is important to examine the performance of BSA in small samples. Accordingly, Tables A1 and A2 of the supplementary materials present additional results for $n = 1000$ and 5000. The pattern results are similar to those given in Tables 5 and 6. When the true bias parameters are sample from a Uniform $(-1, 1)$ distribution, then the 95% interval estimates using NAIVE, the methods of VanderWeele[5] and Lange et al.[27] have less than 95% coverage probability,

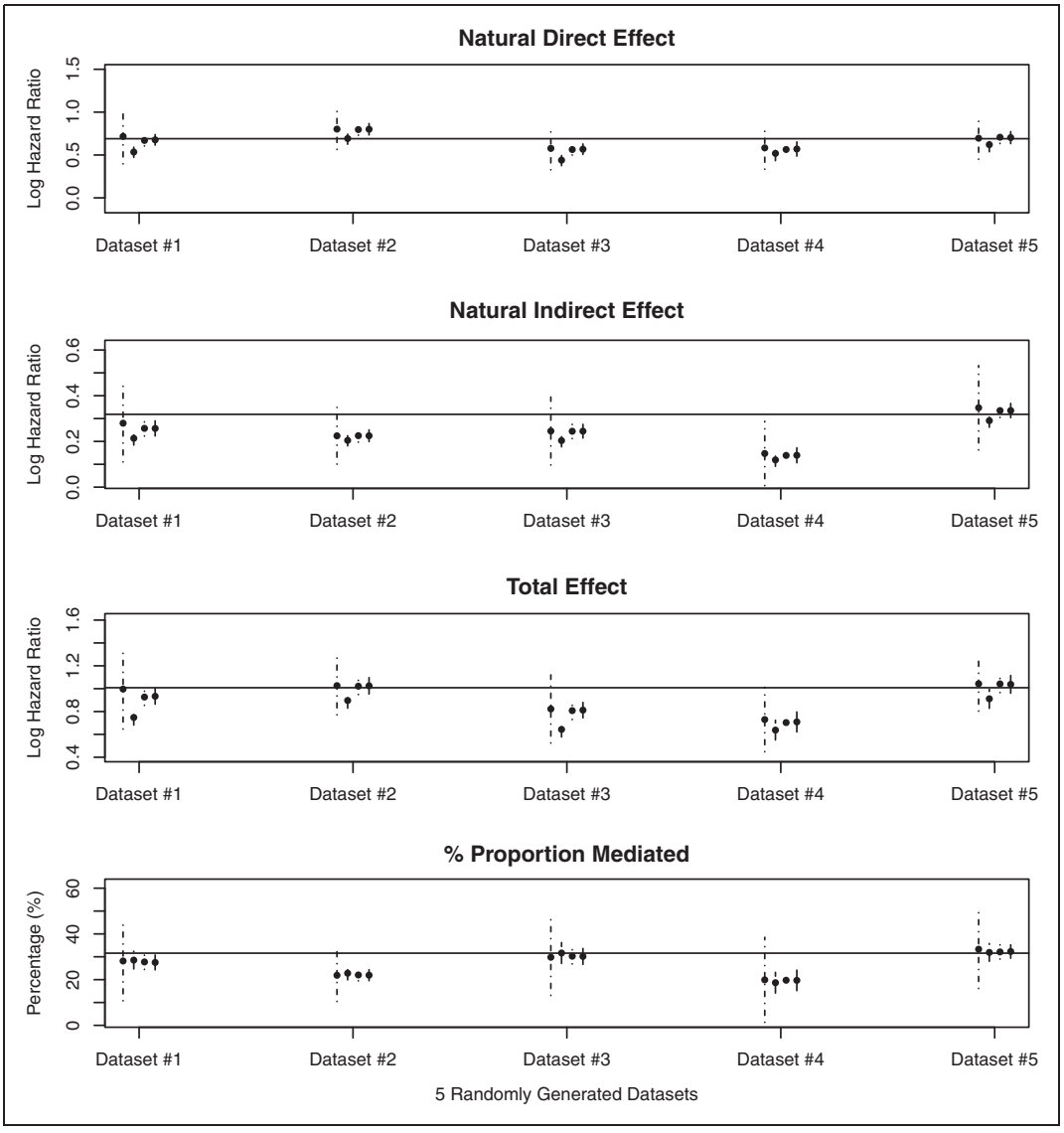

Figure 4. 95% intervals for causal mediation parameters calculated using NAIVE (—), VanderWeele[5] $(\cdots)$, Lange et al.[27] $(- - -)$, or Bayesian sensitivity analysis $(- \cdot -\cdot)$ applied to five simulated data sets (sample size n =50,000). In each data set, there is bias from unmeasured confounding that is generated by randomly sampling the true bias parameters $\beta_U^*, \alpha_U^*, \gamma_0^*, \gamma_X^* \sim$ Uniform$(-1, 1)$. Horizontal lines indicate the true causal mediation parameters.

particularly for the natural indirect effect parameter that is more susceptible to bias from unmeasured confounding. However, if there is only a small amount of bias then BSA intervals are too wide and excessively conservative.

6 Discussion

In this article, we propose a novel Bayesian method to explore sensitivity to unmeasured confounding in causal mediation analysis for natural direct and indirect effects on survival outcomes. Unlike previously published methods, our approach has the unique advantage that it is able to simultaneously assess the effects of unmeasured confounding in the mediator–outcome, exposure–outcome and exposure–mediator relationships. The proposed BSA method is indexed by four bias parameters ($\beta_U, \alpha_U, \gamma_0$ and γ_X) that govern the magnitude of confounding from a binary unmeasured variable U. The investigator assigns prior distributions to each of the parameters to model available information about unmeasured confounding. Because BSA uses a probabilistic

Table 5. Simulation study results based on 100 simulated data sets of sample size $n = 10{,}000$.

	Natural direct effect		Natural indirect effect		Total effect		Proportion mediated (%)	
	Coverage (%)	Length	Coverage (%)	Length	Coverage (%)	Length	Coverage	Length
Results when true bias parameters used to generate the data were sampled $\beta_U^*, \alpha_U^*, \gamma_0^*, \gamma_X^* \sim \text{Unif}(-1, 1)$								
NAIVE	83	0.31	62	0.13	81	0.33	73	13.6
VanderWeele[5]	82	0.31	67	0.14	86	0.34	78	13.6
Lange et al.[27]	79	0.33	53	0.12	71	0.33	77	16.8
BSA[a]	99	0.57	97	0.33	98	0.61	97	31.8
Results when true bias parameters used to generate the data were sampled $\beta_U^*, \alpha_U^*, \gamma_0^*, \gamma_X^* \sim \text{Unif}(-0.75, 0.75)$								
NAIVE	87	0.31	76	0.13	91	0.34	89	13.5
VanderWeele[5]	88	0.31	78	0.14	90	0.34	90	13.6
Lange et al.[27]	84	0.33	64	0.13	73	0.34	88	16.7
BSA	100	0.58	99	0.33	100	0.61	100	31.7
Results when true bias parameters used to generate the data were sampled $\beta_U^*, \alpha_U^*, \gamma_0^*, \gamma_X^* \sim \text{Unif}(-0.5, 0.5)$								
NAIVE	93	0.31	94	0.13	97	0.34	95	13.7
VanderWeele[5]	92	0.31	92	0.14	98	0.34	97	14.0
Lange et al.[27]	89	0.34	68	0.13	76	0.34	93	16.6
BSA[a]	100	0.58	100	0.34	100	0.61	100	31.9
Results when true bias parameters used to generate the data were sampled $\beta_U^*, \alpha_U^*, \gamma_0^*, \gamma_X^* \sim \text{Unif}(-0.25, 0.25)$								
NAIVE	92	0.31	95	0.13	95	0.34	96	13.7
VanderWeele[5]	93	0.32	95	0.14	97	0.34	97	13.9
Lange et al.[27]	90	0.34	71	0.13	77	0.34	95	16.9
BSA[a]	100	0.58	100	0.34	100	0.61	100	32.2
Results when true bias parameters used to generate the data fixed as $\beta_U^*, \alpha_U^*, \gamma_0^*, \gamma_X^* = 0$ (no confounding)								
NAIVE	95	0.31	86	0.13	93	0.34	93	13.5
VanderWeele[5]	98	0.32	89	0.14	95	0.34	95	13.8
Lange et al.[27]	81	0.34	67	0.13	78	0.35	92	17.0
BSA[a]	99	0.58	100	0.34	100	0.62	100	31.8

Coverage probability and average length[b] of 95% interval estimates calculated using BSA, NAIVE and the methods of VanderWeele[5] and Lange et al.[27]
BSA: Bayesian sensitivity analysis.
[a]In all analyses, BSA used the prior $\beta_U, \alpha_U, \gamma_0, \gamma_X \sim \text{Unif}(-1, 1)$ from equation (9).
[b]Simulation standard errors were $\leq$2.2% for coverage probabilities and $\leq$0.01 for average lengths.

approach to sensitivity analysis, it has the advantage of dimension reduction; it gives a single posterior distribution for each of the causal mediation parameters by averaging over uncertainty in the multiple bias parameter inputs.[20]

To illustrate the method, we applied BSA to a data example from epidemiology to examine the mediating role of criminal sentencing on mortality among offenders with mental illness and addictions. Previously, McCandless et al.[21] found an indirect effect of criminal sentencing on higher mortality. However, there were concerns about the possibility of bias from unmeasured confounding variables that predict criminal behaviour and mortality. Our BSA examined the role of unmeasured confounding by assigning uniform prior distributions for $(\beta_U, \alpha_U, \gamma_0, \gamma_X)$ that spanned a broad range of possible bias due to an omitted binary variable U. The natural indirect effects were particularly sensitive to unmeasured confounding. However, the 95% credible intervals in Table 4 clearly support the conclusion of McCandless et al.[21] that criminal sentencing has an indirect effect on mortality among offenders with mental illness and addictions.

A concern with the BSA approach is that the analysis results are very sensitive to the choice of prior distributions for bias parameters. Figure 2 shows that changing the prior distribution will affect the posterior distribution of the causal mediation parameters, and furthermore, this sensitivity occurs despite the large sample size of the data set. However, our simulation results in Section 5 show that if the prior distribution for bias parameter approximates the true parameter generating distribution that was used to sample the data, then BSA credible intervals will have better frequentist coverage probability compared to standard mediation techniques that ignore unmeasured confounding.

A further concern is that because we assign default uniform prior distributions for model parameters then the resulting posterior distributions could be improper. However, Chen et al.[34] describes sufficient conditions for the propriety of the posterior with improper uniform priors on the regression coefficients for general types of regression models and parametric survival models with right censoring. They show that when the design matrix

Table 6. Simulation study results based on 100 simulated datasets of sample size $n = 50{,}000$.

	Natural direct effect		Natural indirect effect		Total effect		Proportion mediated (%)	
	Coverage (%)	Length	Coverage (%)	Length	Coverage (%)	Length	Coverage	Length
Results when true bias parameters used to generate the data were sampled $\beta_U^*, \alpha_U^*, \gamma_0^*, \gamma_X^* \sim \text{Unif}(-1, 1)$								
NAIVE	56	0.14	38	0.06	60	0.15	44	6.1
VanderWeele[5]	60	0.14	42	0.06	63	0.15	49	6.2
Lange et al.[27]	50	0.15	24	0.06	23	0.15	59	7.4
BSA[a]	100	0.50	91	0.31	97	0.52	96	29.1
Results when true bias parameters used to generate the data were sampled $\beta_U^*, \alpha_U^*, \gamma_0^*, \gamma_X^* \sim \text{Unif}(-0.75, 0.75)$								
NAIVE	76	0.14	60	0.06	78	0.15	66	6.1
VanderWeele[5]	76	0.14	59	0.06	76	0.15	64	6.2
Lange et al.[27]	42	0.15	26	0.06	19	0.15	74	7.4
BSA[a]	100	0.50	99	0.31	100	0.52	99	29.4
Results when true bias parameters used to generate the data were sampled $\beta_U^*, \alpha_U^*, \gamma_0^*, \gamma_X^* \sim \text{Unif}(-0.5, 0.5)$								
NAIVE	89	0.14	78	0.06	89	0.15	86	6.1
VanderWeele[5]	91	0.14	78	0.06	90	0.15	86	6.1
Lange et al.[27]	37	0.15	23	0.06	15	0.15	91	7.4
BSA[a]	100	0.50	100	0.31	100	0.52	100	28.9
Results when true bias parameters used to generate the data were sampled $\beta_U^*, \alpha_U^*, \gamma_0^*, \gamma_X^* \sim \text{Unif}(-0.25, 0.25)$								
NAIVE	97	0.13	97	0.06	97	0.14	95	5.8
VanderWeele[5]	97	0.14	98	0.06	98	0.15	95	5.9
Lange et al.[27]	40	0.14	24	0.05	12	0.15	97	7.1
BSA[a]	100	0.50	100	0.32	100	0.52	100	29.4
Results when true bias parameters used to generate the data fixed as $\beta_U^*, \alpha_U^*, \gamma_0^*, \gamma_X^* = 0$ (no confounding)								
NAIVE	93	0.13	96	0.06	94	0.14	91	5.9
VanderWeele[5]	94	0.13	95	0.06	93	0.15	92	5.8
Lange et al.[27]	37	0.14	19	0.05	9	0.15	93	7.1
BSA[a]	100	0.50	100	0.31	100	0.53	100	28.1

Coverage probability and average length[b] of 95% interval estimates calculated using BSA, NAIVE and the methods of VanderWeele[5] and Lange et al.[27]
BSA: Bayesian sensitivity analysis.
[a]In all analyses, BSA used the prior $\beta_U, \alpha_U, \gamma_0, \gamma_X \sim \text{Unif}(-1, 1)$ from equation (9).
[b]Simulation standard errors were $\leq$2.2% for coverage probabilities and $\leq$0.01 for average lengths.

for the regression model is full rank, then the posterior distribution is proper provided that the log-likelihood function is log-concave. For Sections 2.2, 2.3 and 2.4, we assign uniform priors in Bayesian analysis of standard identifiable parametric models (e.g. multiple linear regression) with log concave density functions. For Section 3, BSA model given in equation (10) is non-identifiable. However, the MCSA algorithm maximizes the log posterior conditional on the bias parameters, and this posterior is log concave because the likelihood function comprises standard identifiable parametric models for the mediator M and survival time T.

We emphasize that a limitation of this discussion is that we have examined unmeasured confounding within the specific parametric models from equations (6) to (8). We did not consider other bias correction formulas based on less parametric approaches such as double-robust methods or direct modelling of counterfactual outcomes.[35] Furthermore, the foregoing analyses explores sensitivity to bias from a single binary variable U. But in some settings there may be several unmeasured confounders that are correlated with one another, and developing suitable methods is an area of ongoing research.[36,37] Realistic assessments of uncertainty in observational studies should ideally use a multiple bias modelling approach.[20]

A further limitation of our proposed methodology is that the model in equation (8) assumes that U is conditionally independent of $\boldsymbol{C}$ given X. Hernán and Robins[31] and VanderWeele[32] argue that if $\boldsymbol{C}$ are confounders for either the X–M or the X–T relationship then this conditional independence assumption cannot hold. Nonetheless, assuming conditional independence has practical advantages for sensitivity analysis. If U depends on $\boldsymbol{C}$, given X, then there are more bias parameters to consider.[30] In principle, we could modify equation (8) to include an additional linear predictor $\ldots + \gamma_C^T \boldsymbol{C}$ that depends on the bias parameter vector γ_C. But this requires that the user specify a prior distribution for each of the individual components of γ_C. In practice, there is often little information available about the relationship between U and individual components of $\boldsymbol{C}$ in order to guide a sensitivity analysis.

Acknowledgements

The authors gratefully acknowledge input and thoughtful comments from Erica Moodie on an earlier version of the manuscript.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (grant 371518-2013).

References

1. Hafeman DM and Schwartz S. Opening the Black Box: a motivation for the assessment of mediation. *Int J Epidemiol* 2009; **38**: 838–845.
2. Imai K, Keele L and Tingley D (2010). A general approach to causal mediation analysis. *Psychol Methods* 15: 309–334.
3. VanderWeele TJ. *Explanation in Causal Inference: Methods for Mediation and Interaction*. New York: Chapman Hall CRC, 2015.
4. Lange T and Hansen JV. Direct and indirect effects in a survival context. *Epidemiology* 2011; **22**: 575–581.
5. VanderWeele TJ. Causal mediation analysis with survival data. *Epidemiology* 2011; **22**: 582–585.
6. VanderWeele TJ and Chiba Y. Sensitivity analysis for direct and indirect effects in the presence of exposure-induced mediator-outcome confounders. *Epidemiol Biostat Public Health* 2014; **11**: e9027.
7. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997; **127**: 575–763.
8. Schuemie MJ, Ryan PB, DuMouchel W, et al. Interpreting observational studies: why empirical calibration is needed to correct *p*-values. *Stat Med* 2014; **33**: 209–218.
9. VanderWeele TJ. Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology* 2010; **21**: 540–551.
10. VanderWeele TJ. Unmeasured confounding and hazard scales: sensitivity analysis for total, direct, and indirect effects. *Eur J Epidemiol* 2013; **28**: 113–117.
11. Kim C, Daniels MJ, Marcus BH, et al. A framework for Bayesian nonparametric inference for causal effects of mediation. *Biometrics* 2017; **73**: 401–409.
12. Daniels MJ, Roy JA, Kim C, et al. Bayesian inference for the causal effect of mediation. *Biometrics* 2012; **68**: 1028–1036.
13. Imai K, Keele L and Yamamoto T. Identification, inference and sensitivity analysis for causal mediation effects. *Stat Sci* 2010; **25**: 51–71.
14. Albert JM and Wang W. Sensitivity analyses for parametric causal mediation effect estimation. *Biostatistics* 2015; **16**: 339–351.
15. Hafeman DM. Confounding of indirect effects: a sensitivity analysis exploring the range of bias due to a cause common to both the mediator and the outcome. *Am J Epidemiol* 2011; **174**: 710–717.
16. Baron RM and Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psycol* 1986; **51**: 1173–1182.
17. Chiolero A, Paradis G and Kaufman JS. Assessing the possible direct effect of birth weight on childhood blood pressure: a sensitivity analysis. *Am J Epidemiol* 2014; **179**: 4–11.
18. Gustafson P and Greenland S. Interval estimation for messy observational data. *Stat Sci* 2009; **24**: 328–342.
19. McCandless LC, Gustafson P and Levy AR. Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Stat Med* 2007; **26**: 2331–2347.
20. Greenland S. Multiplebias modelling for analysis of observational data. *J Roy Stat Soc Ser A* 2005; **168**: 267–306.
21. McCandless LC, Stewart LC, Rempel ES, et al. Criminal justice system contact and mortality among offenders with mental illness in British Columbia: an assessment of mediation. *J Epidemiol Community Health* 2014; **69**: 460–466.
22. Somers JM, Moniruzzaman A, Rezansoff SN, et al. The prevalence and geographic distribution of complex co-occurring disorders: a population study. *Epidemiol Psychiatr Sci* 2014; **25**: 267–277.
23. Stan Development Team. 2016. *RStan: the R interface to Stan*. R package version 2.12.1. http://mc-stan.org
24. Gelman A, Carlin JB, Stern HS, et al. *Bayesian data analysis*, 3rd ed. New York: Chapman Hall CRC, 2014.
25. Gelman A and Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci* 1992; **7**: 457–472.
26. Yuan Y and MacKinnon DP. Bayesian mediation analysis. *Psychol Methods* 2009; **14**: 301–322.

27. Lange T, Vansteelandt S and Bekaert M. A simple unified approach for estimating natural direct and indirect effects. *Am J Epidemiol* 2012; **176**: 190–195.
28. Austin PC. The performance of different propensity score methods for estimating marginal hazard ratios. *Stat Med* 2013; **32**: 2837–2849.
29. Lin DY, Psaty BM and Kronmal RA. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics* 1998; **54**: 948–963.
30. 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. *J Am Stat Assoc* 2003; **98**: 47–54.
31. Hernán MA and Robins JM. ''Letter to the editor of Biometrics''. *Biometrics* 1999; **55**: 1316.
32. VanderWeele TJ. Sensitivity analysis: Distributional assumptions and confounding assumptions. *Biometrics* 2008; **64**: 645–649.
33. Tremblay P and Paré P. Crime and destiny: patterns in serious offenders' mortality rates. *Can J Criminol Crim Justice* 2003; **45**: 299–326.
34. Chen MH, Ibrahim JG and Shao Q. Propriety of the posterior distribution and existence of the MLE for regression models with covariates missing at random. *J Am Stat Assoc* 2004; **99**: 421–438.
35. Dorie V, Harada M, Carnegie NB, et al. A flexible, interpretable framework for assessing sensitivity to unmeasured confounding. *Stat Med* 2016; **35**: 3453–3470.
36. McCandless LC, Richardson S and Best N. Adjustment for missing confounders using external validation data and propensity scores. *J Am Stat Assoc* 2012; **107**: 40–51.
37. Groenwold RH, Stern JA, Lawlor DA, et al. Sensitivity analysis for the effects of multiple unmeasured confounders. *Ann Epidemiol* 2016; **26**: 605–611.