

Simplified Bayesian Sensitivity Analysis for Mismeasured and Unobserved Confounders

P. Gustafson*

Department of Statistics, University of British Columbia, Vancouver, B.C., VRT 1Z2, Canada

**email:* gustaf@stat.ubc.ca

and

L. C. McCandless*

Faculty of Health Sciences, Simon Fraser University, Burnaby, B.C. V5A 1S6, Canada

**email:* mccandless@sfu.ca

and

A. R. Levy*

Department of Community Health and Epidemiology, Dalhousie University, Halifax, N.S. B3H 1V7, Canada

**email:* adrian.levy@dal.ca

and

S. Richardson*

Centre for Biostatistics and MRC-HPA Centre for Environment and Health,

Division of Epidemiology, Public Health and Primary Care, Imperial College, London SW7 2AZ, U.K.

**email:* sylvia.richardson@imperial.ac.uk

SUMMARY: We examine situations where interest lies in the conditional association between outcome and exposure variables, given potential confounding variables. Concern arises that some potential confounders may not be measured accurately, while others may not be measured at all. Some form of *sensitivity analysis* might be employed, to assess how this limitation in available data impacts inference. A Bayesian approach to sensitivity analysis is straightforward in concept: a

??? 20??

prior distribution is formed to encapsulate plausible relationships between unobserved and observed variables, and posterior inference about the conditional exposure-disease relationship then follows. In practice, though, it can be challenging to form such a prior distribution in both a realistic and simple manner. Moreover, it can be difficult to develop an attendant Markov chain Monte Carlo (MCMC) algorithm which will work effectively on a posterior distribution arising from a highly nonidentified model. In this paper, a simple prior distribution for acknowledging both poorly measured and unmeasured confounding variables is developed. It requires that only a small number of hyperparameters be set by the user. Moreover, a particular computational approach for posterior inference is developed, since application of MCMC in a standard manner is seen to be ineffective in this problem.

KEY WORDS: Bayesian inference; Measurement error; Sensitivity analysis; Unobserved confounder.

1. Introduction

All diseases can be thought of as having multiple component causes (Rothman, 1976). Investigators typically focus on one suspected component cause - the “exposure” - and collect information on other suspected component causes - potential “confounders” - which they think may be part of one or more sufficient causal complexes. However, this process is extremely challenging. It is never possible to identify all component causes, so some potential confounders go unmeasured. Moreover, often the identified causes are difficult to measure accurately on individuals. Thus it is common to regress a disease-outcome variable Y on an exposure of interest X and a collection of potential confounding variables \mathbf{W} , even when \mathbf{W} may be inadequate in several regards. Investigators may find it challenging to identify and measure all potential confounders, so that \mathbf{W} is missing some variables it ought to contain. Moreover, some components of \mathbf{W} may be only surrogates for actual confounders, i.e., covariate measurement error may be at play. Often issues of unobserved and poorly measured covariates might be discussed qualitatively, as an adjunct to quantitative results from regression of Y on (X, \mathbf{W}) . This is not fully satisfactory, however, since intuition on how much to ‘downweight’ or ‘widen’ inferences in light of such concerns is typically lacking.

A more formal alternative is *sensitivity analysis*, where a list of different scenarios for the nature of unobserved and poorly-measured confounders is constructed. Inferences about the target parameter are then reported for each scenario. Sensitivity analysis for unobserved confounding has a long history dating back to Cornfield et al. (1959); see Greenland (1996) and Rosenbaum (2002) for reviews. On the other hand, dealing with poorly measured confounders is not typically regarded as sensitivity analysis, as corrections for covariate measurement error are often undertaken in settings where the magnitude of measurement error is known or estimable, in which case the problem is regarded as one of inference.

Outside of such settings, however, some form of sensitivity analysis, considering a variety of measurement-error magnitudes, may be required or desirable.

An inherent challenge with sensitivity analysis is how to summarize and interpret different inferences arising from various scenarios for the nature of missing and/or poorly-measured variables. In addition to concern about ‘table explosion,’ it is difficult to synthesize sampling error on the one hand (within-scenario uncertainty about the target parameter arising from having only a finite amount of data), and across-scenario variation on the other hand.

Bayesian methods offer a route to synthesizing within-scenario and across-scenario uncertainty, via a prior distribution which weights possible scenarios. Then the posterior distribution of the target parameter reflects both uncertainties simultaneously. Recently, McCandless et al. (2007; 2008) propose Bayesian sensitivity analysis (BSA) to deal with unobserved confounders specifically, while Greenland (2003; 2005) uses a related approach, Monte Carlo Sensitivity Analysis (MCSA), to deal with unobserved confounding and other study limitations simultaneously. MCSA has a Bayesian flavour, in that values for parameters describing the data limitations are sampled from prior distributions, with the view that such parameters are not informed by the observed data, so that their posterior distribution equates with their prior distribution. This simplifies computation considerably. In some settings though, it is hard to verify that the data are completely uninformative for particular parameters without actually undertaking a fully Bayesian analysis.

Even when BSA is instantiated for unobserved confounders alone, it become quite complicated, in terms of the complexity of prior specification, and the number of *hyperparameters* which must be set by the user. Adding poorly-measured confounders into the mix has the potential to exacerbate this problem severely. Thus the focus of the present paper is to pursue realistic modeling of the link between observed and unobserved variables, while striving for simplicity in terms relying on only a few readily comprehended assumptions, and

not requiring an excessive number of inputs from the user. The latter point is particularly important. When many tuning parameters are required in a sensitivity analysis, it becomes more difficult to disentangle the key quantities driving the sensitivity. We emphasize the fact that a simple approach is required for a method to become adopted in practice, referring to our proposed method as *simplified Bayesian sensitivity analysis* (SBSA).

In addition to the general goal of finding a balance between realism and simplicity in conducting sensitivity analysis, we are also motivated by the notion that failing to acknowledge limitations of available confounders (i.e., treating \mathbf{W} as if it were complete and measured accurately) has considerable potential to induce spurious findings from observational studies. Recently, Fewell et al. (2007) underscore this danger in an epidemiological context, and Brunner and Austin (2009) make similar points in a generic context. Thus we seek to investigate the ability of SBSA to mitigate this problem, and the ‘cost’ of such mitigation in terms of interval estimate width and power to detect non-null exposure-disease relationships.

2. Methodological Details

2.1 General Model Specification

Consider a health outcome Y , exposure X , and confounders $\mathbf{Z} = (Z_1, \dots, Z_p)'$ and $\mathbf{U} = (U_1, \dots, U_q)'$. In the idealized situation of having no limitations on data quality or availability, we might fit a generalized linear model relating Y to $(X, \mathbf{Z}, \mathbf{U})$, taking inference about the X coefficient to represent the ‘causal’ effect of X on Y . As typifies practice, however, consider a setting with severe limitations on data quality and availability. Particularly, measurements of \mathbf{U} are unavailable, while only noisy measurements \mathbf{W} are available in place of \mathbf{Z} . Thus we refer to \mathbf{U} and \mathbf{Z} as unobserved and near-observed confounders respectively.

A model for the observed data (\mathbf{W}, Y, X) can be obtained by marginalizing a model distribution for the complete data $(\mathbf{W}, Y, \mathbf{U}, \mathbf{Z}, X)$, which in turn might be constructed

via a product of conditional models. The factorization we pursue is

$$\begin{aligned} f_{\boldsymbol{\theta}}(\mathbf{W}, Y|X) &= \int f_{\boldsymbol{\theta}}(\mathbf{W}, Y, \mathbf{U}, \mathbf{Z}|X) d\mathbf{U} d\mathbf{Z} \\ &= \int f_{\boldsymbol{\theta}}(\mathbf{W}|Y, \mathbf{U}, \mathbf{Z}, X) f_{\boldsymbol{\theta}}(Y|\mathbf{U}, \mathbf{Z}, X) f_{\boldsymbol{\theta}}(\mathbf{U}|\mathbf{Z}, X) f_{\boldsymbol{\theta}}(\mathbf{Z}|X) d\mathbf{U} d\mathbf{Z}, \end{aligned} \quad (1)$$

where $\boldsymbol{\theta}$ comprises all unknown parameters in the constituent conditional models. Thus (1) defines a likelihood function for $\boldsymbol{\theta}$, which can be combined with a prior distribution to yield *a posteriori* inferences. Typically simplifications arise. For instance, under a *nondifferential* and classical measurement error assumption, $f_{\boldsymbol{\theta}}(\mathbf{W}|Y, \mathbf{U}, \mathbf{Z}, X)$ reduces to $f_{\boldsymbol{\theta}}(\mathbf{W}|\mathbf{Z})$. In the special case that some of the conditional models involve normal linear models for continuous variables, we gain further simplification by treating some observed moments as known, as described in the next subsection.

2.2 The normal, linear case

Say that all the variables are continuous, and that each W_j is a noisy surrogate for Z_j arising via unbiased, homoscedastic, and nondifferential measurement error, i.e., \mathbf{W} and (Y, \mathbf{U}, X) are conditionally independent given \mathbf{Z} . Moreover, the measurement errors for the components of \mathbf{Z} are assumed to be uncorrelated with each other. This situation is modeled as

$$(\mathbf{W}|Y, \mathbf{U}, \mathbf{Z}, X) \sim N_p\left\{\mathbf{Z}, \text{diag}(\tau_1^2, \dots, \tau_p^2)\right\}. \quad (2)$$

One key to making our method simple is to pretend that $\Sigma = \text{Var}\{(X, \mathbf{W}')\}$ is known. That is, in practice we set Σ equal to the corresponding sample covariance matrix. Moreover, it is useful to standardize (X, \mathbf{W}') before applying SBSA, with the consequence that our chosen Σ will have unit diagonal elements. Inferences can then be ‘transformed back’ to the original scale of the X measurements particularly. The pretense that Σ is known does ignore a small amount of uncertainty, but has the advantage of letting us capture the important uncertainties without excessively burdensome prior specification. Note that

under the standardization the intra-class correlation for describing the measurement error magnitude is $ICC_j = \text{Var}(Z_j)/\text{Var}(W_j) = 1 - \tau_j^2$. This is a useful scale on which to regard the measurement error, since ICC_j is invariant to a common affine transformation of Z_j and W_j .

With measurement model (2) we have $\text{Var}(\mathbf{W}) = \text{Var}(\mathbf{Z}) + \text{diag}(\tau_1^2, \dots, \tau_p^2)$ and $\text{Cov}(W_j, X) = \text{Cov}(Z_j, X)$. Thus, with Σ known, if we wish to specify a normal model for the distribution of exposure and near-observed confounders in the population, it must be:

$$\begin{pmatrix} X \\ \mathbf{Z} \end{pmatrix} \sim N_{p+1} \{0, \tilde{\Sigma}(\boldsymbol{\tau}^2)\}, \quad (3)$$

where $\tilde{\Sigma}(\boldsymbol{\tau}^2) = \Sigma - \text{diag}(0, \tau_1^2, \dots, \tau_p^2)$. Note that (3) forces us to restrict the parameter space for $\boldsymbol{\tau}^2$ to values for which $\tilde{\Sigma}(\boldsymbol{\tau}^2)$ is positive definite, a point we return to presently.

Next consider modeling the unobserved confounders \mathbf{U} given the near-observed confounders \mathbf{Z} and the exposure X . For ease of exposition and interpretation we focus on the $q = 1$ case of a single unobserved confounder. Extension to multiple components in \mathbf{U} is possible, but would be more cumbersome to express, and would require modeling assumptions about how the components of \mathbf{U} relate to each other. Most approaches to sensitivity analysis for unobserved confounding do assume a single unobserved confounder, and to some extent this subsumes the more general case. That is, a *linear combination* of multiple unobserved confounders acting on the outcome can be regarded as a single unobserved confounder. This seems reasonable when the methodology is envisioned as a ‘black box,’ whereby the investigator is concerned about the possible existence of one or more important confounders whose identities are unknown. In contrast, explicit modelling of multiple unobserved confounders seems more appropriate if the investigator knows the identity of the confounders (perhaps smoking-status and alcohol-consumption, for instance), but is unable to measure them.

As a tractable model linking U to (X, \mathbf{Z}) we take

$$\left\{ U \left| \begin{pmatrix} X \\ \mathbf{Z} \end{pmatrix} \right. \right\} \sim N(\gamma_x X + \gamma'_z \mathbf{Z}, c^2), \quad (4)$$

to reflect dependencies between the unobserved confounder and exposure, and between the unobserved confounder and the observed confounders. Many suggested sensitivity analyses make the unrealistic but simplifying assumption that unobserved and observed confounders are conditionally independent given exposure (Lin et al., 1998; Greenland, 2003; McCandless et al., 2008), but the realism of this assumption has been questioned (Hernán and Robins, 1999; VanderWeele, 2008). Thus allowing a non-zero γ_z is more realistic. The conditional independence assumption is generally too conservative, as it implies that adjusting for \mathbf{Z} alone has no ability to partially adjust for U . For reasons to be explained shortly, we take $\gamma' = (\gamma_x, \gamma'_z)$ in (4) to be an unknown parameter vector, while c^2 is a fixed hyperparameter.

In the case of a continuous outcome variable Y , the outcome model can be taken, in customary fashion, as

$$(Y|U, \mathbf{Z}, X) \sim N(\alpha_0 + \alpha_x X + \beta_u U + \beta'_z \mathbf{Z}, \sigma^2). \quad (5)$$

We work with (5) at present, but consider extension to the binary outcome case in Section 3.2.

Note that each of (2), (3), (4) and (5) contribute a term in the factorization (1), either directly, or in the case of (3) via the joint model for (X, \mathbf{Z}) implying the conditional model

$$\mathbf{Z}|X \sim N_p \left\{ \boldsymbol{\mu} X, M - \text{diag}(\boldsymbol{\tau}^2) \right\},$$

where $\boldsymbol{\mu} = \Sigma_{zx} \Sigma_{xx}^{-1}$ and $M = \Sigma_{zz} - \Sigma_{zx} \Sigma_{xx}^{-1} \Sigma_{xz}$ are known. Following (1) then, we have a likelihood function for the unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\tau}^2, \sigma^2)$. As a minor point, as

specified the model has a redundancy in the sense that the sign of U is arbitrary. We can remove this redundancy, say by constraining $\gamma_x > 0$.

2.3 Prior Specification

We structure our prior distribution as $\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\alpha})\pi(\boldsymbol{\beta})\pi(\boldsymbol{\gamma}|\boldsymbol{\tau}^2)\pi(\boldsymbol{\tau}^2)\pi(\sigma^2)$. Both $\boldsymbol{\alpha}$ and σ^2 are assigned standard diffuse priors, i.e., a diffuse (bivariate) normal distribution for $\boldsymbol{\alpha}$ and a diffuse inverse-gamma distribution for σ^2 . For $\boldsymbol{\beta}$ we take an exchangeable, but not independent, prior across all $p + 1$ elements, i.e., the near-observable confounders \mathbf{Z} and the unobservable confounder U are treated exchangeably. Particularly, we use a multivariate t distribution, as arising as a scale mixture of independent and identically distributed normal variates. This entails specifying two hyperparameters: a scale parameter r and a degrees-of-freedom d . More explicitly, this prior arises via components of $\boldsymbol{\beta}$ which are independent and identically distributed as $N(0, r^2 S)$ given S , with $S \sim IG(d/2, d/2)$. McCandless et al. (2008) argue for this form of prior in BSA-style problems, showing that sometimes it yields substantially narrower inferences than an independence prior having the same marginal variability, i.e., a prior under which components of $\boldsymbol{\beta}$ are independent and identically distributed as $t_d(0, r^2)$.

Specification of the $\pi(\boldsymbol{\tau}^2)$ and $\pi(\boldsymbol{\gamma}|\boldsymbol{\tau}^2)$ terms is more subtle. Recall that $1 - \tau_j^2$ is the ICC describing the reliability of W_j as a surrogate for Z_j . We *start* by thinking of a prior under which each τ_j^2 is independently distributed as $\text{Beta}(a_j, b_j)$. Let $g(\cdot; a, b)$ denote the $\text{Beta}(a, b)$ density function. In light of the positive definiteness constraint, the *actual* prior distribution used takes the form

$$\pi(\boldsymbol{\tau}^2) \propto \left\{ \prod_{j=1}^p g(\tau_j^2; a_j, b_j) \right\} I_A(\boldsymbol{\tau}^2),$$

where A is the set of all $\boldsymbol{\tau}^2$ for which $\tilde{\Sigma}(\boldsymbol{\tau}^2)$ is positive definite. It is simple to verify that A is the set for which the largest eigenvalue of $M^{-1}\text{diag}(\boldsymbol{\tau}^2)$ is less than one. In practice

we suggest checking the extent to which truncation to A yields properties of the actual prior distribution differing from those of the starting prior distribution. In general, when the starting prior puts most of its weight on values of $\boldsymbol{\tau}^2$ corresponding to modest measurement error, we anticipate little effect of truncation.

An important part of keeping the method simple is specifying a very particular form of *a priori* dependence between $\boldsymbol{\gamma}$ and $\boldsymbol{\tau}^2$. Particularly, we take

$$(\boldsymbol{\gamma}|\boldsymbol{\tau}^2) \sim N_{p+1} \left[0, k^2 \left\{ \tilde{\Sigma}(\boldsymbol{\tau}^2) \right\}^{-1} \right], \quad (6)$$

where k^2 is a fixed hyperparameter. Thus the prior covariance matrix for the $(U|X, \mathbf{Z})$ regression coefficients is proportional to the precision matrix for the regressors. This is actually a common specification, corresponding to the ‘‘g-prior’’ pioneered by Zellner (1986) applied to centered regressors. In contrast to the present setting, g-priors are typically applied when all predictors are observed, so that a sample precision matrix is employed. In the present situation (6) happens to be particularly tractable and interpretable, as (3) and (4) imply that $\text{Var}(U) = c^2 + \boldsymbol{\gamma}'\tilde{\Sigma}(\boldsymbol{\tau}^2)\boldsymbol{\gamma}$. Thus from (6) we have that $\text{Var}(U)$ is *a priori* distributed as $c^2 + k^2 R$, where $R \sim \chi_{p+1}^2$. We can use this fact to choose hyperparameters under which U and \mathbf{Z} are of comparable magnitude *a priori*, which in turn supports the use of an exchangeable prior on $\boldsymbol{\beta}' = (\beta_u, \boldsymbol{\beta}_z')$.

Similarly, note that $\text{Cov}\{(X, \mathbf{Z}')', U\} = \tilde{\Sigma}(\boldsymbol{\tau}^2)\boldsymbol{\gamma}$. Now, with respect to the specified prior distribution we have

$$\begin{aligned} E \left\{ \|\tilde{\Sigma}(\boldsymbol{\tau}^2)\boldsymbol{\gamma}\|^2 \right\} &= EE \left\{ \|\tilde{\Sigma}(\boldsymbol{\tau}^2)\boldsymbol{\gamma}\|^2 | \boldsymbol{\tau}^2 \right\} \\ &= k^2 E \left[\text{tr} \left\{ \tilde{\Sigma}(\boldsymbol{\tau}^2) \right\} \right] \\ &= k^2 E \left\{ 1 + \sum_{j=1}^p (1 - \tau_j^2) \right\} \\ &\approx k^2 \left\{ 1 + \sum_j b_j / (a_j + b_j) \right\}, \end{aligned} \quad (7)$$

where the final expression involves approximation because of the truncation of the prior distribution on $\boldsymbol{\tau}^2$. Hence the average squared covariance between U and an element of (X, \mathbf{Z}') is approximately $k^2(p+1)^{-1} [1 + p \text{avg}_j\{b_j/(a_j + b_j)\}] \approx k^2 \text{avg}_j\{b_j/(a_j + b_j)\}$, where the averaging is both with respect to the prior distribution and across the elements of (X, \mathbf{Z}') . The expression (7) will be useful in the determination of default hyperparameter values.

2.4 Hyperparameter Specification

The SBSA method becomes operational once values are assigned to hyperparameters (\mathbf{a}, \mathbf{b}) , (d, r^2) , c^2 and k^2 . The choice of (\mathbf{a}, \mathbf{b}) is highly problem-specific, since this corresponds to belief about the magnitude of measurement error for the near-observed confounders. In some setting the same prior judgement might be made about the measurement error for each confounder, i.e., $(a_j, b_j) = (a, b)$ for $j = 1, \dots, p$. The choice of (d, r^2) corresponds to belief about the magnitude of confounding effects on the outcome variable, and hence is also problem-specific. To some extent though it may be possible to argue for default values. The choice of d governs the concentration of the prior on the common variance of the confounding effects, with a more concentrated prior limiting the ability of the data to inform this variance, and hence discouraging ‘borrowing of strength’ in estimating the effect magnitudes. In the $d \rightarrow \infty$ limit we recover an *iid* prior which precludes such borrowing. Choosing $d > 4$ ensures a finite variance for the inverse-gamma mixing distribution giving rise to the multivariate t distribution. McCandless et al. (2008) use $d = 10$ as a sufficiently small value to encourage borrowing, without incurring computational difficulties which can arise with a heavy-tailed prior. By some quantitative measures, $d = 10$ does not induce a strong dependence between the magnitudes of the components of $\boldsymbol{\beta}$. For instance, consider $\text{Cor}(\beta_r^2, \beta_s^2) = (d - 1)^{-1}$ for $r \neq s$. However, McCandless et al. do demonstrate that the $d = 10$ and $d \rightarrow \infty$ specifications can lead to marked differences via the borrowing of strength in the former case. Given a choice for d , the choice of r simply expresses a plausible range for the magnitudes of confounding

effects. For instance, McCandless et al. set $r = \log(6)/2$ in an example involving a logistic regression outcome model, on the grounds that odds ratios exceeding six are uncommon in most epidemiological contexts. More formally, it might be possible to base specification of r on a review of available literature. For instance, see Viallefont et al. (2001, Sec. 4.2.2) for a distribution of odds-ratios encountered in studies reported in a specific journal and calendar year.

For c^2 and k^2 we do suggest default values that can be argued for in a general manner. Again the thrust of the approach is to treat the unobservable and near-observable confounders as symmetrically as possible. That is, we wish U to be typical both in terms of its range, and in terms of its association with other confounders. Specifically, we aim to (i) equate the prior mean for $Var(U)$ with the (approximate) average prior mean for $Var(Z_j)$, and (ii), equate the (approximate) average prior mean for the squared covariance between U and an element of (X, \mathbf{Z}') with the average squared covariance between distinct elements of (X, \mathbf{Z}') .

Thus we wish to choose (c^2, k^2) to solve

$$\begin{aligned} c^2 + (p + 1)k^2 &= \text{avg}_j \{b_j / (a_j + b_j)\} \\ k^2 \text{avg}_j \{b_j / (a_j + b_j)\} &= \text{avg}_{r \neq s} \{(\Sigma_{rs})^2\}. \end{aligned} \tag{8}$$

Note that there may not be a solution (with $c^2 > 0$) to this system of equations. In particular, sufficiently large off-diagonal elements in Σ with sufficiently large p could produce this problem. In the settings we have considered, however, this has not been an issue.

The role and default settings for hyperparameters can thus be summarized as follows.

<i>Hyperparameter</i>	<i>Function</i>	<i>Default setting</i>
(a_j, b_j)	Prior for magnitude of measurement error on confounder Z_j .	User-specified.
(d, r^2)	Exchangeable prior for confounder coefficients in outcome model.	$d = 10$, r scale-dependent.
c^2	Residual variance for $(U X, \mathbf{Z})$.	Solution to (8),
k^2	Magnitude of prior uncertainty about $(U X, \mathbf{Z})$ regression coefficients.	to treat U and \mathbf{Z} as similarly as possible.

3. Computation

3.1 Continuous Outcome Variable

Our general experience is that computing Bayesian inferences in problems such as these is surprisingly challenging. This is based on experience with the present model, plus the related models used in McCandless et al. (2007; 2008; 2009). The difficulty does not stem from a large number of parameters or datapoints. Rather, the shapes of likelihood functions arising from strongly nonidentified models do not mesh well with ‘off-the-shelf’ MCMC algorithms. To give some insight, we describe two unsuccessful algorithms followed by a successful one. Then we describe extension of the successful algorithm to the case of a binary outcome variable.

In general it is very common to exploit ‘latent structure’ by applying MCMC techniques to the joint posterior distribution of parameters and unobserved latent variables. This seems superficially appealing in the present context, since the full conditional distributions for latent variables U and \mathbf{Z} are normal distributions. Thus our first sampler simply uses Gibbs sampling steps for tractable full conditionals and random-walk Metropolis-Hastings (RWMH) steps for intractable full conditionals. However, this algorithm converges very slowly and mixes very poorly (i.e., the MCMC output is too highly autocorrelated to be useful). This is briefly illustrated later in the example of Section 4.2.

Given that MCMC sampling of latent variables seems ineffective, our second sampler takes advantage of the multivariate normal structure to integrate the unobservables (U, \mathbf{Z}) out of

the likelihood analytically. It is recognized that in general integrating out parameters in the linear normal model can lead to more efficient samplers. Then MCMC can be applied over the parameter space alone, i.e., on $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\tau}^2, \sigma^2)$ which has $3p + 5$ elements. This is done via random-walk Metropolis-Hastings (RWMH) updates, since all closed-form full conditionals are lost upon marginalization. Unfortunately, even though there is a gain in efficiency by avoiding the sampling of unobservables, this algorithm still mixes poorly in some situations. This indicates that reparameterization of the problem should be investigated.

The third sampler, which proves to be quite effective, involves reparameterizing from $\boldsymbol{\theta}$ to $\boldsymbol{\theta}^*$ having components:

$$\begin{aligned}\alpha_x^* &= \alpha_x + \beta_u \gamma_x \\ \boldsymbol{\beta}_z^* &= \boldsymbol{\beta}_z + \beta_u \boldsymbol{\gamma}_z \\ \sigma^{*2} &= \sigma^2 + c^2 \beta_u^2 \\ (\alpha_0^*, \boldsymbol{\tau}^{*2}, \boldsymbol{\gamma}_z^*, \gamma_x^*, \beta_u^*) &= (\alpha_0, \boldsymbol{\tau}^2, \boldsymbol{\gamma}_z, \gamma_x, \beta_u).\end{aligned}$$

To give some insight into this parameterization, note that $(\alpha_0^*, \alpha_x^*, \boldsymbol{\beta}_z^*)$ comprise the regression coefficients for the $(Y|X, \mathbf{Z}) = (Y|X, \mathbf{Z}, \mathbf{W})$ relationship. By integrating with respect to the conditional distribution of $(\mathbf{Z}|\mathbf{W}, X)$ we then obtain the normal linear model describing the observable relationship $(Y|X, \mathbf{W})$ and thereby defining the likelihood function. In particular, this likelihood depends on $\boldsymbol{\theta}^*$ only through $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}_z^*, \boldsymbol{\tau}^{*2}, \sigma^{*2})$, since

$$\begin{aligned}E(Y|X, \mathbf{W}) &= \alpha_0^* + \{\alpha_1^* + \boldsymbol{\beta}_z^{*'} \text{diag}(\boldsymbol{\tau}^{*2}) M^{-1} \boldsymbol{\mu}\} X + \\ &\quad \boldsymbol{\beta}_z^{*'} \{I_p - \text{diag}(\boldsymbol{\tau}^{*2}) M^{-1}\} \mathbf{W}, \\ \text{Var}(Y|X, \mathbf{W}) &= \sigma^{*2} + \boldsymbol{\beta}_z^{*'} \text{diag}(\boldsymbol{\tau}^{*2}) \{I_p - M^{-1} \text{diag}(\boldsymbol{\tau}^{*2})\} \boldsymbol{\beta}_z^*.\end{aligned}$$

In essence we are trying to improve posterior sampling by implementing MCMC in a parameterization involving combinations of parameters which are strongly informed by the

data. This resembles the rationale for ‘hierarchical centering’ in hierarchical models (see, for instance, Gelfand et al. 1996) and also connects with the discussion of MCMC for nonidentified models in Gelfand and Sahu (1999).

We carry out MCMC in the θ^* parameterization by applying RWMH updates to the following six blocks of parameters: (α^*) , (β_z^*) , (τ^{*2}) , (σ^{*2}) , (γ_z^*) , (γ_x^*, β_u^*) . The rationale for blocking γ_x and β_u together is that these two parameters govern the strength of unobserved confounding. Since there is likely little information in the data about these parameters, updating them simultaneously might help to move through the parameter space more quickly. More generally, by blocking this way we hope to make appropriately small jumps for those parameters which are considerably informed by the data (i.e., the first four blocks), and much larger jumps for parameters governed only by the prior (i.e., the last two blocks). We also gain computational efficiency in that likelihood evaluations are not required for the updates to the last two blocks. As a technical point, all updates require evaluation of the prior density for θ^* , whereas the prior is specified in the original θ parameterization. This is readily implemented since the mapping from θ to θ^* has a closed-form inverse, and moreover it is easy to verify that $|\partial\theta^*/\partial\theta| = 1$.

We find that the third algorithm does indeed mix sufficiently well to be relied upon in practice. This suggests reparameterization as a general MCMC strategy for nonidentified models, in such a way that as much as possible parameter components involved in the likelihood function are updated separately from components which are not involved.

3.2 Extension to Binary Outcome Variable

The majority of studies of exposure-disease relationships involve a binary outcome variable, thus it is important to extend SBSA to this case. Consequently we replace the outcome model (5) with

$$\text{logit}Pr(Y = 1|U, \mathbf{Z}, X) = \alpha_0 + \alpha_x X + \beta_u U + \beta_z' \mathbf{Z}.$$

Given the experience with MCMC algorithms in the case of continuous Y , we wish to again marginalize to obtain a likelihood based on $(Y|X, \mathbf{W})$. Unfortunately, now this marginalization must be done numerically. Since

$$\begin{aligned} Pr(Y = 1|X, \mathbf{W}) &= E\{Pr(Y = 1|U, X, \mathbf{Z})|X, \mathbf{W}\} \\ &= E\{\text{expit}(\alpha_0 + \alpha_x X + \beta_u U + \boldsymbol{\beta}'_z \mathbf{Z})|X, \mathbf{W}\}, \end{aligned}$$

we have the opportunity to express the expectation as a one-dimensional integral, by identifying the conditional distribution of a linear combination of (X, U, \mathbf{Z}) given (X, \mathbf{W}) . Straightforward but tedious manipulations with the multivariate normal density give that

$$(\alpha_0 + \alpha_x X + \beta_u U + \boldsymbol{\beta}'_z \mathbf{Z}|X, \mathbf{W}) \sim N(\omega_0 + \omega_x X + \boldsymbol{\omega}'_w \mathbf{W}, \nu^2),$$

where

$$\begin{aligned} \omega_0 &= \alpha_0 \\ \omega_x &= \alpha_x + \beta_u \gamma_x + (\boldsymbol{\gamma}_z \beta_u + \boldsymbol{\beta}_z)' \left[\{M - D(\boldsymbol{\tau}^2)\}^{-1} + D(\boldsymbol{\tau}^2)^{-1} \right]^{-1} \{M - D(\boldsymbol{\tau}^2)\}^{-1} \boldsymbol{\mu} \\ \boldsymbol{\omega}_z &= (\boldsymbol{\gamma}_z \beta_u + \boldsymbol{\beta}_z)' \left[\{M - D(\boldsymbol{\tau}^2)\}^{-1} + D(\boldsymbol{\tau}^2)^{-1} \right]^{-1} D(\boldsymbol{\tau}^2)^{-1} \\ \nu^2 &= \beta_u^2 c^2 + (\boldsymbol{\gamma}_z \beta_u + \boldsymbol{\beta}_z)' \left[\{M - D(\boldsymbol{\tau}^2)\}^{-1} + D(\boldsymbol{\tau}^2)^{-1} \right]^{-1} (\boldsymbol{\gamma}_z \beta_u + \boldsymbol{\beta}_z). \end{aligned}$$

Now, define $q(\delta, \sigma) = E\{\text{expit}(\delta + \sigma S)\}$, where $S \sim N(0, 1)$, as a function which can be evaluated numerically via one-dimensional quadrature. Likelihood evaluations can then be implemented via

$$Pr(Y = 1|X, \mathbf{W}) = q(\omega_0 + \omega_x X + \boldsymbol{\omega}'_w \mathbf{W}, \nu).$$

Since the problem structure is similar to that for continuous Y , we apply MCMC updates using the same reparameterization as before, with the proviso that the original parameter vector $\boldsymbol{\theta}$ no longer includes σ^2 . Commensurately the transformed parameter vector $\boldsymbol{\theta}^*$ no longer includes σ^{*2} . Thus RWMH updates are now applied to five blocks rather than six.

As well, all five updates now involve likelihood evaluations, since we no longer have a clear delineation of which components of θ^* appear, or don't appear, in the likelihood.

4. Empirical Examples

Here we illustrate the application of SBSA to both simulated and real data. This empirical work is complemented by McCandless et al. (2009), in which a version of SBSA for categorical confounders is developed and applied to data from a large cohort study on the relationship between beta-blocker usage and mortality in a population of heart-failure patients. As well, a paper in preparation will give simulation results on how well the binary outcome version of SBSA controls false discoveries in the settings examined by Fewell et al. (2007). Present attention then focusses on (i) a simulation study to highlight the operating characteristics of the continuous-outcome version of SBSA, and (ii) application of this version of SBSA to real data from a small study.

4.1 Simulated Data

Data are simulated as follows. Raw (i.e., not standardized) versions of the exposure and $p+1 = 5$ true confounders are drawn from an equi-correlated multivariate normal distribution with correlation 0.25 and standard marginals. Given these, the observed outcome Y is then generated via regression coefficients of 0.5 for all five (raw) confounders, an effect coefficient of either 0, 0.165, 0.33, or 0.5 for (raw) exposure, and a residual variance of one. The mismeasured confounders \mathbf{W} are obtained by adding measurement errors to the first four true confounders, with ICC values of 0.8, 0.85, 0.9, and 0.95 respectively, and then standardizing. The fifth true confounder is taken as unobserved in the data. The observed exposure X is taken to be the standardized version of the raw exposure. Samples of size $n = 250$, $n = 1000$, and $n = 4000$ are generated in this fashion.

For each dataset, SBSA is applied using hyperparameters $a_j = 4.3$ and $b_j = 30.7$ for

$j = 1, \dots, 4$, giving a prior mode for each ICC_j of 0.9, and high probability (90.8%) that $ICC_j > 0.8$. With this choice of prior there is effectively no truncation issue regarding the prior on τ^2 . (Using the population version of the variance matrix M , a Monte Carlo estimate of the mass assigned outside of set A by the initial prior is zero.) Also, we set $(d, r^2) = (10, 1.6^2)$ as a somewhat informative prior which states that very large confounding effects are unlikely. All other hyperparameters are set at default values. SBSA is implemented via 20,000 MCMC iterations following 1000 burn-in iterations. Some tuning is required to yield mid-range acceptance rates for the RWMH updates to parameter blocks. Since α_x is the regression coefficient for the standardized exposure X in the model for outcome Y , the posterior distribution of α_x is transformed back to the original scale of a coefficient acting on raw exposure, for comparison with the specified true effect size.

For comparison each dataset is also subjected to *ideal* and *naive* analyses. Ideal analysis, a luxury available only with simulated data, involves regressing the outcome on the exposure and the true confounders. Naive analysis, via regression of the outcome on exposure and the surrogates for the first four confounders, is what might typically be done in practice.

For each sample size and exposure effect, 400 datasets are generated. Table 1 indicates the empirical coverage of nominal 95% intervals for each method, where the target of inference is the actual exposure effect. These are standard linear-model confidence intervals for the ideal and naive analyses, and 95% equal-tailed credible intervals for SBSA. As theory dictates, the ideal intervals have empirical coverage within simulation error of nominal. Conversely, the naive intervals undercover very substantially at the smaller sample size, and catastrophically at the larger sample sizes. This matches the findings of Fewell et al. (2007) concerning the impact of unchecked poor measurement and omission of confounders. It is also a reminder that these problems worsen as we collect larger samples and are drawn to being ever more

precisely wrong. On the other hand, the SBSA intervals are highly conservative, with 100% coverage in all cases. In this regard SBSA appears to work as a sensitivity analysis tool.

Of course we expect to pay for the high coverage of SBSA intervals via their width. This is evidenced in Table 2. At the smallest sample size of $n = 250$, the SBSA intervals are roughly three times wider than their ideal and naive counterparts. Moreover, the intervals narrow only slightly upon quadrupling to $n = 1000$, and then exhibit essentially no further narrowing upon quadrupling again to $n = 4000$. Conversely, the ideal and naive intervals are governed by regular asymptotics, tending to shrink by a factor of two when the sample size quadruples. With SBSA, by $n = 1000$ the uncertainty driven by sampling variability is already dwarfed by ‘scenario uncertainty’ about the extent of measurement error and unobserved confounding, and it is not cost-effective to collect further data to reduce sampling variability further. Such sample-size considerations around nonidentified models and study limitations are addressed in some generality by Gustafson (2006).

Finally, Table 3 reports the proportion of intervals for the exposure-disease relation which exclude zero. We see the nice behaviour expected of perfect data, as the ideal intervals have nominal coverage when the true effect is zero, but quickly become powerful as the true effect deviates from zero. Conversely, the results for the naive analysis underscore the huge potential for type I errors to occur as a result of poorly measured and missing confounders. For SBSA we do see an increased chance of correctly concluding there is an exposure-disease relationship, both as the true effect size increases, and as the sample size increases. With enough data on strong exposure-disease relationships it is possible to express confidence in the existence of the relationship, notwithstanding the limited data on confounders. For a given magnitude of effect and sample size, however, the power to detect the effect via SBSA applied to limited data is much reduced compared to the power arising from full data.

To add breadth to the simulation findings, the supplementary web material includes

commensurate results for two alternate scenarios. The first alternate scenario is as described above, but with the confounder coefficients in the outcome model for Y reduced from 0.5 to 0.25. Thus this scenario is characterized as involving weaker confounders. The behaviour of the methods in this scenario is very similar to that in the main scenario. The second alternate scenario is as per the main scenario, but with the correlations between confounders increased from 0.25 to 0.85 (while the exposure-confounder correlations remain at 0.25). Unobserved confounding should be less damaging when the unobserved confounder is more highly correlated with the near-observed confounders, since adjustment for the latter will better deal with the former. Indeed, we see lower Type I error rates for the naive analysis than in the other scenarios. However, this comes at the cost of lost power, as the SBSA intervals are now too wide to detect an exposure-disease relationship, even for the largest effect size and sample size considered.

4.2 *Illustrative data analysis*

To illustrate SBSA we consider data from a study of bone mineral density reported on by Hopper and Seeman (1994), and also considered in Rosner (2000). While the original study focussed on the relationship between bone mineral density and tobacco use, for illustrative purposes we consider the relationship between bone mineral density and height, considering tobacco use and other variables as possible confounders for this relationship.

The study involved $n = 41$ pairs of female twins, hence all variables are taken to be within-pair differences. The outcome Y is the standardized difference in bone density at the femoral neck. The exposure X is the standardized difference in height. The $p = 6$ potential confounders \mathbf{W} are standardized differences in cumulative (pack-years) and current (cigarettes per day) tobacco consumption, weight, alcohol consumption, coffee consumption, and tea consumption (with the last three initially recorded as a per-week value). Note that

in this framework X is likely measured with little error, while the elements of \mathbf{W} may indeed be measured with considerable error.

The naive analysis of regressing Y on (X, \mathbf{W}) gives an estimated X coefficient of 0.39, with a standard error of 0.18. Thus there is evidence for a positive association between Y and X given \mathbf{W} . Given the standardization of variables, the estimated coefficient can be interpreted as one SD upward change in height being associated with a 0.39 SD upward change in bone mineral density. It is worth noting that none of the estimated \mathbf{W} coefficients are large or significant. Of the six, cumulative tobacco consumption has both the largest magnitude of estimated coefficient (-0.18), and the largest z-score (1.1).

SBSA is implemented using the same hyperparameter choices as earlier in Section 4.1. The posterior sampling uses 200,000 iterations after 10,000 burn-in iterations; traceplots for some key parameters appear in Figure 1. To underscore a key point made in Section 3.1, this figure also shows output from the first sampler, based on Gibbs sampling updates to the latent variables, which is extremely ineffective.

The point estimate of the exposure-disease coefficient (posterior mean of α_x) is 0.40, which is virtually the same as the naive estimate. The posterior standard deviation of 0.32, however, is much larger than the standard error of the naive estimate. That is, incorporating uncertainty concerning measurement error in confounders and a potential unobserved confounder leads to almost twice as much uncertainty about the parameter of interest. The upper-left panel of Figure 2 gives both the point estimates and 95% equal-tailed interval estimates for the two analyses. The SBSA interval crosses zero, so the ‘significance’ seen in the naive analysis is lost upon incorporation of the further uncertainty.

We gain some further insight by looking at *a posteriori* dependencies between the target parameter α_x and other parameters (Figure 2). Surprisingly, there seems to be virtually no dependence between α_x and $\sum_{i=1}^p \tau_i^2$ (which represents the overall extent of confounder

measurement error). This suggests that, for these data, the extra width in the SBSA interval compared to the naive interval is driven by potential unobserved confounding, rather than by the mismeasured confounding. This seems plausible given the lack of evidence for any strong confounding effects of \mathbf{W} in the naive analysis. The clear posterior dependence seen between α_x and β_u , and between α_x and γ_x , also bears this out, i.e., variation in (β_u, γ_x) is seen to drive variation in α_x . Moreover, the posterior distribution of $(\alpha_x | \beta_u = 0)$ closely matches the naive inference, and this is also true for the posterior of $(\alpha_x | \gamma_x = 0)$. Thus inference which acknowledges measurement error but conditions on U *not* being a confounder is very close to the inference which ignores *both* measurement error and unobserved confounding.

5. Discussion

We have demonstrated the feasibility of a Bayesian approach to simultaneously acknowledging confounding variables which are not measured well and confounding variables which are not measured at all. While we have focussed on a very specific context, with precisely measured continuous exposure and near-observed and unobserved continuous confounders, the style of approach to the problem is generic. The two key features are (i) treating confounders exchangeably without regard to which are observed and unobserved, and (ii) trying to reduce prior specifications to a few key elements.

Extensions of these key features to more complex settings will vary in terms of technical and computational challenge. The related work on unobserved (but not poorly measured) binary confounders in McCandless et al. (2009) is one example in this regard. Another extension would be to Berkson measurement error. In this case the factorization (1) would not be natural, but obtaining $f_\theta(Y | \mathbf{W}, X)$ by marginalizing

$$f_\theta(Y, U, \mathbf{Z} | \mathbf{W}, X) = f_\theta(Y | U, \mathbf{Z}, \mathbf{W}, X) f_\theta(U | \mathbf{Z}, \mathbf{W}, X) f_\theta(\mathbf{Z} | \mathbf{W}, X)$$

would be more suitable. Some terms would simplify, i.e., \mathbf{W} would actually be assumed to

be absent from the first two terms, which could then be treated as in Section 2. Furthermore, under a nondifferential error assumption X could be assumed absent from the third term comprising the measurement error model.

There is a general challenge in moving beyond situations where normal linear models can be applied, particularly given the simplicity gained here by taking the variance of (X, \mathbf{W}) as known and exploiting the implications of this arising from normality and linearity assumptions. One approach would be to further explore the use of loglinear models as in (McCandless et al., 2009), while another would be to consider probit style models where categorical variables are modelled as arising from latent variables following normal linear models.

Acknowledgements

PG acknowledges support from the Natural Sciences and Engineering Research Council of Canada, and from the Canadian Institutes of Health Research (Funding Reference Number 62863). AL acknowledges salary support provided by the Michael Smith Foundation for Health Research. SR acknowledges support from the ESRC National Centre for Research Methods (BIAS II node, grant number RES-576-25-0015).

References

- Brunner, J. and Austin, P. (2009). Inflation of type I error rate in multiple regression when independent variables are measured with error. *Canadian Journal of Statistics* **37**, 33–46.
- Cornfield, J., Haenszel, W., Hammond, E., Lilienfeld, A., Shimkin, M., and Wynder, E. (1959). Smoking and lung cancer: recent evidence and a discussion of some questions. *Journal of the National Cancer Institute* **22**, 173–203.
- Fewell, Z., Davey Smith, G., and Sterne, J. (2007). The impact of residual and unmea-

- sured confounding in epidemiologic studies: a simulation study. *American Journal of Epidemiology* **166**, 646–655.
- Gelfand, A. and Sahu, S. (1999). Identifiability, improper priors, and Gibbs sampling for generalized linear models. *Journal of the American Statistical Association* **94**, 247–253.
- Gelfand, A., Sahu, S., and Carlin, B. (1996). Efficient parametrizations for generalized linear mixed models. In *Bayesian statistics 5: Proceedings of the Fifth Valencia Conference*, pages 165–180. Oxford.
- Greenland, S. (1996). Basic methods for sensitivity analysis of biases. *International Journal of Epidemiology* **25**, 1107–1116.
- Greenland, S. (2003). 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* **98**, 47–55.
- Greenland, S. (2005). Multiple-bias modelling for analysis of observational data. *Journal of the Royal Statistical Society, Series A* **168**, 267–306.
- Gustafson, P. (2006). Sample size implications when biases are modelled rather than ignored. *Journal of the Royal Statistical Society, Series A* **169**, 883–902.
- Hernán, M. and Robins, J. (1999). Letter to editor. *Biometrics* **55**, 1316–1317.
- Hopper, J. and Seeman, E. (1994). The bone density of female twins discordant for tobacco use. *New England Journal of Medicine* **330**, 3897–392.
- Lin, D., Psaty, B., and Kronmal, R. (1998). Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics* **54**, 948–963.
- McCandless, L., Gustafson, P., and Levy, A. (2007). Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Statistics in Medicine* **26**, 2331–2347.
- McCandless, L., Gustafson, P., and Levy, A. (2008). A sensitivity analysis using information about measured confounders yielded improved assessments of uncertainty from unmea-

- sured confounding. *Journal of Clinical Epidemiology* **61**, 247–255.
- McCandless, L., Gustafson, P., Levy, A., and Richardson, S. (2009). Hierarchical priors for bias parameters in Bayesian adjustment for unmeasured confounding. *Submitted*.
- Rosenbaum, P. (2002). *Observational Studies*. Springer.
- Rosner, B. (2000). *Fundamentals of Biostatistics, 5th ed.* Duxbury Press.
- Rothman, K. (1976). Causes. *American Journal of Epidemiology* **104**, 587–592.
- VanderWeele, T. (2008). Sensitivity analysis: distributional assumptions and confounding assumptions. *Biometrics* **64**, 645–649.
- Viallefont, V., Raftery, A., and Richardson, S. (2001). Variable selection and Bayesian model averaging in case-control studies. *Statistics in Medicine* **20**, 3215–3230.
- Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior Distributions. In *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, pages 333–243. North-Holland/Elsevier.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

[Figure 1 about here.]

[Figure 2 about here.]

[Table 4 about here.]

[Table 5 about here.]

[Table 6 about here.]

[Table 7 about here.]

[Table 8 about here.]

[Table 9 about here.]

Received February 2009. Revised August 2009.

Accepted

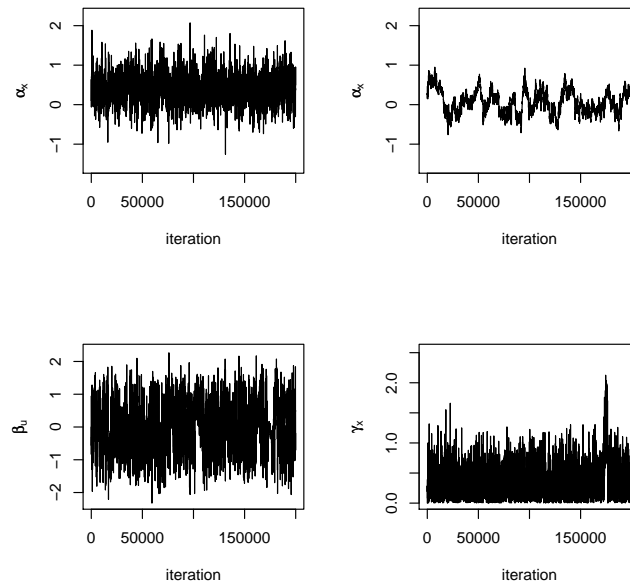


Figure 1. MCMC traceplots for α_x , β_u and γ_x in the bone mineral density example. These arise from the third sampler described in Section 3, except for the α_x traceplot in the upper-right panel. This arises from the first sampler which uses Gibbs sampling updates to the latent variables.

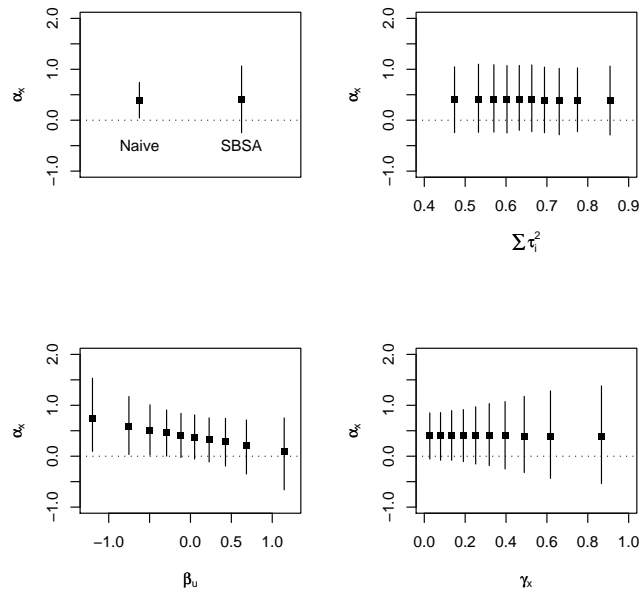


Figure 2. Posterior summaries for the bone mineral density example. The upper left panel contrasts naive and SBSA 95% interval estimates for the target parameter α_x . The remaining panels give posterior intervals conditioned on *a posteriori* deciles of (i) $\sum_{i=1}^p \tau_i^2$, (ii) β_u , (iii) γ_x .

Table 1

Coverage (%) of nominal 95% intervals based on 400 simulated datasets. Note that with this many realizations the empirical coverage of a true 95% interval has simulation SE of 1.1%

exposure effect	ideal				naive				SBSA			
	0.00	0.17	0.33	0.50	0.00	0.17	0.33	0.50	0.00	0.17	0.33	0.50
$n = 250$	95	94	94	94	71	66	70	73	100	100	100	100
$n = 1000$	94	96	92	95	24	20	21	18	100	100	100	100
$n = 4000$	92	94	93	95	0	0	0	0	100	100	100	100

Table 2
Average length of nominal 95% intervals

exposure effect	ideal				naive				SBSA			
	0.00	0.17	0.33	0.50	0.00	0.17	0.33	0.50	0.00	0.17	0.33	0.50
$n = 250$	0.27	0.27	0.27	0.28	0.32	0.32	0.32	0.32	0.91	0.91	0.91	0.91
$n = 1000$	0.14	0.14	0.14	0.14	0.16	0.16	0.16	0.16	0.84	0.84	0.84	0.83
$n = 4000$	0.07	0.07	0.07	0.07	0.08	0.08	0.08	0.08	0.82	0.83	0.82	0.82

Table 3
Percentage of nominal 95% intervals excluding zero

exposure effect	ideal					naive					SBSA		
	0.00	0.17	0.33	0.50	0	0.17	0.33	0.50	0	0.17	0.33	0.50	
$n = 250$	5	68	100	100	29	93	100	100	0	2	40	86	
$n = 1000$	6	100	100	100	76	100	100	100	0	0	49	100	
$n = 4000$	8	100	100	100	100	100	100	100	0	0	56	100	

Table 4
SUPPLEMENTARY MATERIAL - WEB ONLY!!! Coverage (%) of nominal 95% intervals in the weaker confounder setting. The format is as per Table 1.

exposure effect	ideal				naive				SBSA			
	0.00	0.17	0.33	0.50	0.00	0.17	0.33	0.50	0.00	0.17	0.33	0.50
$n = 250$	95	94	94	95	86	86	87	88	100	100	100	100
$n = 1000$	95	96	96	96	65	64	69	65	100	100	100	100
$n = 4000$	94	94	94	95	10	10	13	14	100	100	100	100

Table 5
SUPPLEMENTARY MATERIAL - WEB ONLY!!! Average length of nominal 95% intervals in the weaker confounder setting. The format is as per Table 2.

exposure effect	ideal				naive				SBSA			
	0.00	0.17	0.33	0.50	0.00	0.17	0.33	0.50	0.00	0.17	0.33	0.50
$n = 250$	0.27	0.27	0.28	0.27	0.28	0.28	0.28	0.28	0.85	0.85	0.87	0.85
$n = 1000$	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.79	0.79	0.79	0.78
$n = 4000$	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.77	0.78	0.77	0.78

Table 6
SUPPLEMENTARY MATERIAL - WEB ONLY!!! Percentage of nominal 95% intervals excluding zero in the weaker confounding setting. The format is as per Table 3.

exposure effect	ideal				naive				SBSA			
	0.00	0.17	0.33	0.50	0	0.17	0.33	0.50	0	0.17	0.33	0.50
$n = 250$	5	63	100	100	14	87	100	100	0	2	28	87
$n = 1000$	5	100	100	100	35	100	100	100	0	0	32	100
$n = 4000$	6	100	100	100	90	100	100	100	0	0	35	100

Table 7
SUPPLEMENTARY MATERIAL - WEB ONLY!!! Coverage (%) of nominal 95% intervals in the more correlated confounder setting. The format is as per Table 1.

exposure effect	ideal				naive				SBSA			
	0.00	0.17	0.33	0.50	0.00	0.17	0.33	0.50	0.00	0.17	0.33	0.50
$n = 250$	94	94	96	94	91	93	95	96	100	100	100	100
$n = 1000$	95	95	94	95	89	88	89	87	100	100	100	100
$n = 4000$	95	93	94	95	65	70	65	66	100	100	100	100

Table 8
SUPPLEMENTARY MATERIAL - WEB ONLY!!! Average length of nominal 95% intervals in the more correlated confounder setting. The format is as per Table 2.

exposure effect	ideal				naive				SBSA			
	0.00	0.17	0.33	0.50	0.00	0.17	0.33	0.50	0.00	0.17	0.33	0.50
$n = 250$	0.26	0.26	0.26	0.26	0.29	0.29	0.29	0.29	1.75	1.77	1.78	1.75
$n = 1000$	0.13	0.13	0.13	0.13	0.14	0.14	0.14	0.14	1.68	1.68	1.68	1.67
$n = 4000$	0.06	0.06	0.06	0.06	0.07	0.07	0.07	0.07	1.65	1.65	1.65	1.63

Table 9
SUPPLEMENTARY MATERIAL - WEB ONLY!!! Percentage of nominal 95% intervals excluding zero in the more correlated confounding setting. The format is as per Table 3.

exposure effect	ideal					naive					SBSA		
	0.00	0.17	0.33	0.50	0	0.17	0.33	0.50	0	0.17	0.33	0.50	
$n = 250$	6	70	100	100	9	76	100	100	0	0	0	1	
$n = 1000$	5	100	100	100	11	100	100	100	0	0	0	1	
$n = 4000$	5	100	100	100	35	100	100	100	0	0	0	0	