1-15-2008

A BAYESIAN APPROACH TO EFFECT ESTIMATION ACCOUNTING FOR ADJUSTMENT UNCERTAINTY

Chi Wang
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, chwang@jhsph.edu

Giovanni Parmigiani
The Sydney Kimmel Comprehensive Cancer Center, Johns Hopkins University & Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

Ciprian Crainiceanu
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics

Francesca Dominici
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics

Submitted to the Annals of Applied Statistics

A BAYESIAN APPROACH TO EFFECT ESTIMATION ACCOUNTING FOR ADJUSTMENT UNCERTAINTY

By Chi Wang, Giovanni Parmigiani*, Ciprian Crainiceanu and Francesca Dominici†

Johns Hopkins University

Adjustment for confounding factors is a common goal in the analysis of both observational and controlled studies. The choice of which confounding factors should be included in the model used to estimate an effect of interest is both critical and uncertain. For this reason it is important to develop methods that estimate an effect, while accounting not only for confounders, but also for the uncertainty about which confounders should be included. In a recent article, Crainiceanu et al. (2008) have identified limitations and potential biases of Bayesian Model Averaging (BMA) (Raftery et al., 1997; Hoeting et al., 1999) when applied to adjustment uncertainty, that arise because BMA weights models by their ability to make predictions and this may not reflect the models' ability to correctly adjust for confounding.

An important remaining question is whether it is possible to design approaches that account for adjustment uncertainty by treating the selection of variables as an unknown parameter, as BMA does, but do not suffer from the same limitations. In this paper, we propose a novel Bayesian formulation, called "Bayesian Confounding Adjustment" (BCA) to account for adjustment uncertainty in effect estimation from a Bayesian perspective. BCA uses a different weighting mechanism than BMA, wherein effect estimation is obtained by weighting effect estimates from models, all of which attempt to be fully adjusted for confounding. In simulation studies we show that BCA provides estimates of the exposure effect that have lower mean squared error than BMA and correct coverage. We then compare BCA, the approach of Crainiceanu et al. (2008), and traditional BMA in a time series data set of hospital admissions, air pollution levels and weather variables in Nassau, NY for the period 1999-2005. Using each approach, we estimated the short-term effects of $PM_{2.5}$ on emergency admissions for cardiovascular diseases, accounting for confounding. This application illustrates the potentially significant pitfalls of misusing variable selection methods in the context of adjustment uncertainty.

*Supported in part by NCI grant 2P30CA006973-44S4

†Supported in part by the National Institute for Environmental Health Sciences (ES012054-03)

Keywords and phrases: Adjustment uncertainty, Bayesian model averaging, Treatment effect, Exposure effect

1. Introduction. Estimating the effect of a predictor on a response, while properly adjusting for confounding factors, is a common goal in biomedical research. A prominent and controversial example arises in observational studies of the health effects of environmental contaminants, where the choice of potential confounders is challenging, and major policy decisions can depend on it. The most common practice is currently to select a model for the estimation of the effect, and report effect estimates and confidence intervals that are conditional on that model being correct. This does not account for "adjustment uncertainty", that is uncertainty about which covariates should be included in the model to properly adjust for confounding. It is sometimes possible to effectively convey this uncertainty by sensitivity analyses, showing the variation of the effect estimate and its interval over a range of plausible choices of confounders (Dominici et al., 2004; Peng et al., 2006). For example, when effect estimates are stable over a range of plausible choice, a sensitivity analysis can be sufficient. However, when the effect of inclusions or exclusion of potential confounder is stronger it is important to develop combined estimates that compromise between these choices and properly report the associated uncertainty.

Bayesian Model Averaging (BMA) has been suggested as a formal tool to achieve these goals and account for adjustment uncertainty in effect estimation. Bayesian predictions that account for uncertainty in the selection of predictors (or confounders) (Raftery et al., 1997; Hoeting et al., 1999), are based on treating the vector of indicators of whether each predictor is included as an additional parameter in the analysis. For prediction purposes, this parameter is a nuisance parameter that can be integrated out. This results in a weighted average of predictions whose weights depend on the support that a particular selection receives from the data. This principled approach inherits a number of desirable properties from a frequentist point of view as well, and has performed competitively in out-of-sample prediction comparisons from the artificial intelligence literature (Chipman et al., 2002; Yeung et al., 2005). The conceptual simplicity and solid logic behind treating the unknown confounder subset as a parameter is attractive in adjustment uncertainty as well: for example, Raftery (1995) suggests to estimate the exposure effect by a weighted average of model-specific coefficients, again using the model's posterior probabilities as weights. Examples include applications to air pollution research (Clyde, 2000; Koop and Tole, 2004).

More recently, Crainiceanu et al. (2008) have made the case that it can be dangerous to use approaches originally designed for prediction and mechanically translate their use to effect estimation with adjustment uncertainty. In adjustment uncertainty, the goal of the modeling is to minimize

the MSE of the exposure effect estimate (Dominici et al., 2004; Peng et al., 2006; Crainiceanu et al., 2008). Crainiceanu et al. (2008) introduced a new approach to estimate an exposure effect accounting for adjustment uncertainty to ultimately obtain an estimate of the effect with desirable inferential properties. This approach (denoted by the authors' initials CDP) can be described in two steps. In the first step, CDP regresses exposure on a large set of potential confounders and selects confounders that are strongly associated with exposure. In the second step, CDP regresses outcome on exposure, after including as covariates the confounders identified at the first step. In addition, Crainiceanu et al. (2008) examined the strength and limitation of BMA for estimation and pointed out that BMA is designed for minimizing prediction error but that it might not be appropriate for effect estimation. This happens because the posterior model probabilities used to weight the model-specific estimates of the exposure effect might not reflect the ability of the model to provide an estimate of the exposure effect properly adjusted for confounding. For example, it can occur that large weights are assigned to models that do not adequately adjust for confounders, leading to a biased estimate of the exposure effect. In simulations that are realistic for air pollution research, BMA can be severely biased while the CDP method produces an estimate of the exposure effect that is unbiased and has smaller MSE than BMA.

Here our goal is to develop a Bayesian approach to handle adjustment uncertainty by considering the selection of confounders as a random variable, as in model averaging, while overcoming the pitfalls of BMA identified by Crainiceanu et al. (2008). To this end we introduce a novel BMA approach tailored to estimation of exposure effects accounting for adjustment uncertainty. Our approach estimates the exposure effect by a weighting strategy constructed to assigns high weight to models that are likely to include all the necessary confounders. Specifically, first, we build an exposure model having X as the dependent variable and a large set of potential confounders as independent variables. We introduce model selection parameters $\boldsymbol{\alpha}^X = (\alpha_1^X, ..., \alpha_M^X)^T \in \{0,1\}^M$ such that $\alpha_m^X = 1$ when the mth potential confounder is included. Second, we build an outcome model having Y as the dependent variable, X as independent variable, and also introduce model selection parameters $\boldsymbol{\alpha}^Y$. A key assumption in our approach is that, conditional on the selection indicators $\boldsymbol{\alpha}^X$, any predictors of X for which $\alpha_m^X = 1$ is automatically included in the outcome model. Using this approach we develop a new set of model-specific probabilities $P(\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y | Data)$ that are likely to assign high weights to outcome models that are properly adjusting for confounding.

The paper is organized as follows. In section 2, we introduce the statistical framework and illustrate the different implications of uncertainty in variable selection for effect estimation and prediction. In section 3, we introduce our newly proposed Bayesian methodology to account for adjustment uncertainty in effect estimation. In section 4, we present simulation studies to rigorously compare our approach with the CDP and BMA approaches. In section 5 we apply our methods to time series data in Nassau, NY to investigate the effect of $PM_{2.5}$ on the hospitalization rate of cardiovascular disease. In section 6, we discuss the strength and weakness of the methods and future work.

2. Concepts and Notation. Consider building a model for estimating the effect of exposure X on outcome Y. We assume that a set of L covariates $\boldsymbol{Z} = \{Z_1, \ldots, Z_L\}$ are always included in the model. We also consider M potential confounders $\boldsymbol{U} = \{U_1, \ldots, U_M\}$ selected a priori so that they are likely to affect the response Y, although the effect could potentially be weak. These are either included or excluded from the model, depending on a vector of indicators $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)^T \in \{0,1\}^M$. Here $\alpha_m = 1$ whenever U_m is included in the model. Covariates that are highly associated with X but not with Y are not potential confounder and ideally should not be included in the candidate set $\boldsymbol{U}$. Consider the following linear regression model:

$$E\{y_i\} = \beta^{\boldsymbol{\alpha}} x_i + \sum_{l=1}^{L} \gamma_l^{\boldsymbol{\alpha}} z_{il} + \sum_{m=1}^{M} \alpha_m \delta_m^{\boldsymbol{\alpha}} u_{im}, \tag{2.1}$$

where x_i and y_i are the exposure and outcome levels at observation i. The unknown parameters are the binary vector of unknown model selection parameters $\boldsymbol{\alpha}$, the effect of interest $\beta^{\boldsymbol{\alpha}}$, the additional coefficients $\gamma_l^{\boldsymbol{\alpha}}$ and $\delta_m^{\boldsymbol{\alpha}}$. The intercept term can be included among the covariates in $\boldsymbol{Z}$. In Equation (2.1), and throughout, we use a notation that explicitly keeps track of the fact that regression coefficients differ in meaning as we change $\boldsymbol{\alpha}$. This is especially important when one attempts to make inferences that involve estimates of the exposure effect obtained using different models.

When studying confounding adjustment it is useful to consider the smallest model that includes all the necessary confounders. We denote this by $\boldsymbol{\alpha}^*$, and refer to it as the minimal model. The true effect of X on Y is the coefficient of X in this model. We denote it by $\beta^* = \beta^{\boldsymbol{\alpha}^*}$. A key observation here is that all models that contain at least as many predictors as the minimal model will provide estimates of the exposure effect that are also interpretable as estimate of β^*, while models that do not, as for example

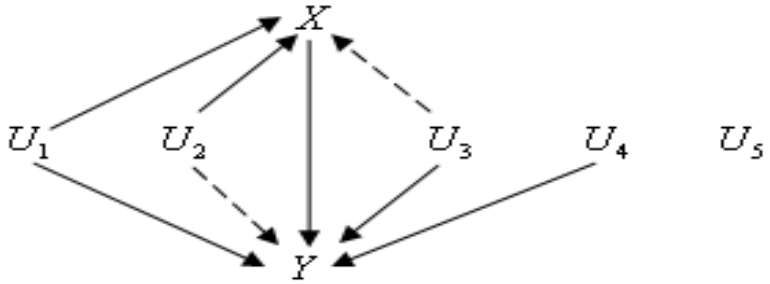

FIG 1. *An illustrative example. Solid lines indicate strong correlations while dashed lines indicate weak correlations.*

models that do not include a key confounder, simply estimate a parameter having a different interpretation. Formally, we say $\boldsymbol{\alpha} \subseteq \boldsymbol{\alpha}'$ if the model $\boldsymbol{\alpha}$ is nested within model $\boldsymbol{\alpha}'$. Then all models nesting $\boldsymbol{\alpha}^*$ estimate the same true effect, so $\beta^{\boldsymbol{\alpha}} = \beta^*$ whenever $\boldsymbol{\alpha}^* \subseteq \boldsymbol{\alpha}$. In this setting, our goal is the estimation of β^* when $\boldsymbol{\alpha}^*$ is unknown.

Example 1: Before we describe the methodology, it is useful to frame the issues using a simple example. Consider the situation depicted in Figure 1, where $\boldsymbol{U} = \{U_1, U_2, U_3, U_4, U_5\}$ and $\boldsymbol{Z} = \emptyset$. The U's are related as represented in Figure 1: U_1 is highly correlated with both exposure and outcome (solid arrows); U_2 is highly correlated with the exposure, but weakly correlated with the outcome; U_3 is highly correlated with outcome and weakly correlated with exposure; U_4 is highly correlated with outcome and uncorrelated with exposure; and lastly U_5 is uncorrelated with both exposure and outcome. In this scenario, U_1,U_2, and U_3 are potential confounders of the association between X and Y and must be included into the regression model as covariates. The minimal model that can provide a correctly adjusted effect of X on Y is $\alpha_1^* = \alpha_2^* = \alpha_3^* = 1$, $\alpha_4^* = \alpha_5^* = 0$. The true model is $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 1$, $\alpha_5 = 0$: this includes the minimal model and also allows to estimate the correct exposure, potentially with greater accuracy than the minimal model. However neither the definition of β^* nor the success of our methodology depend on whether the true model is included in the set of models considered. The full model ($\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 1$) also contains $\boldsymbol{\alpha}^*$ and a correctly defined coefficient. On the other hand, models that do not nest $\boldsymbol{\alpha}^*$ will estimate parameters that are not properly adjusted by confounding. For example, the model $\boldsymbol{\alpha}'(\alpha_1' = \alpha_3' = \alpha_4' = 1, \alpha_2' = \alpha_5' = 0)$ will not provide a proper estimate for the exposure effect. However, it may still be a useful model in terms of model fitting and prediction.

To further illustrate this point, we construct a simulated data set where

the covariates satisfy the relationships in Figure 1, that is:

$$\begin{aligned} X_i &= \delta_1^X U_{1i} + \delta_2^X U_{2i} + \delta_3^X U_{3i} + \epsilon_i^X \\ Y_i &= \beta X_i + \delta_1^Y U_{1i} + \delta_2^Y U_{2i} + \delta_3^Y U_{3i} + \delta_4^Y U_{4i} + \epsilon_i^Y, \end{aligned} \tag{2.2}$$

where $i = 1, ..., 1000$ and $\epsilon_i^X, \epsilon_i^Y$ independently follow $N(0, \sigma_X^2)$ and $N(0, \sigma_Y^2)$ respectively. The variables U_{ji}s are independently distributed as $N(0, \sigma_U^2)$. In our simulation, we set $\delta_1^X = \delta_2^X = 1, \delta_3^X = 0.1, \delta_1^Y = \delta_3^Y = \delta_4^Y = 1, \delta_2^Y = 0.1$, $\beta = 0.1$, $\sigma_X = \sigma_Y = \sigma_U = 1$. The correlation matrix is:

$$\begin{pmatrix} & X & U_1 & U_2 & U_3 & U_4 & U_5 & Y \\ X & 1.00 & 0.57 & 0.58 & 0.04 & 0.01 & -0.01 & 0.41 \\ U_1 & 0.57 & 1.00 & 0.00 & -0.06 & 0.03 & -0.03 & 0.51 \\ U_2 & 0.58 & 0.00 & 1.00 & -0.02 & 0.01 & 0.04 & 0.09 \\ U_3 & 0.04 & -0.06 & -0.02 & 1.00 & 0.02 & -0.03 & 0.48 \\ U_4 & 0.01 & 0.03 & 0.01 & 0.02 & 1.00 & -0.01 & 0.50 \\ U_5 & -0.01 & -0.03 & 0.04 & -0.03 & -0.01 & 1.00 & -0.02 \\ Y & 0.41 & 0.51 & 0.09 & 0.48 & 0.50 & -0.02 & 1.00 \end{pmatrix},$$

which reflects the correlation structure in Figure 1. Using this data set, we estimate β using maximum likelihood estimation under two models: one is the true model for Y in (2.2), and the other is:

$$Y_i = \beta X_i + \delta_1^Y U_{1i} + \delta_3^Y U_{3i} + \delta_4^Y U_{4i} + \epsilon_i^Y, \tag{2.3}$$

which, unlike (2.2) does not include the confounder U_2. The results are shown in Table 1 and in Table 2. The values of AIC (Akaike, 1973) for the true model described in Equation (2.2) and for the model described in Equation (2.3) are very similar indicating that these two models fit the data equally well. We also report the Bayesian Information Criterion (BIC) (Schwarz, 1978) for comparison. The likelihood ratio test for the difference between the model in Equation (2.2) and the model in Equation (2.3) is not significant (p-value=0.087). However, the two models provide widely different estimate of the exposure coefficient. The two models are estimating exposure effects with different interpretation and only the larger model in (2.2) provides a properly adjusted effect.

This simple example illustrates that effective model selection approaches for adjustment uncertainty in effect estimation are not necessarily the same as model selection approaches whose goal is prediction of the response. In the former, models are valuable to the extent that they estimate correctly a single parameter of interest. In the latter, models are valuable to the extent they predict the reponse well —which can often be achieved even by models that provide systematically biased estimate of the exposure effect β.

	model (2.2)	model (2.3)
$\hat{\beta}$	0.121 (0.059, 0.183)	0.160 (0.116, 0.203)
SE($\hat{\beta}$)	0.032	0.022
AIC	2847.874	2848.803
BIC	2882.228	2878.249

TABLE 1
Comparison of model fitting and the estimation of β from model (2.2) and model (2.3). The true value is 0.1.

3. Methods.

3.1. *Bayesian Model Averaging (BMA).* In the context of effect estimation, Raftery (1995); Hoeting et al. (1999) suggests to calculate the posterior distribution of β by taking average over all models weighted by their posterior probabilities:

$$P(\beta|D) = \sum_{\boldsymbol{\alpha}} P(\beta, \boldsymbol{\alpha}|D) = \sum_{\boldsymbol{\alpha}} P(\beta|\boldsymbol{\alpha}, D)P(\boldsymbol{\alpha}|D), \tag{3.1}$$

where D denotes the observed data. We will refer to this approach as BMA.

In practice, this approach can be interpreted and implemented in two ways: one is to force exposure in the model (denoted by FBMA) and only take summation over models that always include the exposure covariate. The other is not to force exposure in the model (denoted by NBMA). For a model that does not include exposure, the distribution of β is simply a point mass at zero.

The approach described in Equation (3.1) is intuitive because of its analogy to the BMA approach for adjusting for model uncertainty in prediction. However, adjustment uncertainty and model uncertainty are different and we need to investigate this approach more carefully. Assume the minimal model is $\boldsymbol{\alpha}^*$, then Equation (3.1) can be decomposed into two parts:

$$P(\beta|D) = \sum_{\boldsymbol{\alpha} \supseteq \boldsymbol{\alpha}^*} P(\beta|\boldsymbol{\alpha}, D)P(\boldsymbol{\alpha}|D) + \sum_{\boldsymbol{\alpha} \not\supseteq \boldsymbol{\alpha}^*} P(\beta|\boldsymbol{\alpha}, D)P(\boldsymbol{\alpha}|D). \tag{3.2}$$

The second term of Equation (3.2) averages $P(\beta|\boldsymbol{\alpha}, D)$ across models $\boldsymbol{\alpha}$ that do not include $\boldsymbol{\alpha}^*$. This leads to assigning potentially large weights to models estimating β's that are different from β^*. The ratio of the weights given to models $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ is

$$\frac{P(\boldsymbol{\alpha}_1|D)}{P(\boldsymbol{\alpha}_2|D)} = \frac{P(D|\boldsymbol{\alpha}_1)}{P(D|\boldsymbol{\alpha}_2)} \frac{P(\boldsymbol{\alpha}_1)}{P(\boldsymbol{\alpha}_2)}.$$

model	$\hat{\beta}$	$SE(\hat{\beta})$	AIC	BIC	FBMA weight	BCA weight
(1,1,1,1,0) (model (2.2))	0.121	0.032	2847.874	2882.228	0.0602	0.9852
(1,0,1,1,0) (model (2.3))	0.160	0.022	2848.803	2878.249	0.9269	0.0000
(1,1,1,1,1)	0.122	0.032	2849.572	2888.834	0.0013	0.0148
(1,0,1,1,1)	0.160	0.022	2850.416	2884.771	0.0117	0.0000
(1,1,1,0,0)	0.096	0.044	3515.806	3545.253	0.0000	0.0000

TABLE 2
Comparison of model posterior from FBMA and BCA, estimate from FBMA is 0.158 with standard error 0.025, that from BCA is 0.122 with standard error 0.031.

When the prior is the same for all models,

$$\frac{P(\boldsymbol{\alpha_1}|D)}{P(\boldsymbol{\alpha_2}|D)} = \frac{P(D|\boldsymbol{\alpha}_1)}{P(D|\boldsymbol{\alpha}_2)}, \tag{3.3}$$

where $\frac{P(D|\boldsymbol{\alpha}_1)}{P(D|\boldsymbol{\alpha}_2)}$ is the Bayes Factor (Kass and Raftery, 1995). This emphasizes the well known fact that that posterior model probabilities in BMA are determined by a model's ability to make prediction, which may be different from is ability to properly adjust for confounding in effect estimation. The sixth column in Table 2 lists the model weights used by BMA in Example 1, highlighting that most of the weight (92.7%) is assigned to model (2.3), which provides an estimation of β equal to 0.160 and a 95% C.I. (0.116, 0.203). In contrast, only 6% of the weight is assigned to the true model (2.2) which estimates the correct β^*. Thus, the BMA estimate for β becomes seriously biased with an associated 95% C.I. (0.109, 0.207) that does not cover the true value of 0.1.

3.2. *Bayesian Confounding Adjustment (BCA).* In developing a model for effect estimation, when a true confounder is added or removed from the regression model, the interpretation of the exposure coefficient changes; however, when a model includes all true confounders, and one adds an additional covariate that is not associated with X or that is not associated with X nor Y, the interpretation of the exposure coefficient does not change. This is in contrast to the prediction framework, in which the outcome typically maintains the same interpretation across models. This important difference suggests that a novel BMA-like approach, that acknowledges the fact that only a fraction of the models harbor the coefficient of interest, could be successful in addressing adjustment uncertainty from a Bayesian standpoint.

We propose to pursue this idea via a weighting mechanism called Bayesian Confounding Adjustment (BCA) that considers jointly the exposure and

outcome models. The exposure model is:

$$E\{x_t\} = \sum_{l=1}^{L} \gamma_l^{\boldsymbol{\alpha}^X} z_{tl} + \sum_{m=1}^{M} \alpha_m^X \delta_m^{\boldsymbol{\alpha}^X} u_{tm} \,, \tag{3.4}$$

where the unknown $\boldsymbol{\alpha}^X$ indicates which potential confounders are included and therefore predictive of the exposure. Those are the requisite confounders for appropriate adjustment in the outcome model described in Equation (2.1). We therefore propose to define the outcome model as

$$E\{y_t\} = \beta^{\boldsymbol{\alpha}^Y} x_t + \sum_{l=1}^{L} \gamma_l^{\boldsymbol{\alpha}^Y} z_{tl} + \sum_{m=1}^{M} \alpha_m^Y \delta_m^{\boldsymbol{\alpha}^Y} u_{tm}. \tag{3.5}$$

where $\boldsymbol{\alpha}^Y$ indicates which potential confounders are included and is dependent on $\boldsymbol{\alpha}^X$ in that $P(\alpha_m^Y = 1 | \alpha_m^X = 1) = 1$, $m = 1, \ldots, M$. In this way, requisite confounders will be automatically included in the outcome model, while additional ones may be included to improve fit. For brevity, we will often refer to $\boldsymbol{\alpha}$'s as "models".

The posterior distribution of β^* (exposure coefficient under the minimal model) may be written as:

$$P(\beta^*|D) = \sum_{\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y} P(\beta^*, \boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y | D) = \sum_{\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y} P(\beta^* | \boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y, D) P(\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y | D). \tag{3.6}$$

For $\boldsymbol{\alpha}^Y$ nesting the minimal model, $\beta^{\boldsymbol{\alpha}^Y} = \beta^*$ so that $P(\beta^* | \boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y, D) = P(\beta^{\boldsymbol{\alpha}^Y} | \boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y, D)$. BCA is designed to assign large weights to models that are nesting the minimal model, so that, approximately:

$$P(\beta^*|D) \doteq \sum_{\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y} P(\beta^{\boldsymbol{\alpha}^Y} | \boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y, D) P(\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y | D), \tag{3.7}$$

which may be estimated from observed data.

This idea can be illustrated using the simulation based on Figure 1 as an example. The seventh column of Table 2 listed the model posterior weights based on BCA: 98.5% of the weight is assigned to the true model, compared to only 6% assigned to the same model by FBMA. No weight is assigned to models not nesting the minimal model, compared to 92.9% in total assigned by FBMA. This result illustrates that BCA can assigns large weights models including the minimal model while at the same time BMA can fail to do so. The large difference between BCA and FBMA is due to the fact that BCA incorporates additional information from the exposure models, and thus is

more likely to include covariates that are correlated with both exposure and outcome, which are usually important in effect estimation but may not be as important in model prediction.

Algorithmically, Monte Carlo methods are used for computing. Equations (3.4) and (3.5) jointly define the likelihood function, and via Bayes rule, the posterior distribution $P(\beta^*, \boldsymbol{\alpha}^Y, \boldsymbol{\alpha}^X|D)$. According to (3.6), the sampled βs approximate a draw from $P(\beta^*|D)$ and thus may be used to estimate $P(\beta^*|D)$. The parameters $(\beta^*, \boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y)$ can be sampled using, for example, the Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990). According to (3.7), β^* may be replaced by $\beta^{\boldsymbol{\alpha}^Y}$ when sampling from its conditional distribution given $(\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y)$.

In our implementation, we assume the following priors:

$$(\boldsymbol{\gamma}^{\boldsymbol{\alpha}^X}, \boldsymbol{\delta}^{\boldsymbol{\alpha}^X}) \sim N(\boldsymbol{\mu}_{0\boldsymbol{\alpha}^X}, (\tau_X)^{-1}\boldsymbol{\Sigma}_{0\boldsymbol{\alpha}^X}), \quad (\beta^{\boldsymbol{\alpha}^Y}, \boldsymbol{\gamma}^{\boldsymbol{\alpha}^Y}, \boldsymbol{\delta}^{\boldsymbol{\alpha}^Y}) \sim N(\boldsymbol{\mu}_{0\boldsymbol{\alpha}^Y}, (\tau_Y)^{-1}\boldsymbol{\Sigma}_{0\boldsymbol{\alpha}^Y}),$$

$$\tau_X, \tau_Y \sim Gamma\left(\frac{\nu}{2}, \frac{\nu\lambda}{2}\right),$$

where ν, λ, the $(L+M)$-vector $\boldsymbol{\mu}_{0\boldsymbol{\alpha}^X}$, the $(L+M+1)$-vector $\boldsymbol{\mu}_{0\boldsymbol{\alpha}^Y}$, the $(L+M)\times(L+M)$-matrix $\boldsymbol{\Sigma}_{0\boldsymbol{\alpha}^X}$ and the $(L+M+1)\times(L+M+1)$-matrix $\boldsymbol{\Sigma}_{0\boldsymbol{\alpha}^Y}$ are chosen hyperparameters. In practice, we chose those hyperparameters following Raftery et al. (1997).

We assume that $(\beta^{\boldsymbol{\alpha}^Y}, Y)$ are independent of $\boldsymbol{\alpha}^X$ given $\boldsymbol{\alpha}^Y$, and that X is independent of $\boldsymbol{\alpha}^Y$ given $\boldsymbol{\alpha}^X$. We also assume $(\beta^{\boldsymbol{\alpha}^Y}, X)$ and $\boldsymbol{\alpha}^Y$ are independent given $Y^* = Y - \beta^{\boldsymbol{\alpha}^Y} X$, and $\boldsymbol{\alpha}^X$ and Y^* are independent given $\boldsymbol{\alpha}^Y$. For a sample with n observations, if $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Y}^*$ denote the vectors of observed data for X ,Y and Y^*, then the three full conditional distributions of interest are:

$$P(\boldsymbol{\alpha}^X|\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{\alpha}^Y, \beta^*) \doteq \frac{P(\boldsymbol{\alpha}^X)P(\boldsymbol{\alpha}^Y|\boldsymbol{\alpha}^X)P(\boldsymbol{X}|\boldsymbol{\alpha}^X)}{P(\boldsymbol{X}, \boldsymbol{\alpha}^Y)} \propto P(\boldsymbol{\alpha}^X)P(\boldsymbol{\alpha}^Y|\boldsymbol{\alpha}^X)P(\boldsymbol{X}|\boldsymbol{\alpha}^X)$$

$$P(\boldsymbol{\alpha}^Y|\boldsymbol{\alpha}^X, \beta^*, \boldsymbol{Y}, \boldsymbol{X}) \doteq P(\boldsymbol{\alpha}^X)P(\boldsymbol{\alpha}^Y|\boldsymbol{\alpha}^X)P(\boldsymbol{Y}^*|\boldsymbol{\alpha}^Y) \propto P(\boldsymbol{Y}^*|\boldsymbol{\alpha}^Y)P(\boldsymbol{\alpha}^Y|\boldsymbol{\alpha}^X)$$

$$\beta^*|\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y, \boldsymbol{X}, \boldsymbol{Y} \dot{\sim} t_{n+\nu}(\beta_{n\boldsymbol{\alpha}^Y}, \sigma^2_{n\boldsymbol{\alpha}^Y}).$$

The definition of $\beta_{n\boldsymbol{\alpha}^Y}$, $\sigma^2_{n\boldsymbol{\alpha}^Y}$ and the derivation of these conditional distributions are shown in Appendix. Using results in Raftery et al. (1997), we

obtain

(3.8)

$$P(\boldsymbol{X}|\boldsymbol{\alpha}^X) = \frac{\Gamma(\frac{\nu+n}{2})(\nu\lambda)^{\nu/2}}{\pi^{n/2}\Gamma(\frac{\nu}{2})|\boldsymbol{I}_n + \boldsymbol{W}_{\boldsymbol{\alpha}^X}\boldsymbol{\Sigma}_{0\boldsymbol{\alpha}^X}\boldsymbol{W}'_{\boldsymbol{\alpha}^X}|^{1/2}} \times [\lambda\nu + (\boldsymbol{X} - \boldsymbol{W}_{\boldsymbol{\alpha}^X}\boldsymbol{\mu}_{0\boldsymbol{\alpha}^X})'(\boldsymbol{I}_n + \boldsymbol{W}_{\boldsymbol{\alpha}^X}\boldsymbol{\Sigma}_{0\boldsymbol{\alpha}^X}\boldsymbol{W}'_{\boldsymbol{\alpha}^X})^{-1}(\boldsymbol{X} - \boldsymbol{W}_{\boldsymbol{\alpha}^X}\boldsymbol{\mu}_{0\boldsymbol{\alpha}^X})]^{-(\nu+n)/2}$$

$$P(\boldsymbol{Y}^*|\boldsymbol{\alpha}^Y) = \frac{\Gamma(\frac{\nu+n}{2})(\nu\lambda)^{\nu/2}}{\pi^{n/2}\Gamma(\frac{\nu}{2})|\boldsymbol{I}_n + \boldsymbol{W}_{\boldsymbol{\alpha}^Y}\boldsymbol{\Sigma}_{0\boldsymbol{\alpha}^Y}\boldsymbol{W}'_{\boldsymbol{\alpha}^Y}|^{1/2}} \times [\lambda\nu + (\boldsymbol{Y}^* - \boldsymbol{W}_{\boldsymbol{\alpha}^Y}\boldsymbol{\mu}_{0\boldsymbol{\alpha}^Y})'(\boldsymbol{I}_n + \boldsymbol{W}_{\boldsymbol{\alpha}^Y}\boldsymbol{\Sigma}_{0\boldsymbol{\alpha}^Y}\boldsymbol{W}'_{\boldsymbol{\alpha}^Y})^{-1}(\boldsymbol{Y}^* - \boldsymbol{W}_{\boldsymbol{\alpha}^Y}\boldsymbol{\mu}_{0\boldsymbol{\alpha}^Y})]^{-(\nu+n)/2}.$$

where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix, $\boldsymbol{W}_{\boldsymbol{\alpha}^X}$, $\boldsymbol{W}_{\boldsymbol{\alpha}^Y}$ are the design matrices of the exposure and outcome regressions respectively.

In this way, samples from the first two conditional distributions can be generated using the MC^3 method proposed by Madigan and York (1995). For the third conditional distribution, we can directly sample from a t-distribution. Software for BCA is available upon request.

3.3. *The CDP method.* Crainiceanu et al. (2008) developed a two-stage frequentist method, here called CDP, to account for adjustment uncertainty in exposure effect estimation in regression modeling with large number of potential confounders. Their method has two steps. In the first step, the CDP method regresses X on potential confounders $\boldsymbol{U}$ and identifies predictors of the exposure. More precisely, the exposure model space is divided into $M+1$ subsets, or orbits, where the mth orbit is the set of all regression models $\boldsymbol{\alpha}^X$ with the the same number, m, of potential confounders. Within each orbit, the maximum likelihood model is selected and denoted by $\boldsymbol{\alpha}^X_m$. The region where the deviance difference function, $D(\boldsymbol{\alpha}^X_m) - D(\boldsymbol{\alpha}^X_{m+1})$, becomes small identifies a range for the required dimensionality of the exposure model. This identifies a set of confounders highly correlated with X. In the second step, CDP identifies strong predictors of outcome among the remaining confounders, which are weak predictors of exposure. The same orbit searching procedure is applied to regression models for outcome with the constrain that covariates selected from the first step are always included. Software is available at `www.biostat.jhsph.edu/~ccrainic/webpage/software/STEADy.zip`.

The BCA method proposed in section 3.2 parallels the CDP method, but has some important differences. For example, by averaging models according to their ability to estimate the exposure effect, BCA summarizes information in a Bayesian way and arguably considers uncertainty more fully. In contrast, CDP considers the problem from a frequentist prespective and models are evaluated based on the change in deviance between adjacent orbits. However, both methods have the same ultimate goal, which is effect estimation

accounting for adjustment uncertainty. In this paper we show that, when implemented correctly, both philosophies can produce the similar results.

4. Simulations. In this section, we conduct simulations studies to illustrate and compare the practical properties of the Bayesian Confounding Adjustment (BCA) procedure versus the method developed by Crainiceanu et al. (2008) (CDP), and versus Bayesian Model Averaging (BMA). For BMA, we consider two different implementations: the first is forcing the exposure to always be in the model (FBMA), while the second one (NBMA) is not.

Our first scenario is similar to the one in Crainiceanu et al. (2008) and considers the following true model:

$$Y_i = \beta X_i + \delta_1 U_{1i} + \delta_2 U_{2i} + \epsilon_i, \tag{4.1}$$

where $i = 1, ..., 1000$. (X_i, U_{1i}, U_{2i}) are independent normal vectors with mean zero and a covariance matrix, $\Sigma = (\sigma_{kl})_{3\times 3}$, where $\sigma_{kk} = 1, k = 1, 2, 3$, $\sigma_{12} = \sigma_{21} = \rho$, and $\sigma_{13} = \sigma_{23} = \sigma_{31} = \sigma_{32} = 0$. The set of potential confounders $\boldsymbol{U}$ includes U_1, U_2 as well as 49 additional independent $N(0, 1)$ random variables. In our simulation, ρ is set to 0.7 and $\beta = \delta_1 = \delta_2 = 0.1$. We generated 25 data sets from model (4.1). For each data set, we calculated the Maximum Likelihood Estimate (MLE) of β from the true model and compared it with the estimates from the four estimation methods: BCA, CDP, FBMA and NBMA. The point estimates of β from these methods are shown in Figure 2 for all 25 simulated data sets. BCA (dashed line with triangles) and CDP (dotted line with plus signs) produce very similar estimates, both close to the estimates obtained from the true model, shown as a solid line with filled dots. As expected, the true model's estimates oscillate around 0.1, the true value of β. In contrast, most of the point estimates based on FBMA (dot-dashed line with asterisk sign) are larger than 0.1, indicating that FBMA systematically overestimates the exposure effect in this example. The point estimates based on NBMA (long dashed line with diamonds) are much more variable than those base on BCA: the standard error of those point estimates based on NBMA is 0.082, compared to 0.045 for estimates based on BCA.

The difference between BCA and CDP on one side, and BMA approaches on the other, is even more pronounced when comparing confidence intervals (CI), shown in Figure 3. The 50% CIs based on BCA and CDP cover the true value of the parameter, 0.1, in roughly 50% of the simulations (11 of 25), whereas the C.I.s based on FBMA cover 0.1 only in 7 out of the 25 simulations. A close inspection of the lengths of CIs indicates that the variability of FBMA point estimates is not typically larger than that for BCA and CDP.

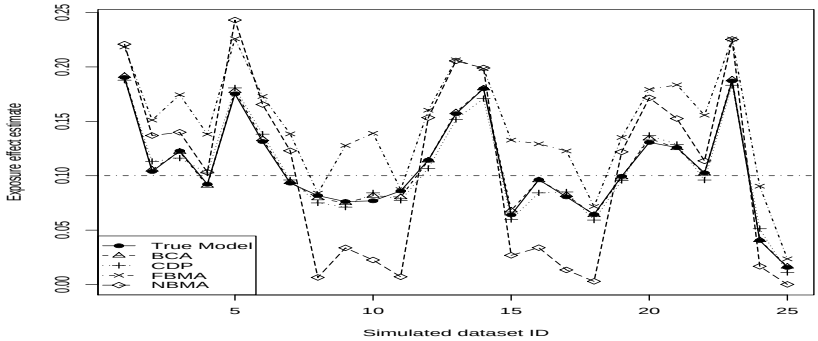

FIG 2. *Estimation of β from four methods along with the true model in 25 simulated data sets in the first scenario (β's true value is shown by a horizonal dot-dashed line)*

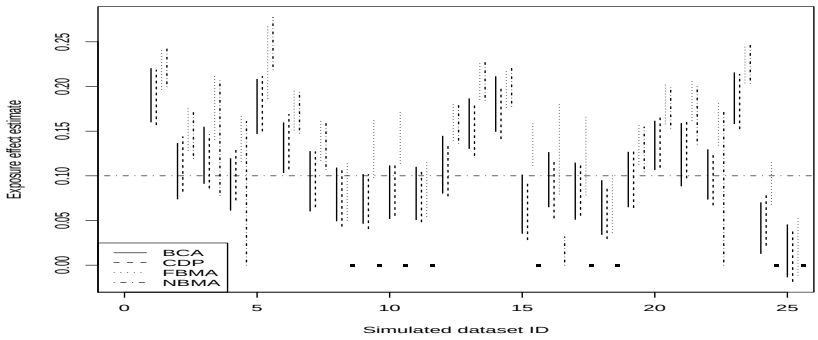

FIG 3. *50% C.I.s of β from the four methods in 25 simulated data sets in the first scenario (β's true value is shown by a horizonal dot-dashed line). The solid dots indicating C.I.s of length zero.*

In other words, FBMA estimates have larger bias and similar variance with respect to BCA and CDP estimates. For the CIs based on NBMA, some of the confidence intervals collapsed to a single point, 0, corresponding to no exposure effect. The reason is that more than 50% of the weight is assigned to models that do not include exposure, which estimate β as zero.

To gain more insight into these differences, it is helpful to inspect the results for one specific simulation. For example, in the second simulated data set, BCA provides a point estimate of 0.106 with 50% C.I. (0.074, 0.137), which covers the true value, 0.1, while FBMA estimate β to be 0.151 with 50% C.I. (0.128, 0.176). The reason why FBMA overestimates the exposure effect in this example is that FBMA averages over models that do not include U_1, while in this case U_1, which is strongly associated with the exposure X and less strongly associated with outcome Y, is a critical

variable in confounding adjustment. Ignoring the joint distribution of (Y, X) and focusing only on the distribution of Y leads to overestimation. While in this case BMA overestimates the exposure effect, in general, the sign of the bias is hard to predict without conducting a more appropriate adjustment.

Table 3 further summarizes the simulation results. If $\hat{\beta}_i$ is the point estimate based on the ith simulated data set and β_0 is the true value of β, we define the mean, $M(\hat{\beta})$, and mean square error, $MSE(\hat{\beta})$, for a given estimation procedure as

$$M(\hat{\beta}) = \frac{1}{25}\sum_{i=1}^{25}\hat{\beta}_i, \qquad MSE(\hat{\beta}) = \frac{1}{25}\sum_{i=1}^{25}(\hat{\beta}_i - \beta_0)^2. \tag{4.2}$$

We also define the mean standard error, $\overline{SE}(\hat{\beta})$, and the standard error across point estimates, $\widehat{SE}(\hat{\beta})$, as

$$\overline{SE}(\hat{\beta}) = \sqrt{\frac{1}{25}\sum_{i=1}^{25}\mathrm{Var}(\hat{\beta}_i)}, \qquad \widehat{SE}(\hat{\beta}) = \sqrt{\frac{1}{25-1}\sum_{i=1}^{25}(\hat{\beta}_i - M(\hat{\beta}))^2}, \tag{4.3}$$

where $\mathrm{Var}(\hat{\beta}_i)$ is the variance of $\hat{\beta}_i$.

We conclude that in this simulation, the MSE of BCA and CDP estimates are roughly the same as the MSE of estimates based on the true model. In addition these MSE are much smaller than the MSE of FBMA estimates. The mean of point estimates based on NBMA is 0.110, which is close to the means for BCA and CDP. Despite this good average behavior, NBMA produces the worst results. Indeed, the MSE for NBMA is 0.007, which is much higher than 0.002 for BCA and CDP. Moreover, the C.I.s based on NBMA cover 0.1 only in 4 out of the 25 simulations. The histogram of the point estimates from NBMA (Figure 4) reveals why NBMA has small bias and large MSE: the distribution has three modes, and while it is centered roughly around the true value, this value falls in a region of low mass. Thus, NBMA rarely provides an estimate close to the true value, even though the average of the point estimates across data sets is close to it.

Our second simulation scenario considered a larger number of covariates that are correlated with the exposure variable and also associated with the outcome. We considered both covariates highly and weakly correlated with exposure and assumed the following true outcome model:

$$Y_i = \beta X_i + \delta_1 U_{1i} + ... + \delta_{14} U_{14i} + \epsilon_i, \tag{4.4}$$

where $i = 1, ..., 1000$ and $(X_i, U_{1i}, ..., U_{7i})$ are independent normal vectors with mean zero and a covariance matrix, $\Sigma = (\sigma_{kl})_{8\times 8}$, corresponding to

	True model	BCA	CDP	FBMA	NBMA
M($\hat{\beta}$)	0.108	0.108	0.107	0.147	0.106
SE($\hat{\beta}$)	0.045	0.045	0.045	0.044	0.052
SE of $\hat{\beta}$ across data sets	0.045	0.045	0.045	0.050	0.082
MSE($\hat{\beta}$)	0.002	0.002	0.002	0.005	0.006
Coverage rate of 50% C.I.	14/25	15/25	14/25	7/25	3/25

TABLE 3
Comparison of estimation of β from four methods along with the true model in the first simulation scenario. The target coverage rate is 12.5/25.

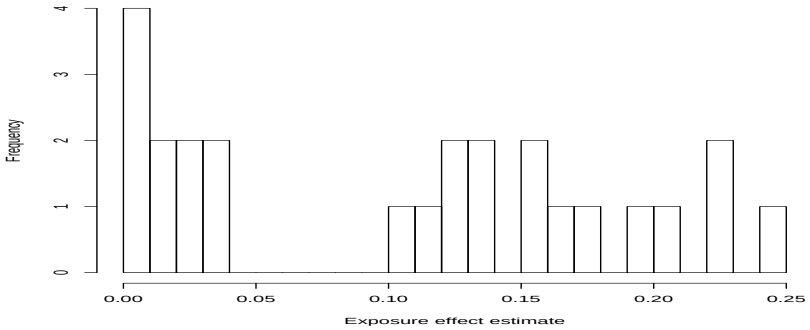

FIG 4. *Histogram of the point estimates based on NBMA in 25 simulated data sets in the first scenario*

an $AR(1)$ process. More precisely, $\sigma_{kk} = 1, \sigma_{kl} = \rho^{|k-l|}, 1 \leq k, l \leq 8$. We also assumed that the rest of the confounders $U_{8i}, ..., U_{14i}$ follows $N(0,1)$ distribution and are independent of the other covariates in model 4.4. The set of potential confounders $\mathbf{U}$ includes $U_1, ...U_{14}$ as well as 43 additional independent $N(0,1)$ random variables which are independent with both X and Y. In our simulation, β is set to be 0.1, $\delta_1 = \ldots = \delta_{14} = 0.1$ and $\rho = 0.7$.

Similarly to the first scenario, we generated 25 data sets from model (4.4). For each simulated data set, we calculated the Maximum Likelihood Estimate (MLE) of β from the known true model and compared it to the estimates from the four methods: BCA, CDP, FBMA and NBMA. Similarly to Figure 2, 3 and Table 2 in the first scenario, we obtained the results shown in Figure 5, 6 and Table 3.

The differences we noted between BCA and CDP on one side, and BMA on the other, are now even more pronounced in this more complex example. In Figure 5, the point estimate obtained using FBMA is biased and larger than

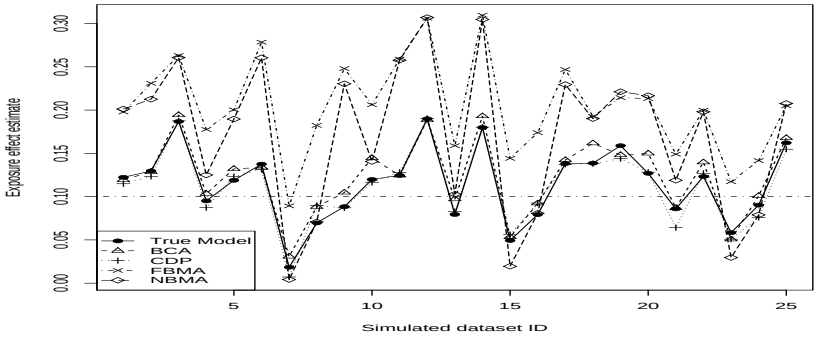

FIG 5. *Estimation of β from four methods along with the true model in 25 simulated data sets in the second scenario (β's true value is shown by a horizonal dot-dashed line)*

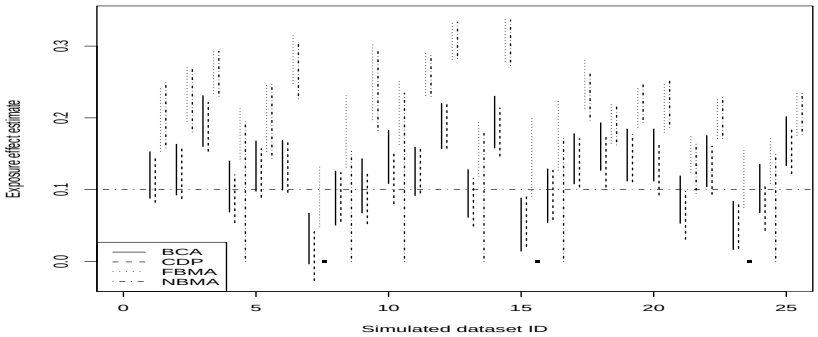

FIG 6. *50% C.I. β from the four methods in 25 simulated data sets in the second scenario (β's true value is shown by a horizonal dot-dashed line).The solid dots indicating C.I.s of length zero.*

	True Model	BCA	CDP	FBMA	NBMA
M($\hat{\beta}$)	0.115	0.123	0.113	0.204	0.170
SE($\hat{\beta}$)	0.050	0.051	0.051	0.055	0.069
SE of $\hat{\beta}$ across data sets	0.043	0.043	0.044	0.055	0.088
MSE($\hat{\beta}$)	0.002	0.002	0.002	0.014	0.012
Coverage rate of 50% C.I.	14/25	12/25	14/25	3/25	7/25

TABLE 4

Comparison of estimation of β from the four methods along with the gold standard in the second simulation scenario. The target coverage rate is 12.5/25.

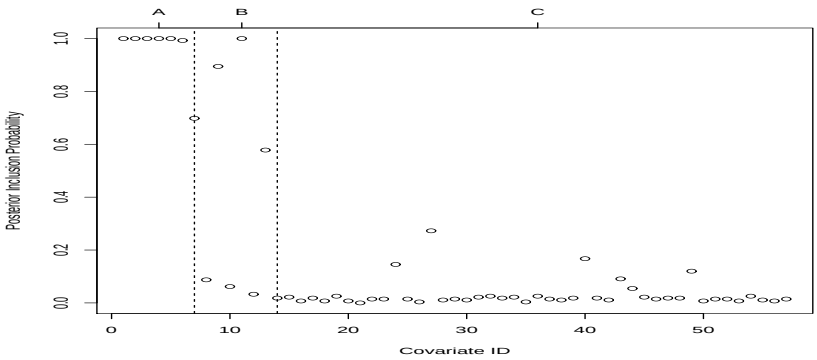

FIG 7. *Posterior inclusion probability of covariates (the covariates are seperated into three groups by two vertical dashed lines, where the first 7 covariates (group A) are in the true model and correlated with X, the next 7 covariates (group B) are in the true model but independent with x, the rest (group C) are covariates not in the true model and independent with X)*

the point estimates based on the true model. Only 3 of the 50% confidence interval in the 25 simulated data sets cover the true value. In contrast, the point estimates based on BCA and CDP are close to those based on the true model, and the corresponding 50% C.I.s cover the true parameter value in roughly 50% of the simulations.

We also computed the posterior inclusion probability (Barbieri and Berger, 2004) defined, for the ith covariate, as:

$$p_i = \sum_{\boldsymbol{\alpha}:\alpha_i=1} P(\boldsymbol{\alpha}|D). \tag{4.5}$$

p_i can be estimated by the proportion of appearances of covariate i in the chain of outcome models. Figure 7 shows the estimated posterior inclusion probability for all the covariates based on a simulated data set from our second scenario, using BCA. The first seven covariates have high posterior inclusion probability, indicating that they are important for estimating the exposure effect β. This is consistent with their high correlation with X.

5. Estimating the Effect of $PM_{2.5}$ on CVD Hospitalization Rate. In this section we applied BCA, CDP, FBMA, and NBMA to daily time series data on emergency hospital admissions, weather variables, and daily $PM_{2.5}$ levels in Nassau County NY for the period 1999-2005. A more extensive description of this data set can be found in Dominici et al. (2006). The goal is to estimate the increase in hospitalization rate of cardiovascular disease (CVD) associated with a $10\mu_g/m^3$ increase in $PM_{2.5}$. To start,

	Full model	BCA	CDP	FBMA	NBMA
$\hat{\beta}$	0.291	0.220	0.221	0.140	0.007
SE($\hat{\beta}$)	0.092	0.081	0.089	0.077	0.033
95% C.I.	(0.110, 0.471)	(0.068, 0.373)	(0.045, 0.396)	(-0.008, 0.298)	(0.000, 0.131)

TABLE 5
Comparison of estimation of $PM_{2.5}$ effect on CVD hospitalization rate based on BCA, CDP, FBMA, NBMA and the full model. The data is a time series data set from Nassau, NY for the period 1999-2005.

we consider a full model that is large enough to include all the necesssary confounders (Dominici et al., 2004; Peng et al., 2006):

$$\begin{aligned} Y_t &= PM_{2.5t} + Dow + Age \\ &+ ns(Temp_t, df_{Temp}) + ns(Temp_{t1-3}, df_{Temp}) + ns(Dew, df_{Dew}) \\ &+ ns(Dew_{t1-3}, df_{Dew}) + ns(t, df_t) + ns(t, 4) \times Age + \epsilon_t \end{aligned} \tag{5.1}$$

where $Y_t = \sqrt{\frac{\text{number of hospital admissions of CVD on day t}}{\text{size of population at risk on day t}}}$ serves as outcome. $PM_{2.5t}$ denotes the level of particulate matter having diameter less than 2.5 micrometer on day t, DOW is a categorical variable indicating the day of the week, Age is age group indicators ($\geq$75 or not). $Temp_t$ and $Temp_{t1-3}$ are temperature on day t and the three day running mean respectively. Dew_t and Dew_{t1-3} are the dew point on day t and the three day running mean. $ns(, .df)$ is a natural cubic spline with df degree of freedom. $ns(t, df_t)$ is to capture the seasonal effect and $ns(t, 4) \times Age$ is the interaction terms of seasonal effect and age. In this example, df_{Temp} is set to 12 and df_{Dew} are set to 12 while df_t are set to 16 per year. The residues ϵ_t are assumed to be independent and identically distributed with a normal distribution ($N(0, \sigma^2)$). After dropping some covariates due to collinearity, there were 164 covariates left, which forms the set of potential confounders.

As in the simulations, we consider 4 approaches: BCA, CDP, FBMA, and NBMA. The estimated $PM_{2.5}$ effect ($\times 10,000$) denoted by $\hat{\beta}$ is listed in Table 5: BCA and CDP provide an estimate of the short-term effect of $PM_{2.5}$ on CVD hospital admissions with a 95% posterior interval and 95% confidence interval that do not include 0. The point estimates of the exposure effect obtained under the full model, BCA and CDP are similar. Moreover, BCA and CDP provide smaller standard errors than the one obtained under the full model. In comparison, FBMA and NBMA provide a very different and non statistically significant estimate of the exposure effect. This illustrates that in practical applications BMA and BCA can lead to different conclusions.

6. Discussion. Estimating an exposure effect accounting for the uncertainty in the adjustment for confounding is complex, and direct application of methods designed to handle model uncertainty in prediction can give misleading results. Building upon work by Crainiceanu et al. (2008), in this paper we further explore the difference between model uncertainty and adjustment uncertainty, and develop a fully Bayesian solution to adjustment uncertainty called Bayesian Confounding Adjustment (BCA). By providing examples and simulation studies we illustrate that adjustment uncertainty and model uncertainty need to be addressed using different inferential procedures.

Given a set of potential confounders, the approach presented in this paper is to specify joint models for both the outcome and the exposure of interest, and constrain confounders that are included in the exposure model to also appear in the outcome model. In simulation studies, we show that BCA provides posterior distributions of the parameter of interest based on averaging across models that are highly likely to include the necessary confounders for adjustment. While we discussed our methods in the setting of linear models, BCA is a general concept and is not constrained to the linear case. For example, at least conceptually, it can be easily extended to generalized linear models.

We also show that the standard BMA method is not adequate to address adjustment uncertainty in effect estimation, possibly because the model-specific estimates of the parameter of interest are weighted based on the model ability of making good prediction, which can be different from their ability to properly adjust for confounding. We conduct simulation studies under two scenarios with different degree of confounding. Our results show that BCA and CDP can provide estimate for exposure effect with small MSE in both scenarios. In contrast, BMA may lead to highly biased estimate. We also compare different methods by conducting a data analysis of time series data from Nassau, NY. In this specific example BCA and CDP provide statistically significant positive estimates for the $PM_{2.5}$ effect on CVD hospitalization rate, while BMA does not.

As BMA, BCA will take weighted average over models rather than making inference based on a single model. However BCA attempts to provide an estimate of the exposure effect by combining information across regression models that all include a model including all the requisite confounders, to ensure that the regression coefficient of interest maintains the same interpretation across models. A nice feature of BMA that is retained by BCA is that the importance of confounders can be evaluated based on posterior inclusion probability. This information may reveal underlying connections

between exposure and confounders, which may become of interest for future research. BCA is more computationally intensive than CDP and BMA.

As with CDP, the successful application of BCA may rely on assumptions that are hard to verify. First, we need to assume that there are no unobserved confounders. Second, we need to define a large enough model to include all the potential observed confounders. Third we must assume that this large set of potential confounders do not include covariates that are highly predictive of X but that are not associated with Y. Scientific knowledge is required to ensure that these assumptions are valid. Statistical methods may also help to check whether there is evidence for the existence of unmeasured confounders. For example, one can decompose the association between exposure and outcome into distinct spatio-temporal scales and check for the consistency in the estimation of exposure effect across these spatio-temporal scales (Janes et al., 2007).

If there are no unmeasured confounders , the full model, that is the model including all variables that are correlated with X and Y, and those that are correlated with Y only, as well as potentially others that are not associated with either, can provide unbiased estimate for the exposure effect. However, using the full model we will generally have wider confidence intervals compared to BCA. By combining estimation from different smaller models, especially from models that only include requisite confounders but do not include many unnecessary covariates, BCA provides more precise inference than the full model.

In summary, we hope to have provided a well motivated and practical tool for accounting for uncertainty in the selection of confounders in effect estimation. Our approach adopts the fully probabilistic structure of BMA without suffering from the pitfalls we highlighted in BMA, and is likely to contribute to a more reasoned and quantitative approach to the specification of models used to determine health effects of common exposures, and the reporting of the associated uncertainty.

APPENDIX A

A.1. Full conditional distribution of $\boldsymbol{\alpha}^X$

Assume $(\beta^{\boldsymbol{\alpha}^Y}, Y)$ and $\boldsymbol{\alpha}^X$ are independent given $\boldsymbol{\alpha}^Y$,

$$\begin{aligned} & P(\boldsymbol{\alpha}^X|\boldsymbol{X},\boldsymbol{Y},\boldsymbol{\alpha}^Y,\beta^*) \\ \doteq\ & P(\boldsymbol{\alpha}^X|\boldsymbol{X},\boldsymbol{Y},\boldsymbol{\alpha}^Y,\beta^{\boldsymbol{\alpha}^Y}) \\ =\ & P(\boldsymbol{\alpha}^X|\boldsymbol{X},\boldsymbol{\alpha}^Y) \\ =\ & \frac{P(\boldsymbol{\alpha}^X)P(\boldsymbol{\alpha}^Y|\boldsymbol{\alpha}^X)P(\boldsymbol{X}|\boldsymbol{\alpha}^X,\boldsymbol{\alpha}^Y)}{P(\boldsymbol{X},\boldsymbol{\alpha}^Y)}. \end{aligned}$$

Assume X and $\boldsymbol{\alpha}^Y$ are independent given $\boldsymbol{\alpha}^X$, then

$$\begin{aligned} & P(\boldsymbol{\alpha}^X|\boldsymbol{X},\boldsymbol{Y},\boldsymbol{\alpha}^Y,\beta^*) \\ \doteq\ & \frac{P(\boldsymbol{\alpha}^X)P(\boldsymbol{\alpha}^Y|\boldsymbol{\alpha}^X)P(X|\boldsymbol{\alpha}^X)}{P(\boldsymbol{X},\boldsymbol{\alpha}^Y)}. \end{aligned}$$

A.2. Full conditional distribution of $\boldsymbol{\alpha}^Y$

Assume $(\beta^{\boldsymbol{\alpha}^Y}, X)$ and $\boldsymbol{\alpha}^Y$ are independent given Y^*,

$$\begin{aligned} & P(\boldsymbol{\alpha}^Y|\boldsymbol{X},\boldsymbol{Y},\boldsymbol{\alpha}^X,\beta^*) \\ \doteq\ & P(\boldsymbol{\alpha}^Y|\boldsymbol{X},\boldsymbol{Y},\boldsymbol{\alpha}^X,\beta^{\boldsymbol{\alpha}^Y}) \\ =\ & P(\boldsymbol{\alpha}^Y|\boldsymbol{Y}^*,\boldsymbol{\alpha}^X) \\ =\ & \frac{P(\boldsymbol{\alpha}^Y)P(\boldsymbol{\alpha}^X|\boldsymbol{\alpha}^Y)P(\boldsymbol{Y}^*|\boldsymbol{\alpha}^Y,\boldsymbol{\alpha}^X)}{P(\boldsymbol{Y}^*,\boldsymbol{\alpha}^X)}. \end{aligned}$$

Assume $\boldsymbol{\alpha}^X$ and Y^* are independent given $\boldsymbol{\alpha}^Y$,

$$\begin{aligned} & P(\boldsymbol{\alpha}^Y|\boldsymbol{X},\boldsymbol{Y},\boldsymbol{\alpha}^X,\beta^*) \\ \doteq\ & P(\boldsymbol{\alpha}^Y)P(\boldsymbol{\alpha}^X|\boldsymbol{\alpha}^Y)P(\boldsymbol{Y}^*|\boldsymbol{\alpha}^Y). \end{aligned}$$

A.3. Full conditional distribution of β

Assume $\beta^{\boldsymbol{\alpha}^Y}$ and $\boldsymbol{\alpha}^X$ are independent given $\boldsymbol{\alpha}^Y$, then

$$P(\beta^{\boldsymbol{\alpha}^Y}|\boldsymbol{\alpha}^X,\boldsymbol{\alpha}^Y,\boldsymbol{X},\boldsymbol{Y}) = P(\beta^{\boldsymbol{\alpha}^Y}|\boldsymbol{\alpha}^Y,\boldsymbol{X},\boldsymbol{Y}).$$

Assuming the observations of X are placed on the first column of $\boldsymbol{W}_{\boldsymbol{\alpha}^Y}$, from Bernardo and Smith (2000), we obtain

$$\beta^{\boldsymbol{\alpha}^Y}|\boldsymbol{\alpha}^Y,\boldsymbol{X},\boldsymbol{Y} \sim t_{n+\nu}(\beta_{n\boldsymbol{\alpha}^Y},\sigma^2_{n\boldsymbol{\alpha}^Y}),$$

where $\beta_{n\boldsymbol{\alpha}^Y}$ is the first element of $\boldsymbol{\theta}_{n\boldsymbol{\alpha}^Y}$, $\sigma^2_{n\boldsymbol{\alpha}^Y}$ is the (1,1) element of $\widetilde{S}_{n\boldsymbol{\alpha}^Y}(\boldsymbol{W}'_{\boldsymbol{\alpha}^Y}\boldsymbol{W}_{\boldsymbol{\alpha}^Y}+\boldsymbol{\Sigma}^{-1}_{0\boldsymbol{\alpha}^Y})^{-1})$ and

$$\boldsymbol{\theta}_{n\boldsymbol{\alpha}^Y} = (\boldsymbol{W}'_{\boldsymbol{\alpha}^Y}\boldsymbol{W}_{\boldsymbol{\alpha}^Y}+\boldsymbol{\Sigma}^{-1}_{0\boldsymbol{\alpha}^Y})^{-1}(\boldsymbol{\Sigma}^{-1}_{0\boldsymbol{\alpha}^Y}\boldsymbol{\mu}_{0\boldsymbol{\alpha}^Y}+\boldsymbol{W}'_{\boldsymbol{\alpha}^Y}\boldsymbol{Y})$$

$$\widetilde{S}_{n\boldsymbol{\alpha}^Y} = (n+\nu)^{-1}[\nu\lambda+(\boldsymbol{Y}-\boldsymbol{W}_{\boldsymbol{\alpha}^Y}\boldsymbol{\theta}_{n\boldsymbol{\alpha}^Y})'\boldsymbol{Y}+(\boldsymbol{\mu}_{0\boldsymbol{\alpha}^Y}-\boldsymbol{\theta}_{n\boldsymbol{\alpha}^Y})'\boldsymbol{\Sigma}^{-1}_{0\boldsymbol{\alpha}^Y}\boldsymbol{\mu}_{0\boldsymbol{\alpha}^Y}].$$

Thus,

$$\beta^*|\boldsymbol{\alpha}^Y,\boldsymbol{X},\boldsymbol{Y} \dot{\sim} t_{n+\nu}(\beta_{n\boldsymbol{\alpha}^Y},\sigma^2_{n\boldsymbol{\alpha}^Y}).$$

REFERENCES

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika* **60,** 255–265.

Barbieri, M.M. and Berger, J.O. (2004). Optimal Predictive Model Selection. *The Annals of Statistics* **32,** 870–897.

Bernardo, J. M. and Smith, A. F. M. (2000). Bayesian Theory. *John Wiley & Sons,* England.

Chipman, H., George E. and McCulloch R. (2002). Bayesian Treed Models. *Machine Learning* **48,** 299–320.

Clyde, M. (2000). Model Uncertainty and Health Effects Studies for Particulate Matter. *Environmetrics* **11,** 745–763.

Crainiceanu, C. M., Dominici, F. , and Parmigiani, G. (2008). Adjustment Uncertainty in Effect Estimation. *Biometrika*, to appear.

Dominici, F., McDermott, A. and Hastie, T. (2004). Improved Semi-Parametric Time Series Models of Air Pollution and Mortality. *Journal of the American Statistical Association* **468,** 938–948.

Dominici, F., Peng, R.D. , Bell, M. , Pham, L. , McDermott, A. , Zeger, S.L., and Samet, J. M. (2006). Fine Particulates Air Pollution and Hospital Admission for Cardiovascular and Respiratory Diseases. *Journal of the American Medical Association* **295,** 1127–1135.

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association* **85,** 398–409.

Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6,** 721–741.

Hoeting, J., Madigan, D., Raftery, A., and Volinsky, C. (1999). Bayesian model averaging: a tutorial. *Statistical Science* **14,** 382–417.

Janes, H., Dominici, F. and Zeger, S.L. (2007). Trends in Air Pollution and Mortality: An Approach to the Assessment of Unmeasured Confounding. *Epidemiology* **18(4),** 416–423.

Kass, R.E. and Raftery, A.E. (1995). Bayes factors. *Journal of the American Statistical Association* **90,** 773–795.

Koop, G. and Tole, L. (2004). Measuring the health effects of air pollution: to what extent can we really say taht people are dying of bad air. *Journal of Environmental Economics and Management* **47,** 30–54.

Madigan, D. and York, J. (1995). Bayesian graphical models for discrete data. *International Statistical Review* **63,** 215–232.

Peng, R.D., Dominici, F. and Louis, T. (2006). Model Choice in Multi-Site Time Series Studies of Air Pollution and Mortality. *Journal of the Royal Statistical Society Series A (with discussion)* **169(2),** 179–203.

Raftery, A.E. (1995). Bayesian model selection in social research with discussions. *Sociological Methodology*, Marsden, P.V., (Ed.).

Raftery, A.E., Madigan, D., and Hoeting, J.A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* **92,** 179–191.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6,** 461–464.

Yeung, K., Bumgarner, R. , and Raftery, A.E. (2005). Bayesian Model Averaging: development of improved multi-class, gene selection and classification tool for microarray data. *Bioinformatics* **21,** 2394–2402.

Department of Biostatistics
Johns Hopkins University
615 N.Wolfe Street
Baltimore, MD 21205 USA

E-mail: chwang@jhsph.edu
gp@jhu.edu
ccrainic@jhsph.edu
fdominic@jhsph.edu
URL: http://www.biostat.jhsph.edu