Harvard University
Harvard University Biostatistics Working Paper Series

Year 2011 *Paper* 127

Bayesian Effect Estimation Accounting for Adjustment Uncertainty

Chi Wang* Giovanni Parmigiani†
Francesca Dominici‡

*Univ. of Kentucky, chi.wang@uky.edu
†Dana-Farber Cancer Institute and Harvard School of Public Health, gp@jimmy.harvard.edu
‡Harvard School of Public Health, fdominic@hsph.harvard.edu

Bayesian Effect Estimation Accounting for Adjustment Uncertainty

Chi Wang,[1,2,*] Giovanni Parmigiani,[3,4] and Francesca Dominici[4]
[1]Markey Cancer Center, University of Kentucky, Lexington, KY 40536, USA
[2]Department of Biostatistics, College of Public Health, University of Kentucky, Lexington, KY 40536, USA
[3]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02115, USA
[4]Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA
*email: chi.wang@uky.edu

April 5, 2011

Abstract

Model-based estimation of the effect of an exposure on an outcome is generally sensitive to the choice of which confounding factors are included in the model. We propose a new approach, which we call Bayesian Adjustment for Confounding (BAC), to estimate the effect on the outcome associated with an exposure of interest while accounting for the uncertainty in the confounding adjustment. Our approach is based on specifying two models: 1) the outcome as a function of the exposure and the potential confounders (the outcome model); and 2) the exposure as a function of the potential confounders (the exposure model). We consider Bayesian variable selection on both models and link the two by introducing a dependence parameter ω denoting the prior odds of including a predictor in the outcome model, given that the same predictor is in the exposure model. In the absence of dependence ($\omega = 1$), BAC reduces to traditional Bayesian Model Averaging (BMA). In simulation studies we show that BAC with $\omega > 1$ estimates the exposure effect with smaller bias than traditional BMA, and improved coverage. We then compare BAC, a recent 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 estimate 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.
Key words: Adjustment uncertainty; Bayesian model averaging; Exposure effects; Treatment effects.

1 Introduction

Estimating the effect of an exposure on an outcome, 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 statistical 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 variables should be included in the model to properly adjust for confounding.

It is possible to effectively convey this uncertainty by sensitivity analysis, 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). Bayesian Model Averaging (BMA) has been suggested as a more formal tool to account for model uncertainty. 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 indicators of whether each predictor is included in the model as unknown nuisance parameters. This results in a weighted average of predictions whose weights depend on the support that each selection receives from the data. This principled approach enjoys a number of desirable properties from a frequentist point of view as well, and has performed

competitively in out-of-sample prediction comparisons (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. Raftery (1995), Hoeting et al. (1999) suggested to estimate the exposure effect by a weighted average of model-specific effect estimates, again using the model's posterior probabilities as weights. Viallefont et al. (2001) applied this method to estimate an exposure's odds ratio in case-control studies. Other applications include air pollution research (Clyde, 2000; Koop and Tole, 2004).

However, while effective in some cases, traditional implementations of BMA can face severe limitations in effect estimation. Most of these can be traced to the fundamental difficulty arising with the fact that regression coefficients may have a different interpretation across models, a fact only recently being introduced explicitly in the specification of prior distributions (Consonni and Veronese, 2008). Crainiceanu et al. (2008) noted that model uncertainty methods useful in prediction may not generally perform well in adjustment uncertainty. They introduced a two-step approach (CDP) to estimate an exposure effect accounting for adjustment uncertainty. In the first step, it regresses exposure on a large set of potential confounders and selects confounders that are associated with exposure. In the second step, it regresses outcome on exposure, after including the confounders identified in the first step. Compared to this approach, traditional BMA with vague priors on the model space did not perform well. This is because the posterior model probabilities used to weight the model-specific estimates of the exposure effect might not reflect the model's ability to estimate the exposure effect, properly adjusting for confounding. For example, it can be that large weights are assigned to models that do not adequately adjust for confounders, leading to a biased estimate of the exposure effect. This problem may become more serious when limited prior information is available on the effect of interest.

Here we develop a novel Bayesian approach to adjustment uncertainty, which we call "Bayesian Adjustment for Confounding" (BAC). We consider the selection of confounders as a random variable, as in BMA, while overcoming the pitfalls described above. Our method makes explicit allowance for the fact that the interpretation of the effects can vary across models. BAC addresses this by explicitly focusing on models which are fully adjusted for confounding. Our technique generalizes BMA to simultaneous modeling of the exposure and the outcome. Our approach is based on specifying two models: 1) the outcome as a function of the exposure and the potential confounders (the outcome model); and 2) the exposure as a function of the potential confounders (the exposure model). The key to our approach is the specification of prior distribution such that, conditional on a predictor's inclusion in the model having exposure as the dependent variable, the same predictor should also have a higher probability to be included in the outcome model. We characterize this prior information by specifying a dependent parameter ω denoting the odds of including a predictor in the outcome model given that the same predictor is in the exposure model. This leads to a model-weighting strategy for effect estimation accounting for adjustment uncertainty. This strategy assigns high weights to models that are likely to include all the necessary confounders. Our method is explicitly designed to provide competitive results even without strong prior information on the magnitude of the effect.

While we do not take a causal inference perspective, our method has points of contacts with causal inference methodologies that are based on joint modeling of exposure and outcome as functions of confounders (Rosenbaum and Rubin, 1983; Robins et al., 1992) and with their Bayesian counterparts (McCandless et al., 2009). This literature strongly emphasizes, as we do, the critical role of model specification and the need for robustness to the choice of confounders (Rubin, 1997; Bang and Robins, 2005; Greenland, 2008). From this perspective, our methodology achieves a combination of three desirable properties: effect estimation efficiency, via the exposure model; variable selection robustness, achieved by allowing the selection to be a random variable; and bias reduction, achieved by including prior information to favor predictors of exposure in the selection of variables for the outcome model.

2 Bayesian Adjustment for Confounding (BAC)

2.1 Models

We build a model for estimating the effect of exposure, or treatment, X on outcome Y. We also have information on a set of M potential confounders $\boldsymbol{U} = \{U_1, \ldots, U_M\}$ identified because they are likely to affect Y, though their effects could be weak. A priori, there may be uncertainty about whether potential confounders should be adjusted for in effect estimation.

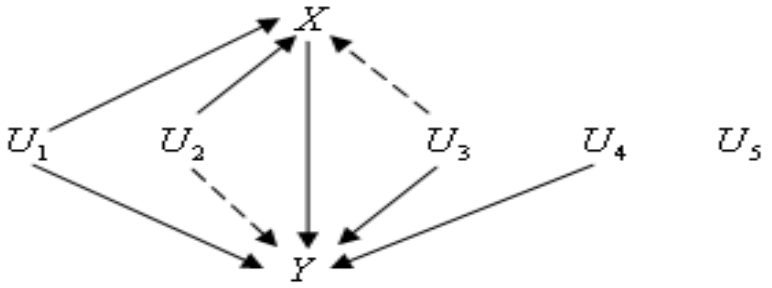

Figure 1: An illustrative example. Solid arrows indicate strong correlation, and dashed arrows indicate weak correlation.

Though many of our ideas are more general, we discuss our approach in the context of simultaneous linear regression models with two equations: one for exposure and one for outcome. In each equation, potential confounders are either included or excluded, depending on unknown vectors of indicators $\boldsymbol{\alpha}^X \in \{0,1\}^M$ and $\boldsymbol{\alpha}^Y \in \{0,1\}^M$. Here $\alpha_m^X = 1$ (or $\alpha_m^Y = 1$) whenever U_m is included in the exposure (or outcome) model. For brevity, we refer to the parameters $\boldsymbol{\alpha}$s as "models". Conditional on unknown parameters (indicated by Greek letters), and confounders, the regression equations for exposure X_i and outcome Y_i are

$$E\{X_i\} = \sum_{m=1}^{M} \alpha_m^X \delta_m^{\boldsymbol{\alpha}^X} U_{im} \quad (1)$$

$$E\{Y_i|X_i\} = \beta^{\boldsymbol{\alpha}^Y} X_i + \sum_{m=1}^{M} \alpha_m^Y \delta_m^{\boldsymbol{\alpha}^Y} U_{im} \quad (2)$$

where i indexes the sampling unit. For regression coefficients, β and δ, we use a notation that explicitly keeps track of the fact that those coefficients differ in meaning with the $\boldsymbol{\alpha}$s. This is especially important when one attempts to make inferences that involve estimates of the exposure effect obtained using different models. Intercept columns can be included among the Us. Some α_m^Ys can be set to one if confounders are deemed required.

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 variable 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 prediction, where the predicted quantities typically maintain the same interpretation across models.

Thus, when studying confounding adjustment, it is useful to consider the smallest outcome model that includes all the necessary confounders. We denote it by $\boldsymbol{\alpha}_*^Y$, and refer to it as the minimal model. The estimand of interest —the true effect of X on Y, is the coefficient of X in this model, or $\beta_* = \beta^{\boldsymbol{\alpha}_*^Y}$. If there are interactions between exposure and confounders, the estimands are model coefficients of both the main effect and the interaction terms. Without loss of generality, we will focus on the situation where there are no interaction terms. Our goal is estimation of β_* when $\boldsymbol{\alpha}_*^Y$ is unknown. A key observation is that all models that contain at least as many confounders as the minimal model will provide estimates of the exposure effect that are also interpretable as estimates of β_*. On the other hand, a model that does not include the minimal model, that is, a model that excludes at least one true confounder, will provide estimates of a parameter that is not the estimand of interest.

2.2 A basic illustration

It is useful to illustrate our approach using a simple example. Consider the situation depicted in Figure 1: U_1 is strongly correlated with both exposure and outcome; U_2 is strongly correlated with exposure, but weakly with outcome; U_3 is strongly correlated with outcome and weakly correlated with exposure; U_4 is strongly correlated with outcome and uncorrelated with exposure; and lastly U_5 is uncorrelated with both.

In this example, U_1, U_2 and U_3 are the true confounders of the effect of X on Y and the minimal model that can provide a correctly adjusted effect is $\boldsymbol{\alpha}_*^Y = (1,1,1,0,0)$. The true model is $\boldsymbol{\alpha}^Y = (1,1,1,1,0)$: this "includes" $\boldsymbol{\alpha}_*^Y$, that is it includes all the variables in $\boldsymbol{\alpha}_*^Y$. In addition, the true model also includes U_4. Since U_4 is not correlated with X, the interpretation of $\beta^{\boldsymbol{\alpha}^Y}$ is the same as that of $\beta^{\boldsymbol{\alpha}_*^Y}$. Therefore, the true model also allows for correct adjustment. Since U_4 is correlated with Y, including it can improve overall model fitting, which may yield smaller standard error of the X coefficient estimate. Thus, the true model may potentially lead to greater efficiency than the minimal model, though greater efficiency is not guaranteed in a finite sample. The full model $\boldsymbol{\alpha}^Y = (1,1,1,1,1)$ also contains $\boldsymbol{\alpha}_*^Y$ and a correctly defined coefficient. On the other hand, a model that does not include $\boldsymbol{\alpha}_*^Y$ will estimate a parameter that is not properly adjusted for confounding. For example, the model $\boldsymbol{\alpha}^Y = (1,0,1,1,0)$ will estimate a $\beta^{\boldsymbol{\alpha}^Y}$ that is not adjusted by U_2, which is an important confounder. Nonetheless, it may still be a useful model for prediction and may receive relatively strong support from the data.

Table 1: The correlation matrix of the simulated data set in Section 2.2

	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

To illustrate, we construct a simulated data set where the variables satisfy the relationships of Figure 1, using the correlations of Table 1, and regressions:

$$\begin{aligned} X_i &= \delta_1^X U_{i1} + \delta_2^X U_{i2} + \delta_3^X U_{i3} + \epsilon_i^X \\ Y_i &= \beta X_i + \delta_1^Y U_{i1} + \delta_2^Y U_{i2} + \delta_3^Y U_{i3} + \delta_4^Y U_{i4} + \epsilon_i^Y, \end{aligned} \tag{3}$$

where $i = 1, \ldots, 1000$, $\epsilon_i^X, \epsilon_i^Y$ are independent $N(0, \sigma_X^2)$ and $N(0, \sigma_Y^2)$ respectively, and the U_ms are independent $N(0, \sigma_U^2)$. We set $\delta^X = (1,1,0.1)$, $\delta^Y = (1,0.1,1,1)$, $\beta = 0.1$, and $\sigma_X^2 = \sigma_Y^2 = \sigma_U^2 = 1$. Using data so generated, we estimate β by maximum likelihood using two models: one is the true model and the other is the smaller model $\boldsymbol{\alpha}^Y = (1,0,1,1,0)$, which, unlike (3) does not include the true confounder U_2. Results are summarized in Table 2.

The BICs (Schwarz, 1978) for the true model and the smaller model are similar (2882.228 for true model and 2878.249 for smaller model), indicating that they fit the data comparably. The likelihood ratio test for the difference between them has p-value 0.087. However, the two models provide widely different estimates of β. The estimate from true model is 0.121 (95% confidence interval 0.059 to 0.183) while that from smaller model is 0.160 (95% confidence interval 0.116 to 0.204). In fact, the two estimates have different interpretations. In this case only the larger and true model provides an estimate of the exposure effect that is properly adjusted for confounding. This simple example illustrates that model selection approaches for adjustment uncertainty in effect estimation should be different from model selection approaches whose goal is prediction of the outcome. 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 accurately predict the outcome — which can often be achieved even by models that provide systematically biased estimates of the exposure effect.

2.3 Prior Distributions and Implementation of BAC

The importance of including in the outcome model all the potential confounders that belong to the minimal model suggests that an approach that acknowledges the fact that only a fraction of the models harbor the coefficient of interest with the correct interpretation, could be successful in addressing adjustment uncertainty

Table 2: Comparison of model posteriors from BMA, BAC and TBAC. The estimate of β from BMA is 0.157 with 95% credible interval (0.105, 0.203), that from BAC is 0.121 with 95% credible interval (0.059, 0.182), that from TBAC is 0.121 with 95% credible interval (0.059, 0.183). BMA is implemented forcing the exposure to always be in the model (FBMA). The dependence parameters ω in both BAC and TBAC are set to ∞.

Model	$\widehat{\beta}$	95% Confidence Interval	BIC	BMA weight	BAC weight	TBAC weight
(1,1,1,1,0) (true model (3))	0.121	(0.059, 0.183)	2882.228	0.060	0.985	0.970
(1,0,1,1,0)	0.160	(0.116, 0.204)	2878.249	0.927	0.000	0.000
(1,1,1,1,1)	0.122	(0.060, 0.184)	2888.834	0.001	0.015	0.030
(1,0,1,1,1)	0.160	(0.116, 0.204)	2884.771	0.012	0.000	0.000
(1,1,1,0,0)	0.096	(0.009, 0.183)	3545.253	0.000	0.000	0.000

Note: the weight in each of the three methods is defined as $P(\boldsymbol{\alpha}^Y|D)$, the posterior of $\boldsymbol{\alpha}^Y$. This posterior is calculated differently in each method. The posterior from BMA is calculated using a uniform prior on $\boldsymbol{\alpha}^Y$; that from BAC is calculated from the marginal of $P(\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y|D)$ where the prior of $P(\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y)$ is defined in equation (5); that from TBAC is calculated by using $P(\boldsymbol{\alpha}^Y|\boldsymbol{X})$ defined in equation (7) as the prior on $\boldsymbol{\alpha}^Y$.

from a Bayesian standpoint. We propose to pursue this idea via a novel approach called Bayesian Adjustment for Confounding (BAC) that considers jointly the exposure and outcome models, as in equations (1) and (2), and includes unknown model selection parameters $\boldsymbol{\alpha}^X$ and $\boldsymbol{\alpha}^Y$. We specify a prior distribution on $\boldsymbol{\alpha}^Y|\boldsymbol{\alpha}^X$ such that

$$\frac{P(\alpha_m^Y = 1|\alpha_m^X = 1)}{P(\alpha_m^Y = 0|\alpha_m^X = 1)} = \omega, \quad \frac{P(\alpha_m^Y = 1|\alpha_m^X = 0)}{P(\alpha_m^Y = 0|\alpha_m^X = 0)} = 1, \quad m = 1, \ldots, M, \tag{4}$$

where $\omega \in [1, \infty]$ is a dependence parameter denoting the prior odds of including U_m into the outcome model when U_m is included in the exposure model. When $\omega = \infty$, the first equation in (4) becomes $P(\alpha_m^Y = 1|\alpha_m^X = 1) = 1$, and requires that any U_m for which $\alpha_m^X = 1$ is automatically included in the outcome model. When $1 < \omega < \infty$, our prior on $\boldsymbol{\alpha}^Y|\boldsymbol{\alpha}^X$ provides a chance to rule out the predictors that are only associated with X but not associated with Y. To account for the feedback effect of $\boldsymbol{\alpha}^Y$ on $\boldsymbol{\alpha}^X$, we also set

$$\frac{P(\alpha_m^X = 1|\alpha_m^Y = 0)}{P(\alpha_m^X = 0|\alpha_m^Y = 0)} = \frac{1}{\omega}, \quad \frac{P(\alpha_m^X = 1|\alpha_m^Y = 1)}{P(\alpha_m^X = 0|\alpha_m^Y = 1)} = 1,$$

to assign low probabilities for predictors not selected by outcome model to be included in exposure model. The joint prior of $(\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y)$ implied by these conditional specifications is

$$\begin{aligned} P(\alpha_m^X = 0, \alpha_m^Y = 0) = P(\alpha_m^X = 0, \alpha_m^Y = 1) = P(\alpha_m^X = 1, \alpha_m^Y = 1) = \omega/(3\omega + 1) \\ P(\alpha_m^X = 1, \alpha_m^Y = 0) = 1/(3\omega + 1). \end{aligned} \tag{5}$$

The conditional prior of $\boldsymbol{\alpha}^Y$ given $\boldsymbol{\alpha}^X$ in (4) plays a key role in approximating the marginal posterior distribution of the exposure coefficient under the minimal model, β_*,

$$P(\beta_*|D) = \sum_{\boldsymbol{\alpha}^Y} P(\beta_*|\boldsymbol{\alpha}^Y, D) P(\boldsymbol{\alpha}^Y|D),$$

where $D = (\boldsymbol{X}, \boldsymbol{Y})$ contains vectors of observed data for X and Y. Our analysis is also conditional on observed data for potential confounders $\boldsymbol{U}$, and they will not be noted in posteriors for simplicity of notation. When ω is large, the conditional prior in (4) greatly increases the chance for predictors strongly correlated with X to be included in the outcome model. These predictors are confounders if they are also correlated with Y. Therefore, the prior leads to a posterior distribution of $\boldsymbol{\alpha}^Y$ ($P(\boldsymbol{\alpha}^Y|D)$) that assigns mass mostly to models which are fully adjusted for confounding, that is, models containing the minimal model. For these models, $\beta^{\boldsymbol{\alpha}^Y} = \beta_*$ so that $P(\beta_*|\boldsymbol{\alpha}^Y, D) = P(\beta^{\boldsymbol{\alpha}^Y}|\boldsymbol{\alpha}^Y, D)$. Therefore, approximately,

$$P(\beta_*|D) \doteq \sum_{\boldsymbol{\alpha}^Y} P(\beta^{\boldsymbol{\alpha}^Y}|\boldsymbol{\alpha}^Y, D) P(\boldsymbol{\alpha}^Y|D), \tag{6}$$

where $P(\beta^{\boldsymbol{\alpha}^Y}|\boldsymbol{\alpha}^Y, D)$ can be directly estimated from observed data. This approximation will be further discussed in Section 3.

Our goal is to calculate the posterior distribution of the parameters of interest $(\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y, \beta_*)$ in equations (1) and (2). In our implementation, we assume the following priors for model parameters: $\boldsymbol{\delta}^{\boldsymbol{\alpha}^X}|(\boldsymbol{\alpha}^X, \tau_X) \sim N(\boldsymbol{\mu}_{0\boldsymbol{\alpha}^X}, (\tau_X)^{-1}\phi^2\boldsymbol{\Sigma}_{0\boldsymbol{\alpha}^X})$, $(\beta^{\boldsymbol{\alpha}^Y}, \boldsymbol{\delta}^{\boldsymbol{\alpha}^Y})|(\boldsymbol{\alpha}^Y, \tau_Y) \sim N(\boldsymbol{\mu}_{0\boldsymbol{\alpha}^Y}, (\tau_Y)^{-1}\phi^2\boldsymbol{\Sigma}_{0\boldsymbol{\alpha}^Y})$, $\tau_X, \tau_Y \sim$ Gamma$(\nu/2, \nu\lambda/2)$, where ν, λ, ϕ, the M-vector $\boldsymbol{\mu}_{0\boldsymbol{\alpha}^X}$, the $(M+1)$-vector $\boldsymbol{\mu}_{0\boldsymbol{\alpha}^Y}$, the $M \times M$-matrix $\boldsymbol{\Sigma}_{0\boldsymbol{\alpha}^X}$ and the $(M+1) \times (M+1)$-matrix $\boldsymbol{\Sigma}_{0\boldsymbol{\alpha}^Y}$ are hyperparameters that are selected as in Raftery et al. (1997). To implement the Markov chain Monte Carlo (MCMC) algorithm, we make the following assumptions:

A1: $(\beta^{\boldsymbol{\alpha}^Y}, X)$ are independent of $\boldsymbol{\alpha}^Y$ given $(\boldsymbol{\alpha}^X, \widetilde{Y})$, where $\widetilde{Y} = Y - \beta^{\boldsymbol{\alpha}^Y}X$.

A2: X is independent of $\boldsymbol{\alpha}^Y$ given $\boldsymbol{\alpha}^X$.

A3: $(\beta^{\boldsymbol{\alpha}^Y}, Y)$ are independent of $\boldsymbol{\alpha}^X$ given $(\boldsymbol{\alpha}^Y, X)$

A4: $\widetilde{Y}$ is independent of $\boldsymbol{\alpha}^X$ given $\boldsymbol{\alpha}^Y$

The assumptions can be interpreted as follows. A1: Given a known $\widetilde{Y}$ and a known exposure model, the selection of the outcome model should no longer depend on the exposure and its effect on Y. A2: given that we know the covariates that are included in the exposure model (i.e. $\boldsymbol{\alpha}^X$), the outcome model should not provide additional information on X. The two remaining assumptions can be interpreted similarly, except that they are conditioning on the outcome model instead of the exposure model.

We use a MCMC algorithm to draw posterior samples of $(\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y, \beta^{\boldsymbol{\alpha}^Y})$ to approximate $P(\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y, \beta_*|D)$. These posterior samples are obtained by iteratively sampling from $P(\boldsymbol{\alpha}^X|\beta^{\boldsymbol{\alpha}^Y}, \boldsymbol{\alpha}^Y, D)$, $P(\boldsymbol{\alpha}^Y|\beta^{\boldsymbol{\alpha}^Y}, \boldsymbol{\alpha}^X, D)$ and $P(\beta^{\boldsymbol{\alpha}^Y}|\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y, D)$. Sampling from the first two full conditionals is achieved by the MC^3 method (Madigan and York, 1995). The derivation of these full conditionals is described in Appendix A.

2.4 Two-stage Bayesian Adjustment for Confounding (TBAC)

In this subsection, we consider a second approach which, when calculating the posterior distribution of $(\beta^{\boldsymbol{\alpha}^Y}, \boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y)$, cuts the feedback from $\boldsymbol{\alpha}^Y$ to $\boldsymbol{\alpha}^X$. This approach, called two-stage BAC (TBAC), treats the exposure and outcome models separately in two stages.

TBAC requires Assumption A2 as well as the following assumption:

A1': $\beta^{\boldsymbol{\alpha}^Y}$ is independent of $\boldsymbol{\alpha}^Y$ given $\widetilde{Y}$.

Assumption A1' is similar to Assumption A1 except that X is not taken into account since TBAC will treat X as fixed when considering the outcome model in its second stage.

In stage one of TBAC, we specify a uniform prior on $\boldsymbol{\alpha}^X$, a conditional prior on $\boldsymbol{\alpha}^Y|\boldsymbol{\alpha}^X$ as defined in equation (4) and use the exposure model only to calculate $P(\boldsymbol{\alpha}^X|\boldsymbol{X})$ and $P(\boldsymbol{\alpha}^Y|\boldsymbol{X})$. These two posterior distributions are calculated as follows:

$$\begin{aligned} P(\boldsymbol{\alpha}^X|\boldsymbol{X}) &\propto P(\boldsymbol{X}|\boldsymbol{\alpha}^X)P(\boldsymbol{\alpha}^X) \\ P(\boldsymbol{\alpha}^Y|\boldsymbol{X}) &= \sum_{\boldsymbol{\alpha}^X} P(\boldsymbol{\alpha}^Y|\boldsymbol{\alpha}^X, \boldsymbol{X})P(\boldsymbol{\alpha}^X|\boldsymbol{X}) \overset{\text{using A2}}{=\!=\!=} \sum_{\boldsymbol{\alpha}^X} P(\boldsymbol{\alpha}^Y|\boldsymbol{\alpha}^X)P(\boldsymbol{\alpha}^X|\boldsymbol{X}), \end{aligned} \tag{7}$$

where the expression of $P(\boldsymbol{X}|\boldsymbol{\alpha}^X)$ is given in Appendix A.

In stage two of TBAC, we use $P(\boldsymbol{\alpha}^Y|\boldsymbol{X})$ as prior on $\boldsymbol{\alpha}^Y$ and approximate $P(\boldsymbol{\alpha}^Y, \beta_*|D)$ by $P(\boldsymbol{\alpha}^Y, \beta^{\boldsymbol{\alpha}^Y}|D)$. We assume the same prior distributions for model parameters as in BAC and implement two separate MCMC algorithms for each of the two stages. Details on the sampling algorithms are described in Appendix A.

TBAC can be considered as a Bayesian model averaging method on the outcome model with an informative model prior $P(\boldsymbol{\alpha}^Y|\boldsymbol{X})$ obtained from stage one. This prior is the key difference between TBAC and traditional Bayesian model averaging, in which a flat uniform prior on the outcome model is typically assumed. In the following section, we will provide a detailed comparison between BAC/TBAC and BMA.

3 Relation to Bayesian Model Averaging (BMA)

In the context of effect estimation, several authors (Raftery, 1995; Hoeting et al., 1999) suggested to calculate the posterior distribution of the effect by taking an average over models, weighted by their posterior

probabilities:

$$\sum_{\boldsymbol{\alpha}^{\boldsymbol{Y}}} P(\beta^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}|\boldsymbol{\alpha}^{\boldsymbol{Y}}, \boldsymbol{Y}) P(\boldsymbol{\alpha}^{\boldsymbol{Y}}|\boldsymbol{Y}). \tag{8}$$

This corresponds to marginalization according to the law of total probabilities, but only if the parameters $\beta^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}$ have the same interpretation.

From the perspective of adjustment uncertainty, (8) can be decomposed into two parts: the sum over models that include the correct estimand, and the rest. That is

$$\sum_{\boldsymbol{\alpha}^Y \supseteq \boldsymbol{\alpha}^Y_*} P(\beta_*|\boldsymbol{\alpha}^Y, \boldsymbol{Y}) P(\boldsymbol{\alpha}^Y|\boldsymbol{Y}) + \sum_{\boldsymbol{\alpha}^Y \not\supseteq \boldsymbol{\alpha}^Y_*} P(\beta^{\boldsymbol{\alpha}^Y}|\boldsymbol{\alpha}^Y, \boldsymbol{Y}) P(\boldsymbol{\alpha}^Y|\boldsymbol{Y}). \tag{9}$$

where $\boldsymbol{\alpha} \supseteq \boldsymbol{\alpha}'$ indicates that model $\boldsymbol{\alpha}$ contains all the variables that are also contained in model $\boldsymbol{\alpha}'$. The second term of (9) averages across models that do not include $\boldsymbol{\alpha}^Y_*$, and therefore do not estimate the same effect.

In BMA one needs to be careful about not assigning large weights to the models in the second term of equation (9). A common practice in traditional implementations of BMA is to use uniform, or highly dispersed, priors on the $\boldsymbol{\alpha}^Y$s and often on the effect of interest as well. When the prior is the same for all models, the ratio of the weights given to models $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ is the Bayes Factor $P(\boldsymbol{Y}|\boldsymbol{\alpha}_1)/P(\boldsymbol{Y}|\boldsymbol{\alpha}_2)$ (Kass and Raftery, 1995) and the posterior model probabilities in BMA are driven by a model's predictive ability, which may differ from its ability to properly adjust for confounding in effect estimation.

To illustrate, the fifth column in Table 2 lists model weights used by BMA in the simulated data set in Section 2.2. We assigned a uniform prior to all models. Most of the weight (92.7%) is assigned to model $(1, 0, 1, 1, 0)$, which does not include all requisite confounders, and estimates the effect at 0.160 (95% confidence interval 0.116 to 0.204). In contrast, only 6.0% of the weight is assigned to the true model (3) which estimates the correct β_*. Thus, the BMA estimate of β (which is equal to 0.157) is severely biased and its associated 95% credible interval (0.105, 0.203) does not cover the true value of 0.1. We repeated the simulation 1,000 times. The coverage rate for the 95% credible interval is only 0.79.

BAC and TBAC are constructed using the same general principles as BMA, but, in our view, offer a far more appropriate prior for the model $\boldsymbol{\alpha}^Y$. The conditional prior $P(\boldsymbol{\alpha}^Y|\boldsymbol{\alpha}^X)$ defined in equation (4) includes BMA as a special case of $\omega = 1$, where a flat uniform prior is assigned to $\boldsymbol{\alpha}^Y$. But when ω is larger than one, the prior of $\boldsymbol{\alpha}^Y|\boldsymbol{\alpha}^X$ is informative and incorporates information on which Us are good predictors of X. TBAC exploits the exposure model to identify confounders highly correlated with X. Some of these confounders, if weakly correlated with Y, may not be identified by the outcome model alone. BAC shares the same property as TBAC, and additionally uses a full Bayesian approach in its implementation, that includes feedback from the outcome model to the exposure model. Therefore, compared to BMA, BAC and TBAC attempt to place most of the posterior weights $P(\boldsymbol{\alpha}^Y|D)$ on the first term in equation (9) and away from the second. To illustrate, Table 2 lists the model posterior weights based on BAC: 98.5% of the weight is assigned to the true model, compared to only 6.0% assigned to the same model as the one selected by BMA. No weight is assigned to models not nesting the minimal model, compared to 93.9% in total assigned by BMA. This result illustrates that linking the two variable selection problems can assign large weights to models including the minimal model, in cases when BMA can fail to do so. This is also the heuristic behind approximation (6).

4 Simulations

In this section, we conduct simulation studies to illustrate and compare the practical properties of BAC, TBAC, CDP (a two-step frequentist approach accounting for adjustment uncertainty by Crainiceanu et al. (2008)), traditional BMA (Raftery, 1995; Hoeting et al., 1999), and standard stepwise selection (Mickey and Greenland, 1989). We consider two simulation scenarios. The first shows that BMA can provide a very biased estimate of the exposure effect even under a very simple setting with only two confounders in the true model. In contrast, BAC can fully adjust for confounding and provide unbiased parameter estimates. The second shows similar results in a more complex setting.

In our simulations, we consider both BAC and TBAC with $\omega = 2, 4, 10$ or ∞. For BMA, we consider two different implementations: the first is forcing the exposure to always be in the model (FBMA), while the second (NBMA) is not. For the stepwise method, the threshold for adding a variable into the model is taken as 0.20, and the threshold for removing a variable is taken as 0.05 (Mickey and Greenland, 1989).

Our first scenario is similar to the one in Crainiceanu et al. (2008) and considers the true model: $Y_i = \beta X_i + \delta_1^Y U_{1i} + \delta_2^Y U_{2i} + \epsilon_i^Y$, where $i = 1, \ldots, 1000$, and ϵ_i^Y are independent $N(0,1)$. (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 $\beta = \delta_1^Y = \delta_2^Y = 0.1$. We generated 500 data sets. For each, we calculated the Maximum Likelihood Estimate (MLE) of β from the true model and compared it with the estimates from six estimation methods: BAC, TBAC, CDP, FBMA, NBMA and stepwise selection. The results are summarized in Table 3.

Table 3: Comparison of estimates of β from six methods, along with the gold standard (true model) in the first simulation scenario. BIAS is the difference between the mean of estimates of β and the true value, SEE is the mean of standard error estimates, SSE is the standard error of the estimates of β, MSE is the mean square error, and CP is the coverage probability of the 95% confidence interval or credible interval.

Method		BIAS	SEE	SSE	MSE	CP
True model		0.000	0.044	0.044	0.002	0.95
BAC	$\omega = \infty$	0.000	0.044	0.044	0.002	0.94
	$\omega = 10$	0.018	0.047	0.050	0.003	0.91
	$\omega = 4$	0.027	0.046	0.052	0.003	0.87
	$\omega = 2$	0.034	0.045	0.052	0.004	0.84
TBAC	$\omega = \infty$	0.000	0.044	0.044	0.002	0.95
	$\omega = 10$	0.018	0.047	0.050	0.003	0.92
	$\omega = 4$	0.026	0.046	0.051	0.003	0.89
	$\omega = 2$	0.034	0.045	0.051	0.004	0.84
CDP		0.000	0.044	0.045	0.002	0.95
FBMA		0.041	0.044	0.051	0.004	0.78
NBMA		−0.009	0.050	0.074	0.006	0.72
Stepwise		0.019	0.039	0.058	0.004	0.72

Unless noted, BAC and TBAC will refer to the special case of $\omega = \infty$ in the rest of this section. BAC, TBAC and CDP produce very similar estimates, both close to the estimates obtained from the true model. All these methods have point estimates around 0.1, the true value of β. Their MSEs are also similar to each other. In contrast, the mean of point estimates based on FBMA are much larger than 0.1, indicating that FBMA systematically overestimates the exposure effect in this example. The MSE of FBMA is also higher. The mean of point estimates based on NBMA is 0.091, which is close to the means from BAC and TBAC. Despite this good average behavior, NBMA produces the worst results. The MSE of NBMA is 0.006, which is much higher than 0.002 for BAC and TBAC. The distribution of the point estimates from NBMA reveals why NBMA has small bias and large MSE: 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. The point estimates based on the stepwise method are systematically larger than 0.1. The MSE is higher than that of the true model.

The difference between BAC, TBAC and CDP on one side, and BMA and stepwise approaches on the other is even more pronounced when comparing confidence intervals or credible intervals (both referred to as CI for brevity). The coverage probabilities of 95% CIs based on BAC, TBAC and CDP are close to 0.95, the desired value. In contrast, the coverage probabilities of FBMA and NMBA are only 0.78 and 0.72 respectively.

It is interesting to investigate the impact of the dependence parameter ω on confounding adjustment in BAC and TBAC. As ω decreases from ∞ to 2, the connection between exposure model and outcome model

becomes weaker. The estimates, therefore, become closer to those from BMA. The biases increase from 0.000 to 0.034, the MSEs increase from 0.002 to 0.004, and the coverage probabilities drop from 0.94 to 0.84 in BAC and 0.95 to 0.84 in TBAC. The results show that ω controls the degrees of confounding adjustment, with $\omega = \infty$ providing the fullest adjustment in this scenario.

Our second simulation scenario considers a larger number of potential confounders that are correlated with the exposure and also with the outcome. We consider both variables that are strongly and weakly correlated with exposure, and assume the following true outcome model: $Y_i = \beta X_i + \delta_1^Y U_{1i} + \ldots + \delta_{14}^Y U_{14i} + \epsilon_i^Y$, where $i = 1, \ldots, 1000$, ϵ_i^Y are independent $N(0,1)$, and $(X_i, U_{1i}, \ldots, U_{7i})$ are independent normal vectors with mean zero and a covariance matrix, $\Sigma = (\sigma_{kl})_{8\times 8}$, where $\sigma_{kl} = 1$ if $k = l$ or $\sigma_{kl} = \rho^{k+l-2}$ if $k \neq l$, $1 \leq k, l \leq 8$. We also assumes that the rest of the confounders $U_{8i}, \ldots, U_{14i}$ independently follow $N(0,1)$ distribution and are independent of the other confounders in the true outcome model. The set of potential confounders $\boldsymbol{U}$ includes $U_1, \ldots, U_{14}$ 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 0.1, $\delta_1 = \ldots = \delta_{14} = 0.1$ and $\rho = 0.7$. Similarly to the first scenario, we generated 500 data sets. For each simulated data set, we calculated the MLE of β from the known true model and compared it to the estimates from the six methods: BAC, TBAC, CDP, FBMA, NBMA, and stepwise. The results are summarized in Table 4.

Table 4: Comparison of estimates of β from six methods, along with the gold standard (true model) in the second simulation scenario. For BAC and TBAC, ϕ is set to 2.85. For FBMA, several different ϕs are considered.

Method		BIAS	SEE	SSE	MSE	CP
True model		0.000	0.051	0.049	0.002	0.96
BAC	$\omega = \infty$	0.009	0.051	0.050	0.003	0.96
	$\omega = 10$	0.045	0.055	0.058	0.005	0.84
	$\omega = 4$	0.064	0.055	0.061	0.008	0.75
	$\omega = 2$	0.080	0.055	0.062	0.010	0.64
TBAC	$\omega = \infty$	0.006	0.051	0.050	0.003	0.97
	$\omega = 10$	0.043	0.055	0.058	0.005	0.85
	$\omega = 4$	0.062	0.055	0.060	0.007	0.76
	$\omega = 2$	0.078	0.055	0.061	0.010	0.66
CDP		0.000	0.051	0.048	0.002	0.97
FBMA	$\phi = 2.85$	0.097	0.054	0.061	0.013	0.55
	$\phi = 1.05$	0.070	0.055	0.060	0.009	0.70
	$\phi = 0.30$	0.039	0.053	0.055	0.005	0.87
	$\phi = 0.10$	0.019	0.046	0.039	0.002	0.96
NBMA		0.056	0.064	0.096	0.012	0.63
Stepwise		0.044	0.043	0.067	0.006	0.66

The differences we noted between BAC, TBAC and CDP on one side, and BMA and stepwise on the other, are now even more pronounced in this more complex example. The point estimate obtained using FBMA is biased and larger than the point estimate based on the true model. The coverage probabilities of 95% CIs are only 0.55 and 0.63 for FBMA and NBMA respectively. The point estimate using the stepwise method is also biased. The coverage probability is only 0.66. In contrast, the point estimates based on BAC, TBAC and CDP are close to those based on the true model, and the coverage probabilities are very close to the desired value. The choice of ω in the priors of BAC and TBAC has a pronounced effect on the estimates. When ω decreases from ∞ to 2, the coverage probability dropped from 0.96 to 0.64 in BAC and 0.97 to 0.66 in TBAC.

The performance of BMA depends strongly on the spread of prior. For the Normal-Gamma prior we considered, the spread can be controlled by hyperparameter ϕ. Following the recommendation by Raftery et al. (1997), we chose $\phi = 2.85$ for BAC, TBAC and BMA in all the examples of this paper. This prior

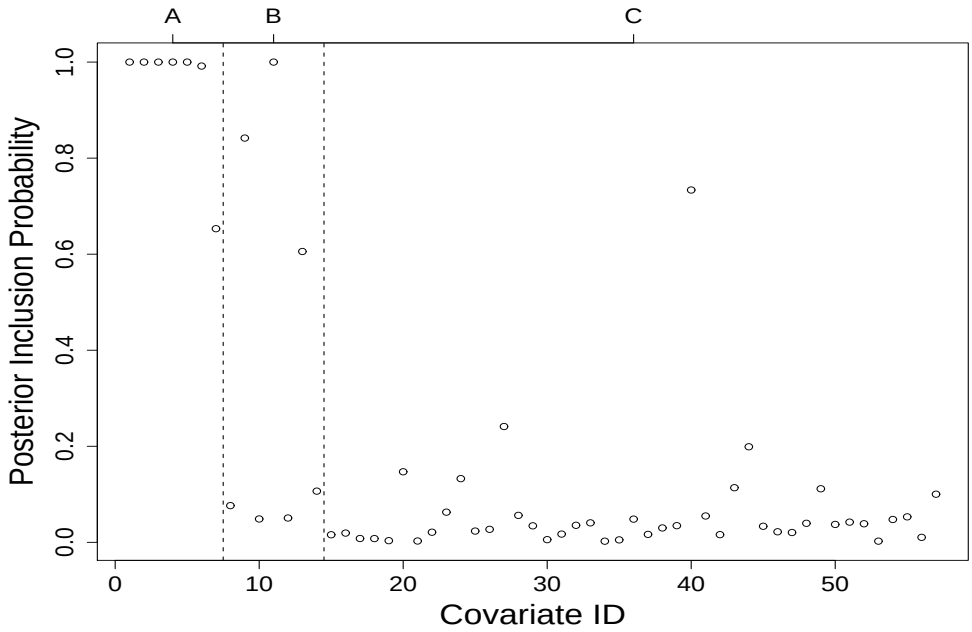

Figure 2: Posterior inclusion probability of potential confounders, separated into three groups by two vertical dashed lines. The first 7 (group A) are in the true model and are correlated with X, the next 7 (group B) are in the true model but are independent of X, the rest (group C) are not in the true model and are independent of X.

is quite spread with 95% of the mass between -5.27 and 5.27. The FBMA estimate under this prior is significantly biased. But the performance of FBMA improves when a more concentrated prior with smaller ϕ is used. Table 4 lists the estimates of FBMA based on different values of ϕ. When $\phi = 0.1$, with 95% of the mass between -0.19 and 0.19, the FBMA estimates are as good as those based on BAC and TBAC. This suggests that strong prior information, concentrating in the region of the true value, is required for FBMA to have good performance. In contrast, BAC and TBAC provide reasonable estimates even under the most spread prior $\phi = 2.85$. This shows that strong prior information is not a requisite for Bayesian approaches for effect estimation as long as appropriate methods are applied.

We also computed the posterior inclusion probability (Barbieri and Berger, 2004) defined, for the mth confounder, as $p_m = \sum_{\boldsymbol{\alpha}^Y : \alpha_m^Y = 1} P(\boldsymbol{\alpha}^Y | D)$, which is estimated by the proportion of appearances of confounder m in the chain of outcome models. Figure 2 shows the estimated posterior inclusion probabilities for all the confounders, in a simulated data set from our second scenario, using TBAC. The first seven confounders 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.

4.1 Additional Simulations

In Appendix B, we describe simulations designed to evaluate and compare the performance of BAC priors with different ωs in the presence of predictors correlated with X but not with Y. These predictors are not confounders since they are not associated with Y given X. Including them in the outcome model will not help for confounding adjustment and may decrease the efficiency of effect estimation. We found that using $\omega = 10$ yields smaller MSE compared to $\omega = \infty$. This is because $\omega = 10$ gives a non-zero probability for a predictor included in the exposure model not to be included in the outcome model. In other words, this prior is able to exclude a predictor of X from the outcome model if that predictor is not correlated with Y. Therefore, in the presence of predictors only correlated with X but not with Y, a prior using a finite ω tends to have higher efficiency than $\omega = \infty$.

In Appendix C, we describe simulations designed to evaluate the performance of BAC and TBAC when the exposure model is misspecified. A disadvantage with BAC/TBAC is that it requires two models while BMA only requires one. However, in our context, this does not come necessarily with an increased risk of model misspecification. Our simulation results show that both BAC and TBAC are robust to misspecification of the exposure model. The key feature in confounding adjustment is to include a sufficient number of confounders. A roughly correct exposure model may often be enough to ensure that this happens.

In Appendix D, we describe simulations designed to compare BAC vs. TBAC when $\omega = \infty$. We found that the two methods behave similarly in the majority of the cases examined here. However, they show some differences when dealing with predictors weakly associated with both X and Y. Compared to TBAC, BAC assigns lower weights to models that include those predictors. As a result, the two methods give somewhat different posterior distributions of $\boldsymbol{\alpha}^X$ and $\boldsymbol{\alpha}^Y$. But since these predictors have limited impact on the estimation of the exposure effect, they still provide very similar exposure effect estimates.

In Appendix E, we provided simulation results to compare BAC/TBAC and BMA under the two simulation scenarios described in this section but with a smaller sample size of 100. We found that the MSE from BMA is smaller than that from BAC and TBAC in the first scenario, but is larger in the second scenario. The results indicate that, although BAC and TBAC in general perform better, BMA may sometimes yield smaller MSE when the sample size is small. Combining with results from Appendix B, we conclude that there is not a single value of ω that is uniformly optimal in terms of MSE. The choice of ω should depend on sample size, complexity of confounding structure, as well as the bias/variance trade off. And the prior with $\omega = \infty$ is usually conservative, which provides unbiased estimates.

5 Air Pollution Example

In air pollution epidemiology, adjusting for confounding bias is probably the biggest challenge when estimating a small health effect associated with exposure to an environmental contaminant. In addition, because of the heavy policy implications associated with the public health impact of air pollution, most of the epidemiological evidence has been severely challenged by the threat of confounding bias.

In this section, we apply the newly proposed methods (BAC, TBAC) to daily time series data for Nassau County NY for the period 1999-2005. Although this data analysis is mainly used as an illustration of our newly proposed approach, the results clearly illustrate the potential application and impact of BAC and TBAC in epidemiology studies of observational data. The data include 1,532 daily records of emergency hospital admissions, weather variables, and $PM_{2.5}$ levels. A more extensive description of this data set can be found in Dominici et al. (2006). The goal is to estimate the increase in the rate of hospitalizations for cardiovascular disease (CVD) associated with a $10\mu g/m^3$ increase in $PM_{2.5}$, while accounting for age-specific longer-term trends, weather and day of the week. The hospitalization rate is calculated separately for each age group ($\geq$ 75 or not) on each day. In our model, to control for longer-term trends due, for example, to changes in medical practice patterns, seasonality and influenza epidemics, we include smooth functions of calendar time. We also include a smooth function to allow seasonal variations to be different in the two age groups. To control for the weather effect, we include smooth functions of temperature and dew point. To start, we consider a full model that is large enough to include all the necessary confounders (Dominici et al., 2000, 2004; Peng et al., 2006):

$$\begin{aligned} Y_{at} = \beta\ \mathrm{PM}_{2.5t} + \ \mathrm{DOW}\ + \text{intercept for age group } a \\ + ns(\mathrm{Temp}_t, df_{\mathrm{Temp}}) + ns(\mathrm{Temp}_{t1-3}, df_{\mathrm{Temp}}) + ns(\mathrm{Dew}, df_{\mathrm{Dew}}) \\ + ns(\mathrm{Dew}_{t1-3}, df_{\mathrm{Dew}}) + ns(t, df_t) + ns(t, df_{at}) \times \text{age group}\ + \epsilon_t, \end{aligned}$$

where the outcome $Y_{at} = \sqrt{\text{CVD hospital admissions/size of population at risk}}$ for each age group a ($\geq$75 or not) on day t (=1, ..., 1532). $\mathrm{PM}_{2.5t}$ denotes the level of particulate matter having diameter less than 2.5 micrometer on day t. DOW are indicator variables for the day of the week. Temp_t and Temp_{t1-3} are the 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. The quantity $ns(., df)$ is a natural cubic spline with df degrees of freedom. We include $ns(t, df_t)$, $ns(\mathrm{Temp}_t, df_{\mathrm{Temp}})$, $ns(\mathrm{Temp}_{t1-3}, df_{\mathrm{Temp}})$, $ns(\mathrm{Dew}, df_{\mathrm{Dew}})$ and $ns(\mathrm{Dew}_{t1-3}, df_{\mathrm{Dew}})$ to adjust for the potential nonlinear confounding effects of seasonal variations,

temperature and dew point. The quantity $ns(t, df_{at}) \times$ age group is a natural cubic spline of t for the ≥ 75 age group to allow its seasonal variation to be different from the other age group. Similar to Crainiceanu et al. (2008), df_{Temp} is set to 12, df_{Dew} is set to 12, df_t is set to 16 per year, and df_{at} is set to 4. These degrees of freedom are considered as sufficiently large for the full model to include all the potential confounders (Crainiceanu et al., 2008). The residuals ϵ_t are assumed to be independent and identically distributed with a normal $N(0, \sigma^2)$ distribution. After dropping some potential confounders due to collinearity, we work with a set of 164 potential confounders.

We consider six approaches: BAC, TBAC, CDP, FBMA, NBMA, and stepwise. For BAC and TBAC, we consider priors with $\omega = 2, 4, 10$ or ∞. The estimated $PM_{2.5}$ effect ($\times 10,000$) denoted by $\widehat{\beta}$ is listed in Table 5: BAC, TBAC (with $\omega = \infty$) and CDP provide estimates of the short-term effect of $PM_{2.5}$ on CVD hospital admissions with 95% CIs that do not include 0. With $\omega = \infty$, both BAC and TBAC provide similar estimates of the exposure effect as CDP. Moreover, all of the three methods provide smaller standard errors than the one obtained under the full model. In comparison, FBMA and NBMA provide a very different and not statistically significant estimate of the exposure effect. Some confounders known to be important, such as temperature and dew point, are downweighted in BMA. Both temperature and dew point are positively correlated with $PM_{2.5}$ and negatively correlated with hospitalization rate. Failure to include them in the model diminishes the $PM_{2.5}$ effect. This illustrates that in practical applications BMA and BAC can lead to different conclusions. The key difference lies in the linking strength between the exposure model and the outcome model. As the strength decreases, which corresponds to smaller value of ω, the estimates from BAC and TBAC become closer to that from BMA.

Table 5: Comparison of estimates of $PM_{2.5}$ effect on CVD hospitalization rate based on BAC, TBAC, CDP, FBMA, NBMA, stepwise, and the full model.

Method		$\widehat{\beta}$	SE($\widehat{\beta}$)	95% CI
Full model		0.291	0.092	(0.110, 0.471)
BAC	$\omega = \infty$	0.226	0.081	(0.067, 0.385)
	$\omega = 10$	0.217	0.079	(0.060, 0.371)
	$\omega = 4$	0.186	0.085	(0.019, 0.351)
	$\omega = 2$	0.155	0.079	(0.007, 0.317)
TBAC	$\omega = \infty$	0.229	0.083	(0.071, 0.403)
	$\omega = 10$	0.216	0.075	(0.071, 0.367)
	$\omega = 4$	0.190	0.080	(0.035, 0.347)
	$\omega = 2$	0.155	0.077	(0.010, 0.313)
CDP		0.221	0.089	(0.045, 0.396)
FBMA		0.140	0.077	(−0.008, 0.298)
NBMA		0.007	0.033	(0.000, 0.131)
Stepwise		0.106	0.066	(−0.023, 0.234)

6 Discussion

Estimating an exposure effect, while accounting for the uncertainty in the adjustment for confounding, is of essential importance in observational studies. Building upon work by Dominici et al. (2004) and Crainiceanu et al. (2008), in this paper we develop Bayesian solutions to the estimation of the association between X and Y accounting for the uncertainty in the confounding adjustment. Given a set of potential confounders, we address simultaneously model selection for both the outcome and the exposure. While we discuss our methods in the setting of linear models, BAC and TBAC are general concepts and are not constrained to the linear case. For example they can be extended to generalized linear models using relatively well understood computational strategies.

As BMA, BAC and TBAC take a weighted average over models rather than making inference based on a single model. However they attempt to provide an estimate of the exposure effect by combining information

across regression models that include 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 BAC and TBAC 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. BAC and TBAC are more computationally intensive than BMA.

Successful application of BAC and TBAC rely on availability of all confounders. 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 correlated with X and Y, those correlated with Y only, as well as potentially others that are not associated with either, will provide unbiased estimates of the exposure effect. However, using the full model will generally yield wider confidence intervals compared to BAC and TBAC. By combining estimation from different smaller models, especially from models that only include requisite confounders but do not include many unnecessary variables, BAC and TBAC can provide more precise inference than the full model.

TBAC parallels CDP in its two-stage structure, and in the inclusion of variables selected from the exposure model into the outcome model. However, there are also important differences. TBAC provides a model-based solution rather than a partially algorithmic one, and also arguably considers uncertainty more fully in a Bayesian framework. BAC further takes into account the feedback effect and considers a full Bayesian approach. Also, in CDP, models are evaluated based on the change in deviance between sets of increasing dimensionality, a criterion that could lead to different conclusions compared to BAC and TBAC. Large spaces of confounders may potentially be required for CDP users to reliably observe the stabilization of the estimated effect that is required for the method to succeed. However, no restrictions on dimensionality apply to BAC and TBAC. Computationally, CDP is clearly faster, and also offers helpful visualizations. The two methods produce results with similar frequentist properties in our simulation studies.

In the propensity score literature, it is recommended to include variables that are strongly correlated with Y but only weakly correlated with X into the model for calculating the propensity score, as the bias resulting from their exclusion would dominate any loss of efficiency in modest or large studies (Rubin, 1997; Brookhart et al., 2006). One of the strengths of our method, shared by others such as doubly robust estimation (Scharfstein et al., 1999), is that we can identify these in a data-based way, rather than having to rely on prior knowledge as required in propensity score adjustment.

An alternative Bayesian variable selection approach is the Bayesian lasso (Park and Casella, 2008), assuming a mixture prior of a point mass at zero and a double exponential distribution for regression coefficients (Hans, 2010). An alternative version of both BAC and TBAC could be constructed using this prior instead. We expect that the use of the Bayesian lasso on the outcome model alone would present similar limitations to traditional BMA, but have not explored this in detail.

In summary, in this paper, we have introduced a well motivated and effective 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 traditional BMA implementations, 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.

Acknowledgements

We thank Ciprian Crainiceanu for his significant input into earlier versions of this manu-script, and Eric Tchetgen for a helpful discussion. We also thank the Co-Editor, the Associate Editor and two referees for insightful comments that have substantially improved the article. The work of Francesca Dominici was supported by grants EPA R83622, EPA RD83241701, NIH/NIEHS R01ES012054, NIH/NIEHS R01ES012044 and NIH/NCI P01CA134294 *Conflict of Interest:* None declared.

References

Bang, H. and Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61,** 962–973.

Barbieri, M. M. and Berger, J. O. (2004). Optimal predictive model selection. *The Annals of Statistics* **32,** 870–897.

Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., and Sturmer, T. (2006). Variable selection for propensity score models. *American Journal of Epidemiology* **163,** 1149–1156.

Chipman, H. A., George, E. I., and McCulloch, R. E. (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.

Consonni, G. and Veronese, P. (2008). Compatibility of prior specifications across linear models. *Statistical Science* **23,** 332–353.

Crainiceanu, C. M., Dominici, F., and Parmigiani, G. (2008). Adjustment uncertainty in effect estimation. *Biometrika* **95,** 635–651.

Dominici, F., McDermott, A., and Hastie, T. J. (2004). Improved semiparametric time series models of air pollution and mortality. *Journal of the American Statistical Association* **99,** 938–948.

Dominici, F., Peng, R. D., Bell, M., Pham, L., McDermott, A., Zeger, S. L., and Samet, J. M. (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *The Journal of the American Medical Association* **295,** 1127–1134.

Dominici, F., Samet, J. M., and Zeger, S. L. (2000). Combining evidence on air pollution and daily mortality from the twenty largest u.s. cities: A hierarchical modeling strategy (with discussion). *Journal of the Royal Statistical Society, Series A: Statistics in Society* **163,** 263–302.

Greenland, S. (2008). Variable selection versus shrinkage in the control of multiple confounders. *American Journal of Epidemiology* **167,** 523–529.

Hans, C. (2010). Model uncertainty and variable selection in bayesian lasso regression. *Statistics and Computing* **20,** 221–229.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: A tutorial (with discussion). *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,** 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 that 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.

McCandless, L. C., Gustafson, P., and Austin, P. C. (2009). Bayesian propensity score analysis for observational data. *Statistics in Medicine* **28,** 94–112.

Mickey, R. M. and Greenland, S. (1989). The impact of confounder selection criteria on effect estimation. *American Journal of Epidemiology* **129,** 125–137.

Park, T. and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association* **103,** 681–686.

Peng, R. D., Dominici, F., and Louis, T. A. (2006). Model choice in time series studies of air pollution and mortality. *Journal of the Royal Statistical Society, Series A: Statistics in Society* **169,** 179–203.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology* **25,** 111–163.

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.

Robins, J. M., Mark, S. D., and Newey, W. K. (1992). Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics* **48,** 479–495.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70,** 41–55.

Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine* **127,** 757–763.

Scharfstein, D. O., Rotnitzky, A., and Robins, J. M. (1999). Rejoinder to "Adjusting for nonignorable drop-out using semiparametric nonresponse models". *Journal of the American Statistical Association* **94,** 1135–46.

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

Viallefont, V., Raftery, A. E., and Richardson, S. (2001). Variable selection and Bayesian model averaging in case-control studies. *Statistics in Medicine* **20,** 3215–3230.

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