Denis Talbot*, Geneviève Lefebvre and Juli Atherton

The Bayesian Causal Effect Estimation Algorithm

DOI 10.1515/jci-2014-0035

Abstract: Estimating causal exposure effects in observational studies ideally requires the analyst to have a vast knowledge of the domain of application. Investigators often bypass difficulties related to the identification and selection of confounders through the use of fully adjusted outcome regression models. However, since such models likely contain more covariates than required, the variance of the regression coefficient for exposure may be unnecessarily large. Instead of using a fully adjusted model, model selection can be attempted. Most classical statistical model selection approaches, such as Bayesian model averaging, do not readily address causal effect estimation. We present a new model averaged approach to causal inference, Bayesian causal effect estimation (BCEE), which is motivated by the graphical framework for causal inference. BCEE aims to unbiasedly estimate the causal effect of a continuous exposure on a continuous outcome while being more efficient than a fully adjusted approach.

Keywords: model selection, causal diagrams, exposure effect estimation, variance reduction

1 Introduction

Estimating causal exposure effects in observational studies demands a vast knowledge of the domain of application. For instance, to estimate the causal effect of an exposure on an outcome, the graphical framework to causality usually involves postulating a causal graph to identify an appropriate set of confounding variables [1]. Specifying such a graph can be difficult, especially in subject areas where prior knowledge is scarce or limited.

Investigators often bypass difficulties related to the identification and selection of confounders through the use of fully adjusted outcome regression models. Such models express the outcome variable as a function of the exposure variable and all available potential confounding variables. A fully adjusted outcome regression model is commonly assumed to yield an unbiased estimator of the true effect of the exposure. However, since such models likely contain more covariates than required, the variance of the regression coefficient for exposure may be unnecessarily large. Instead of using a fully adjusted model, model selection can be attempted.

Most classical statistical model selection approaches do not readily address causal effect estimation. One such approach is Bayesian model averaging (BMA) [2, 3]. BMA averages quantities of interest (e.g. a regression coefficient or the value of a future observation) over all possible models under consideration: in the average, each estimate is weighted by the posterior probability attributed to the corresponding model. When the goal is prediction, BMA accounts for the uncertainty associated with model choice and produces confidence intervals that have adequate coverage probabilities [4]. Unfortunately, BMA can perform poorly when used to estimate a causal effect of exposure [5, 6].

Wang et al. [6] suggested two novel approaches that modify BMA to specifically target causal effect estimation: Bayesian adjustment for confounding (BAC) and two-stage Bayesian adjustment for confounding (TBAC). Graph-based simulations presented in Wang et al. [6] show that the causal effect estimators of BAC and TBAC are unbiased in a variety of scenarios, hence supporting their adequacy for causal inference.

***Corresponding author: Denis Talbot,** Département de médecine sociale et préventive, Université Laval, 1050 avenue de la Médecine, Pavillon Ferdinand-Vandry, Québec, Quebec G1V0A6, Canada; Département de mathématiques, Université du Québec à Montréal, 201 avenue du Président-Kennedy PK-5151, Montréal, Quebec H2X 3Y7, Canada,
E-mail: denis.talbot@fmed.ulaval.ca
Geneviève Lefebvre, Juli Atherton, Département de mathématiques, Université du Québec à Montréal, 201 avenue du Président-Kennedy PK-5151, Montréal, Quebec H2X 3Y7, Canada

A theoretical justification for the use of BAC for causal inference purposes is further discussed in Lefebvre, Atherton, and Talbot [7]. However, some simulations comparing BAC and TBAC to fully adjusted models show little difference in the variance of the causal effect estimators of each method [6, 8]. Moreover, the choice of BAC's hyperparameter ω has been recognized as challenging [9]. The value $\omega = \infty$ has been recommended if one seeks an unbiased causal exposure effect estimator [7]. Lefebvre et al. [7] proposed using cross-validation and bootstrap for selecting an ω value that aims to minimize the mean-square-error (MSE) of the BAC's causal effect of exposure estimator. These results suggest that the optimal ω value not only depends on the data-generating scenario, but also on sample size, thus making it very hard in practice to select an appropriate ω value.

In this paper we propose a new model averaging approach to causal inference: Bayesian causal effect estimation (BCEE). BCEE aims to unbiasedly estimate the causal effect of a continuous exposure on a continuous outcome, while being more efficient than a fully adjusted approach. With a sample of finite size, however, this is an ambitious objective. Hence, through a user-selected hyperparameter, BCEE enables an analyst to consider various degrees of trade-off between bias and variance for the estimator. While BCEE shares some similarities with TBAC, one distinctive feature of our approach is that its motivation lies in the graphical framework for causal inference (e.g. Pearl [1]).

The paper is structured as follows. In Section 2, we present the BCEE algorithm and discuss, in Section 3, a number of aspects of its practical implementation. We compare BCEE to some existing approaches for causal effect estimation in Section 4. In Section 5, we apply BCEE to a real dataset where we estimate the causal effect of mathematical perceived competence on the self-reported average in mathematics for highschool students in the province of Quebec. We conclude in Section 6 with a discussion of our results and provide suggestions for further research.

2 Bayesian causal effect estimation (BCEE)

Before presenting BCEE in Section 2.3, we first describe the modeling framework in Section 2.1 and provide a proposition and corollary concerning directed acyclic graphs (DAGs) in Section 2.2. The description of how the proposition and the corollary are used to develop BCEE is presented in Section 2.4. We conclude, in Section 2.5, with a toy example that sheds light on BCEE's properties. Note that although we refer to BCEE as a Bayesian algorithm, strictly speaking, it is approximately Bayesian since it requires specifying prior distributions only for a subset of the parameters. To simplify the discussion, we motivate BCEE from a frequentist perspective.

2.1 Modeling framework

We consider estimating the causal effect of a continuous exposure on a continuous outcome. Let X be the random exposure variable, Y be the random outcome variable and $\boldsymbol{U} = \{U_1, U_2, ..., U_M\}$ be a set of M available, pre-exposure, potentially confounding random covariates. Let i index the units of observations, $i = 1, ..., n$. Our goal is to estimate the causal effect of exposure using a linear regression model for the outcome with normal, independent and identically distributed errors. Assuming the set $\boldsymbol{U}$ is sufficient to identify the average causal effect and the model is correctly specified, a fully adjusted linear regression model can be used to estimate the causal effect. Under such assumptions, parameter β encodes the average causal effect of a unit increase in X on Y in the linear model

$$\mathbb{E}(Y_i|X_i, \boldsymbol{U_i}) = \delta_0 + \beta X_i + \sum_{m=1}^{M} \delta_m U_{im}, \tag{1}$$

where δ_0 is the intercept and δ_m is the regression coefficient associated with covariate U_m. A disadvantage to using a fully adjusted outcome model is that the variance of the exposure effect

estimator $\hat{\beta}$ can be large. Therefore, one might want to include a reduced number of covariates in the outcome model (1), that is, to adjust for a strict subset of $\boldsymbol{U}$ also sufficient to estimate the causal effect of X on Y.

Consider G an assumed causal directed acyclic graph (DAG) compatible with the distribution of the observed covariates in G, $\{Y, X, U\}$. Let $\boldsymbol{D} = \{D_1, D_2, ..., D_J\} \subset \boldsymbol{U}$ be the set of parents (direct causes) of X in G. Then using Pearl's back-door criterion [1], it is straightforward to show that adjusting for the set $\boldsymbol{D}$ is sufficient to avoid confounding. In other words, the parameter β in the linear model

$$\mathbb{E}(Y_i|X_i, \boldsymbol{D_i}) = \delta_0 + \beta X_i + \sum_{j=1}^{J} \delta_j D_{ij} \tag{2}$$

can also be interpreted as the average causal effect of X on Y. It can also be shown that outcome models adjusting for sets of pre-exposure covariates that *at least* include the direct causes of exposure are unbiased; BAC may be seen to be exploiting this feature [7]. Adjusting for the set of direct causes of X in the outcome model thus seems appealing since $\boldsymbol{D}$ is generally smaller than the full set $\boldsymbol{U}$. However, this approach can also yield an estimator of β, $\hat{\beta}$, whose variance is large unless those direct causes of X are also strong predictors of Y (e.g. Lefebvre et al. [7]).

BAC, TBAC and BCEE all rely on the fact that the set of direct causes of X is sufficient for estimating the causal effect and that this set of covariates can be identified from the data. A differentiating feature of BCEE is that it aims to disfavor outcome models that include one or more direct causes of X that are unnecessary to eliminate confounding. This is viewed as desirable since these variables generally increase the variance of $\hat{\beta}$. By doing so, BCEE targets sufficient models

$$\mathbb{E}(Y_i|X_i, \boldsymbol{Z_i}) = \delta_0 + \beta X_i + \sum_{k=1}^{K} \delta_k Z_{ik} \tag{3}$$

for which the variance of $\hat{\beta}$ is smaller than the variance of $\hat{\beta}$ in model (1) and the variance of $\hat{\beta}$ obtained using BAC or TBAC. In Section 2.2 we present a proposition and a corollary that underlie the functioning of BCEE.

2.2 A motivation based on directed acyclic graphs

The results presented in this section are based on Pearl's back-door criterion and are thus obtained from a graphical perspective to causality using directed acyclic graphs (DAGs). For a brief review of this framework, we refer the reader to the appendix of VanderWeele and Shpitser [10].

Proposition 2.1 presented below gives a sufficient condition to identify a set $\boldsymbol{Z}$ that yields an unbiased estimator $\hat{\beta}$ of the causal effect of X in eq. (3). Corollary 2.1 starts with such a sufficient set $\boldsymbol{Z}$ and provides conditions under which a direct cause of X included in $\boldsymbol{Z}$ can be excluded so that the resulting set $\boldsymbol{Z}'$ is also sufficient. Remark that this corollary is akin to Proposition 1 from VanderWeele and Shpitser [10]. In the sequel, the concept of d-separation is used to entail notions of conditional independence between variables. Moreover, the distribution-free adjustment defined in Pearl [1] relates to the adjustment in the linear model setting introduced in Section 2.1. For instance, see Chapter 5 from Pearl [1] and section 5.3.2 in particular.

Proposition 2.1 *Consider data compatible with a causal DAG G. Let* $\boldsymbol{D} = \{D_1, D_2, ..., D_J\}$ *be the set of direct causes of X and let* $\boldsymbol{Z}$ *be a set of covariates which we consider adjusting for. Adjusting for* $\mathbf{Z}$ *is sufficient to identify the average causal effect of X on Y if*

1) *no descendants of X are in* $\boldsymbol{Z}$ *and*
2) *if for each* $D_j \in \boldsymbol{D}$*, either*
   (a) $D_j \in \boldsymbol{Z}$ *or*
   (b) *if* $D_j \notin \boldsymbol{Z}$ *then Y and* D_j *are d-separated by* $\{X \cup \boldsymbol{Z}\}$.

Proof: see Appendix A.1.

Corollary 2.1 *Consider a $D_j \in \boldsymbol{Z}$ and let $\boldsymbol{Z}' = \boldsymbol{Z} \backslash D_j$.*

1) *If D_j and Y are d-separated by $\{X \cup \boldsymbol{Z}'\}$ then all back-door paths $X \leftarrow D_j \cdots \rightarrow Y$ are blocked by $\boldsymbol{Z}'$.*
2) *If in addition to 1., $\boldsymbol{Z}$ is sufficient to identify the average causal effect according to Proposition 2.1, then $\boldsymbol{Z}'$ is also sufficient to identify the average causal effect of X on Y.*

Proof: see Appendix A.2.

We now address how the proposition and the corollary are used in the linear regression setting presented in Section 2.1. First, Theorem 1.2.4 from Pearl [1] states the quasi-equivalence between d-separation and conditional independence. That is, unless a very precise tuning of parameters occurs, d-separation of Y and D_j by $\{X \cup \boldsymbol{Z}\}$ is equivalent to conditional independence between Y and D_j given $\{X \cup \boldsymbol{Z}\}$. Hence, we can replace d-separation by conditional independence in Proposition 2.1 and in Corollary 2.1. Under the assumption that all variables in the graph G are multivariate normal, we have that conditional independence is equivalent to zero partial correlation and thus to zero regression parameter in the linear model [11]. More specifically, if Y and D_j are conditionally independent given $\{X \cup \boldsymbol{Z}\}$, then the regression parameter associated to D_j in the linear regression of Y on D_j, X and $\boldsymbol{Z}$ is 0; and this parameter is 0 only if Y and D_j are conditionally independent given $\{X \cup \boldsymbol{Z}\}$. The assumption of multivariate normality is quite stringent; a weaker assumption is that model (1) is correctly specified (see Appendix B).

2.3 The BCEE algorithm

BCEE is viewed as a BMA procedure where the prior distribution of the outcome model is informative and constructed by using estimates from earlier steps of the algorithm, including the exposure model. In this section, we introduce BCEE and define the aforementioned prior distribution. The connections between Proposition 2.1, Corollary 2.1 and BCEE's prior distribution are discussed in Section 2.4.

We now define the outcome model using the same model averaging notation as in BAC and TBAC. Let $\boldsymbol{\alpha^Y} = (\alpha_1^Y, \ldots, \alpha_M^Y)$ be an M-dimensional vector for the inclusion of the covariates $\boldsymbol{U}$ in the outcome model, where component α_m^Y equals 1 if covariate U_m is included in the model and α_m^Y equals 0 if covariate U_m is not included, $m = 1, \ldots, M$. Letting i index the units of observation, $i = 1, \ldots, n$, the outcome model is the following normal linear model

$$Y_i = \delta_0^{\boldsymbol{\alpha^Y}} + \beta^{\boldsymbol{\alpha^Y}} X_i + \sum_{m=1}^{M} \alpha_m^Y \delta_m^{\boldsymbol{\alpha^Y}} U_{im} + \epsilon_i^{\boldsymbol{\alpha^Y}}, \tag{4}$$

where $\delta_m^{\boldsymbol{\alpha^Y}}$ and $\beta^{\boldsymbol{\alpha^Y}}$ denote respectively the unknown regression coefficients associated with U_m and X in the outcome model specified by $\boldsymbol{\alpha^Y}$. The parameter $\delta_0^{\boldsymbol{\alpha^Y}}$ denotes the unknown intercept in model $\boldsymbol{\alpha^Y}$ and the distribution of the error terms is given by $\epsilon_i^{\boldsymbol{\alpha^Y}} \overset{iid}{\sim} N(0, \sigma^2_{\boldsymbol{\alpha^Y}})$.

Given model (4) and a prior distribution $P(\boldsymbol{\alpha^Y})$, the use of BMA for the estimation of the exposure effect requires first obtaining the posterior distribution of the outcome model $P(\boldsymbol{\alpha^Y}|Y) \propto P(Y|\boldsymbol{\alpha^Y})P(\boldsymbol{\alpha^Y})$. Standard implementation of BMA often involves selecting a uniform prior distribution $P(\boldsymbol{\alpha^Y}) = 1/2^M \ \ \forall \ \ \boldsymbol{\alpha^Y}$, in which case $P(\boldsymbol{\alpha^Y}|Y) \propto P(Y|\boldsymbol{\alpha^Y})$. The model-averaged exposure effect is then given by

$$\mathbb{E}[\beta|Y] = \sum_{\boldsymbol{\alpha^Y}} \left[\int_{-\infty}^{\infty} \beta^{\boldsymbol{\alpha^Y}} P\left(\beta^{\boldsymbol{\alpha^Y}} | \boldsymbol{\alpha^Y}, Y\right) d\beta^{\boldsymbol{\alpha^Y}} \right] P\left(\boldsymbol{\alpha^Y}|Y\right). \tag{5}$$

In BCEE, we utilize an informative prior distribution rather than the usual non-informative one. This distribution aims to give the bulk of the prior probability to outcome models in which $\beta^{\boldsymbol{\alpha^Y}}$ has a causal interpretation according to Proposition 2.1, and that cannot be reduced according to Corollary 2.1. As will be seen, this prior distribution is constructed by borrowing information from the data.

The first step in the construction of BCEE's prior distribution $P^B(\boldsymbol{\alpha}^{\boldsymbol{Y}})$ is to compute the posterior distribution of the exposure model. This step is also present in TBAC and is performed in BCEE to identify possible causal exposure models and thus likely direct causes of the exposure. Recall that direct causes of exposure play a pivotal role in both Proposition 2.1 and Corollary 2.1. We now introduce the notation for the exposure model. Let $\boldsymbol{\alpha}^{\boldsymbol{X}} = (\alpha_1^X, ..., \alpha_M^X)$ be an M-dimensional vector for the inclusion of the covariates $\boldsymbol{U}$ in the exposure model. The exposure model is the following normal linear model

$$X_i = \delta_0^{\boldsymbol{\alpha}^{\boldsymbol{X}}} + \sum_{m=1}^{M} \alpha_m^X \delta_m^{\boldsymbol{\alpha}^{\boldsymbol{X}}} U_{im} + \epsilon_i^{\boldsymbol{\alpha}^{\boldsymbol{X}}}, \tag{6}$$

where $\delta_m^{\boldsymbol{\alpha}^{\boldsymbol{X}}}$ denotes the unknown regression coefficient of U_m, $m = 1, ..., M$, in the exposure model specified by $\boldsymbol{\alpha}^{\boldsymbol{X}}$. The parameter $\delta_0^{\boldsymbol{\alpha}^{\boldsymbol{X}}}$ denotes the unknown intercept in $\boldsymbol{\alpha}^{\boldsymbol{X}}$ and $\epsilon_i^{\boldsymbol{\alpha}^{\boldsymbol{X}}} \overset{iid}{\sim} N(0, \sigma^2_{\boldsymbol{\alpha}^{\boldsymbol{X}}})$. In this step, each model $\boldsymbol{\alpha}^{\boldsymbol{X}}$ is attributed a weight corresponding to its posterior probability, $P(\boldsymbol{\alpha}^{\boldsymbol{X}}|X) \propto P(X|\boldsymbol{\alpha}^{\boldsymbol{X}})P(\boldsymbol{\alpha}^{\boldsymbol{X}})$. For simplification, $P(\boldsymbol{\alpha}^{\boldsymbol{X}})$ is taken to be uniform (that is, $P(\boldsymbol{\alpha}^{\boldsymbol{X}}) = 1/2^M \ \ \forall \ \ \boldsymbol{\alpha}^{\boldsymbol{X}}$), although other prior distributions could be considered.

We are now ready to define $P^B(\boldsymbol{\alpha}^{\boldsymbol{Y}})$, which depends not only on $P(\boldsymbol{\alpha}^{\boldsymbol{X}}|X)$, but also on the regression coefficients $\delta_m^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}$. Remember that Proposition 2.1 and Corollary 2.1 both require verifying conditional independences. This can be achieved through the examination of the outcome model regression coefficients (see the final remarks of Section 2.2). To simplify the presentation, we assume for now that the true values of the regression coefficients are provided by an oracle. The BCEE prior distribution is as follows:

$$P^B(\boldsymbol{\alpha}^{\boldsymbol{Y}}) = \sum_{\boldsymbol{\alpha}^{\boldsymbol{X}}} P^B(\boldsymbol{\alpha}^{\boldsymbol{Y}}|\boldsymbol{\alpha}^{\boldsymbol{X}}) P(\boldsymbol{\alpha}^{\boldsymbol{X}}|X), \text{where}$$

$$P^B(\boldsymbol{\alpha}^{\boldsymbol{Y}}|\boldsymbol{\alpha}^{\boldsymbol{X}}) \propto \prod_{m=1}^{M} Q_{\boldsymbol{\alpha}^{\boldsymbol{Y}}}(\alpha_m^Y|\alpha_m^X).$$

For vectors $\boldsymbol{\alpha}^{\boldsymbol{Y}}$ and $\boldsymbol{\alpha}^{\boldsymbol{X}}$, $Q_{\boldsymbol{\alpha}^{\boldsymbol{Y}}}(\alpha_m^Y|\alpha_m^X)$ is given by one of the following:

$$\begin{aligned} &Q_{\boldsymbol{\alpha}^{\boldsymbol{Y}}}(\alpha_m^Y = 1|\alpha_m^X = 1) = \frac{\omega_m^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}}{\omega_m^{\boldsymbol{\alpha}^{\boldsymbol{Y}}} + 1}, \ Q_{\boldsymbol{\alpha}^{\boldsymbol{Y}}}(\alpha_m^Y = 0|\alpha_m^X = 1) = \frac{1}{\omega_m^{\boldsymbol{\alpha}^{\boldsymbol{Y}}} + 1}, \\ &Q_{\boldsymbol{\alpha}^{\boldsymbol{Y}}}(\alpha_m^Y = 1|\alpha_m^X = 0) = \frac{1}{2}, \ Q_{\boldsymbol{\alpha}^{\boldsymbol{Y}}}(\alpha_m^Y = 0|\alpha_m^X = 0) = \frac{1}{2}, \end{aligned} \tag{7}$$

where $\omega_m^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}$ is defined in (8). To properly define $\omega_m^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}$ we must first define the notion of an m-nearest neighbor outcome model. For a given model $\boldsymbol{\alpha}^{\boldsymbol{Y}}$ where $\alpha_m^Y = 0$, the m-nearest neighbor model to $\boldsymbol{\alpha}^{\boldsymbol{Y}}$, $\boldsymbol{\alpha}^{\boldsymbol{Y}}(m)$, has exactly the same covariates as $\boldsymbol{\alpha}^{\boldsymbol{Y}}$ except with $\alpha_m^Y = 1$ instead of $\alpha_m^Y = 0$. In the case where $\alpha_m^Y = 1$, there is no need to define an m-nearest neighbor model. We now define a new set of regression parameters:

$$\tilde{\delta}_m^{\boldsymbol{\alpha}^{\boldsymbol{Y}}} = \begin{cases} \delta_m^{\boldsymbol{\alpha}^{\boldsymbol{Y}}} & \text{if } \alpha_m^Y = 1 \\ \delta_m^{\boldsymbol{\alpha}^{\boldsymbol{Y}}(m)} & \text{if } \alpha_m^Y = 0. \end{cases}$$

For example, if $\boldsymbol{U} = \{U_1, U_2\}$ and $\boldsymbol{\alpha}^{\boldsymbol{Y}} = (1, 0)$ then $\tilde{\delta}_1^{\boldsymbol{\alpha}^{\boldsymbol{Y}}} = \delta_1^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}$ can be directly taken from model $\boldsymbol{\alpha}^{\boldsymbol{Y}}$, whereas $\tilde{\delta}_2^{\boldsymbol{\alpha}^{\boldsymbol{Y}}} = \delta_2^{\boldsymbol{\alpha}^{\boldsymbol{Y}}(2)}$ needs to be taken from model $\boldsymbol{\alpha}^{\boldsymbol{Y}}(2) = (1, 1)$.

With this additional notation, we define the hyperparameter $\omega_m^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}$ as:

$$\omega_m^{\boldsymbol{\alpha}^{\boldsymbol{Y}}} = \omega \times \left(\tilde{\delta}_m^{\boldsymbol{\alpha}^{\boldsymbol{Y}}} \frac{\sigma_{U_m}}{\sigma_Y} \right)^2, \tag{8}$$

where $0 \leq \omega \leq \infty$ is a user-defined hyperparameter, σ_{U_m} and σ_Y are respectively, the (true) standard deviations of U_m and Y. Note that $\tilde{\delta}_m^{\boldsymbol{\alpha}^{\boldsymbol{Y}}} \sigma_{U_m}/\sigma_Y$ is a standardization of $\tilde{\delta}_m^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}$ which makes it insensitive to the measurement units of both Y and U_m. In practice, we cannot rely on an oracle to provide $\tilde{\delta}_m^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}$; in the sequel, we use the maximum likelihood estimator of $\tilde{\delta}_m^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}$ instead. Also, the true values of σ_{U_m} and σ_Y are not known and are estimated by s_{U_m} and s_Y. The prior distribution $P^B(\boldsymbol{\alpha}^{\boldsymbol{Y}})$ thus has an empirical Bayes flavor. Once $P^B(\boldsymbol{\alpha}^{\boldsymbol{Y}})$ is obtained, the posterior distribution of the outcome model $P(\boldsymbol{\alpha}^{\boldsymbol{Y}}|Y)$ is computed and the posterior

exposure effect calculated according to eq. (5). In Section 3.2, we discuss how one can account for using the data for the specification of $P^B(\boldsymbol{\alpha}^{\boldsymbol{Y}})$ to obtain appropriate inferences.

2.4 The rationale behind BCEE

In this section, we explain in detail how BCEE's prior distribution $P^B(\boldsymbol{\alpha}^{\boldsymbol{Y}})$ is motivated by causal graphs through Proposition 2.1 and Corollary 2.1.

To begin, recall that the first step of BCEE serves to identify likely exposure models. Classical properties of Bayesian model selection ensure that the true (structural) exposure model, the one including only and all direct causes of X ($\boldsymbol{D} = \{D_1, ..., D_J\}$), is asymptotically attributed all the posterior probability by the first step of BCEE (e.g. Haughton [12], Wasserman [13]). This result follows from assuming that the set of potential confounding covariates $\boldsymbol{U}$ includes all direct causes of X and no descendants of X and that the specification of the model is correct: that is, the true exposure model is indeed a normal linear model of the form $X_i = \delta_0 + \sum_{j=1}^{J} \delta_j D_{ij} + \epsilon_i^X$, with $\epsilon_i^X \overset{iid}{\sim} N(0, \sigma_X^2)$.

The algorithm BCEE aims to give the bulk of the posterior weight to outcome models in which $\beta^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}$ has a causal interpretation according to Proposition 2.1 and that cannot be reduced according to Corollary 2.1. In such outcome models, $\boldsymbol{\alpha}^{\boldsymbol{Y}}$ includes any given direct cause (identified in the first step) only if the inclusion of this direct cause of exposure is necessary for $\beta^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}$ to have a causal interpretation in $\boldsymbol{\alpha}^{\boldsymbol{Y}}$. To do so, $P^B(\boldsymbol{\alpha}^{\boldsymbol{Y}})$ places small prior weight on outcome models which do not respect condition 2 of Proposition 2.1. In such models, some direct causes of X are excluded (condition point 2a from Proposition 2.1) and Y is dependent on those excluded direct causes of X given X and the potential confounding covariates already included (condition point 2b from Proposition 2.1). Moreover, $P^B(\boldsymbol{\alpha}^{\boldsymbol{Y}})$ seeks to limit the prior weight attributed to outcome models that could be reduced according to Corollary 2.1. In such models, some direct causes of X are included, but these are not associated with Y conditionally on X and the other covariates included.

To illustrate how Proposition 2.1 and Corollary 2.1 motivate the formulation of $P^B(\boldsymbol{\alpha}^{\boldsymbol{Y}})$ we provide the following thought experiment. To simplify our presentation, we assume that the direct causes of exposure are known and that the outcome model (1) is correctly specified. Moreover, we order the elements of $\boldsymbol{U}$ so that the first J elements are $\boldsymbol{D}$, that is $\{U_1, ..., U_M\} = \{D_1, ..., D_J, U_{J+1}, ..., U_M\}$. For ease of interpretation, we also assume that the covariates $\boldsymbol{U}$ are standardized, although, due to the way $\omega_m^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}$ is defined, this is not necessary in practice. We consider four different situations to illustrate how BCEE functions. In each situation, a direct cause of exposure $D_j = U_j$ is either included or excluded from the outcome model $\boldsymbol{\alpha}^{\boldsymbol{Y}}$ and the maximum likelihood estimate $|\hat{\tilde{\delta}}_j^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}|$ is either close to 0 or large. The anticipated magnitudes of $Q_{\boldsymbol{\alpha}^{\boldsymbol{Y}}}(\alpha_j^Y | \alpha_j^X)$ and of $P^B(\boldsymbol{\alpha}^{\boldsymbol{Y}} | \boldsymbol{\alpha}^{\boldsymbol{X}})$ for each situation are presented in Table 1. Considering jointly those four situations, we see that only outcome models that both correctly identify the average causal effect of exposure and that solely include necessary direct causes of exposure receive non-negligible prior

Table 1: Magnitudes of $Q_{\boldsymbol{\alpha}^{\boldsymbol{Y}}}(\alpha_j^Y | \alpha_j^X)$ and $P^B(\boldsymbol{\alpha}^{\boldsymbol{Y}} | \boldsymbol{\alpha}^{\boldsymbol{X}})$ for four situations defined by the inclusion of a direct cause of exposure D_j and the magnitude of $|\hat{\tilde{\delta}}_j^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}|$.

Situation	D_j	$\lvert\hat{\tilde{\delta}}_j^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}\rvert$	$\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{D}_j \mid \boldsymbol{X}, \boldsymbol{Z}'$	$\omega_j^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}$	$Q_{\boldsymbol{\alpha}^{\boldsymbol{Y}}}(\boldsymbol{\alpha}_j^{\boldsymbol{Y}} \mid \boldsymbol{\alpha}_j^{\boldsymbol{X}})$	$P^B(\boldsymbol{\alpha}^{\boldsymbol{Y}} \mid \boldsymbol{\alpha}^{\boldsymbol{X}})$
(1)	Excl.	Large	Not likely	Large	Close to 0	Close to 0
(2)	Incl.	Close to 0	Likely	Close to 0	Close to 0	Close to 0
(3)	Incl.	Large	Not likely	Large	Close to 1	Depends
(4)	Excl.	Close to 0	Likely	Close to 0	Close to 1	Depends

Note: $\boldsymbol{Z}'$ denotes the potential confounding covariates included in $\boldsymbol{\alpha}^{\boldsymbol{Y}}$ excluding D_j, Excl. = Excluded, Incl. = Included, Depends = Depends on other D_js

probabilities. In the next paragraph, we describe in detail the first situation, which supposes that direct cause D_j is omitted from $\boldsymbol{\alpha^Y}$ and its associated estimated parameter $|\hat{\delta}_j^{\boldsymbol{\alpha^Y}}|$ is large.

Suppose $\boldsymbol{\alpha^Y}$ does not include D_j. Note that $Q_{\boldsymbol{\alpha^Y}}(\alpha_j^Y = 0|\alpha_j^X = 1)$ depends on $\hat{\delta}_j^{\boldsymbol{\alpha^Y}}$ through $\omega_j^{\boldsymbol{\alpha^Y}}$. Therefore, $P^B(\boldsymbol{\alpha^Y}|\boldsymbol{\alpha^X})$ also depends on $\hat{\delta}_j^{\boldsymbol{\alpha^Y}}$. If $|\hat{\delta}_j^{\boldsymbol{\alpha^Y}}|$ is large, then Y is likely not independent of D_j conditionally on X and the potential confounding covariates included in $\boldsymbol{\alpha^Y}$. It follows that $\boldsymbol{\alpha^Y}$ does not respect condition 2b from Proposition 2.1. Since the value of $\omega_j^{\boldsymbol{\alpha^Y}}$ is large, $Q_{\boldsymbol{\alpha^Y}}(\alpha_j^Y = 0|\alpha_j^X = 1)$ is small and so is $P^B(\boldsymbol{\alpha^Y}|\boldsymbol{\alpha^X})$. In this situation, $P^B(\boldsymbol{\alpha^Y})$ is well behaved: the model $\boldsymbol{\alpha^Y}$ is not sufficient to identify the average causal effect of exposure and hence it receives little prior probability. A similar reasoning can be applied for situation 4 of Table 1. The reasoning for situations 2 and 3 is also quite similar, but requires invoking Corollary 2.1 to determine whether the inclusion of D_j is necessary or not.

Remark in Table 1 that in situations 3 and 4, where $Q_{\boldsymbol{\alpha^Y}}(\alpha_j^Y|\alpha_j^X)$ is close to 1, $P^B(\boldsymbol{\alpha^Y}|\boldsymbol{\alpha^X})$ depends in a large part on the $Q_{\boldsymbol{\alpha^Y}}$ associated with the other direct causes of exposure. If none of the $Q_{\boldsymbol{\alpha^Y}}$ are close to 0, then $P^B(\boldsymbol{\alpha^Y}|\boldsymbol{\alpha^X})$ is non-negligible and hence favors models that identify the causal effect according to Proposition 2.1 and Corollary 2.1. However, if any of the $Q_{\boldsymbol{\alpha^Y}}$ is close to 0, then $P^B(\boldsymbol{\alpha^Y}|\boldsymbol{\alpha^X})$ is close to 0.

2.5 A toy example

We consider a toy example to gain preliminary insights on the finite sample properties of BCEE. We generated a sample of size $n = 500$ satisfying the following relationships:

$$X = U_1 + U_2 + \epsilon_X$$
$$Y = X + 0.1U_1 + \epsilon_Y,$$

with U_1, $U_2 \sim N(0,1)$ and ϵ_X, $\epsilon_Y \sim N(0,1)$, all independent.

The first step of BCEE is to calculate the posterior distribution of the exposure model $P(\boldsymbol{\alpha^X}|X)$. The four possible exposure models in this example are:

$$\boldsymbol{\alpha_1^X} : X| \rightarrow (\alpha_1^X = 0, \alpha_2^X = 0),$$
$$\boldsymbol{\alpha_2^X} : X|U_1 \rightarrow (\alpha_1^X = 1, \alpha_2^X = 0),$$
$$\boldsymbol{\alpha_3^X} : X|U_2 \rightarrow (\alpha_1^X = 0, \alpha_2^X = 1),$$
$$\boldsymbol{\alpha_4^X} : X|U_1, U_2 \rightarrow (\alpha_1^X = 1, \alpha_2^X = 1).$$

We approximate $P(X|\boldsymbol{\alpha^X})$ using $\exp[-0.5BIC(\boldsymbol{\alpha^X})]$ [14], where $BIC(\boldsymbol{\alpha^X})$ is the Bayesian information criterion for exposure model $\boldsymbol{\alpha^X}$. In our example, model $\boldsymbol{\alpha_4^X}$ receives all posterior weight, that is $P(\boldsymbol{\alpha^X} = (1,1)|X) = 1$.

Next, we compute the posterior distribution of the outcome model using $P^B(\boldsymbol{\alpha^Y})$. We take $\omega = 100\sqrt{n}$, a choice that is subsequently discussed in Section 3.1. The four possible outcome models are:

$$\boldsymbol{\alpha_1^Y} : Y|X \rightarrow (\alpha_1^Y = 0, \alpha_2^Y = 0),$$
$$\boldsymbol{\alpha_2^Y} : Y|X, U_1 \rightarrow (\alpha_1^Y = 1, \alpha_2^Y = 0),$$
$$\boldsymbol{\alpha_3^Y} : Y|X, U_2 \rightarrow (\alpha_1^Y = 0, \alpha_2^Y = 1),$$
$$\boldsymbol{\alpha_4^Y} : Y|X, U_1, U_2 \rightarrow (\alpha_1^Y = 1, \alpha_2^Y = 1).$$

Note that only models $\boldsymbol{\alpha_2^Y}$ and $\boldsymbol{\alpha_4^Y}$ correctly identify the causal effect of exposure. We present the calculation of $P^B(\boldsymbol{\alpha^Y}|\boldsymbol{\alpha^X})$ for model $\boldsymbol{\alpha_2^Y}$. Since we obtained $P(\boldsymbol{\alpha^X} = (1,1)|X) = 1$, we only need to calculate $P^B(\boldsymbol{\alpha^Y} = (1,0)|\boldsymbol{\alpha^X} = (1,1)) \propto Q_{\boldsymbol{\alpha_2^Y}}(\alpha_1^Y = 1|\alpha_1^X = 1)Q_{\boldsymbol{\alpha_2^Y}}(\alpha_2^Y = 0|\alpha_2^X = 1)$. We have:

$$Q_{\boldsymbol{\alpha}_2^Y}(\alpha_1^Y = 1|\alpha_1^X = 1) = \frac{\omega_1^{\boldsymbol{\alpha}_2^Y}}{\omega_1^{\boldsymbol{\alpha}_2^Y} + 1},$$

$$Q_{\boldsymbol{\alpha}_2^Y}(\alpha_2^Y = 0|\alpha_2^X = 1) = \frac{1}{\omega_2^{\boldsymbol{\alpha}_2^Y} + 1}.$$

We get $\omega_1^{\boldsymbol{\alpha}_2^Y} = \omega(\hat{\tilde{\delta}}_1^{\boldsymbol{\alpha}_2^Y} \times s_{U_1}/s_Y)^2 = 100\sqrt{500}(0.14 \times 1.00/2.01)^2 = 9.75$. Note that because U_1 is included in $\boldsymbol{\alpha}_2^Y$, $\hat{\tilde{\delta}}_1^{\boldsymbol{\alpha}_2^Y} = \hat{\delta}_1^{\boldsymbol{\alpha}_2^Y}$. Also, we have $\omega_2^{\boldsymbol{\alpha}_2^Y} = \omega(\hat{\tilde{\delta}}_2^{\boldsymbol{\alpha}_2^Y} \times s_{U_2}/s_Y)^2 = 100\sqrt{500}(-0.01 \times 1.04/2.01)^2 = 0.05$. Because U_2 is not in $\boldsymbol{\alpha}_2^Y$, we get the regression parameter estimate for U_2 from its 2-nearest neighbor model, that is $\hat{\tilde{\delta}}_2^{\boldsymbol{\alpha}_2^Y} = \hat{\delta}_2^{\boldsymbol{\alpha}_4^Y}$. Finally, the value of the (unnormalized) prior probability of model $\boldsymbol{\alpha}_2^Y$ is 0.8658.

Following the same process for the three other outcome models, we calculate the prior probabilities. From there, we calculate the posterior distribution of the outcome model using the relationship $P(\boldsymbol{\alpha}^Y|Y) \propto P(Y|\boldsymbol{\alpha}^Y)P^B(\boldsymbol{\alpha}^Y)$. Again, we use $\exp[-0.5BIC(\boldsymbol{\alpha}^Y)]$ to approximate $P(Y|\boldsymbol{\alpha}^Y)$. Table 2 provides the results with the details of the intermediate steps.

Table 2: Calculation of the BCEE outcome model posterior distribution with intermediate steps.

Model	$U.P^B(\boldsymbol{\alpha}^Y)$	$P^B(\boldsymbol{\alpha}^Y)$	BIC	BMA $P(\boldsymbol{\alpha}^Y\|Y)$	$P(\boldsymbol{\alpha}^Y\|Y)$
$\boldsymbol{\alpha}_1^Y$	0.0230	0.0229	1,435.82	0.4602	0.0254
$\boldsymbol{\alpha}_2^Y$	0.8658	0.8618	1,435.81	0.4629	0.9625
$\boldsymbol{\alpha}_3^Y$	0.0749	0.0746	1,440.04	0.0560	0.0101
$\boldsymbol{\alpha}_4^Y$	0.0409	0.0407	1,442.00	0.0209	0.0021

Note: $U.P^B(\boldsymbol{\alpha}^Y)$ is the unnormalized prior probability, $P^B(\boldsymbol{\alpha}^Y)$ is the prior probability, BIC is the Bayesian information criterion, BMA $P(\boldsymbol{\alpha}^Y|Y)$ is the posterior probability the model resulting from a BMA procedure with a non-informative prior distribution, and $P(\boldsymbol{\alpha}^Y|Y)$ is the posterior probability using BCEE.

We see from these results how BCEE, as compared to BMA, shifts the posterior weight toward models that identify the causal effect of exposure. In fact, in this toy example, BCEE puts almost all the posterior weight on the true outcome model. BCEE accomplishes this by using an informative prior distribution for the outcome model that borrows information both from the exposure selection step and from neighboring regression coefficient estimates in the outcome models.

3 Practical considerations regarding BCEE

In this section we discuss practical considerations regarding the usage of the BCEE algorithm. First, we discuss the choice of the hyperparameter ω value in eq. (8), then we suggest two alternative ways of implementing BCEE.

3.1 Choice of ω

Recall that BCEE's prior distribution $P^B(\boldsymbol{\alpha}^Y)$ depends on a user-selected hyperparameter ω. In what follows, we suggest making ω proportional to $\sqrt{n}$ on the basis of asymptotic results related to the quantities $Q_{\boldsymbol{\alpha}^Y}$ in eq. (7). Without loss of generality, we only discuss the case $Q_{\boldsymbol{\alpha}^Y}(\alpha_m^Y = 1|\alpha_m^X = 1)$. Indeed, the cases $Q_{\boldsymbol{\alpha}^Y}(\alpha_m^Y = 1|\alpha_m^X = 0)$ and $Q_{\boldsymbol{\alpha}^Y}(\alpha_m^Y = 0|\alpha_m^X = 0)$ are trivial because the two quantities are both equal to 1/2. Moreover, the case $Q_{\boldsymbol{\alpha}^Y}(\alpha_m^Y = 0|\alpha_m^X = 1)$ is essentially equivalent to the case $Q_{\boldsymbol{\alpha}^Y}(\alpha_m^Y = 1|\alpha_m^X = 1)$ since these quantities are closely (and negatively) associated. Remark that because we consider the case where $\alpha_m^Y = 1$, $\tilde{\delta}_m^{\boldsymbol{\alpha}^Y} = \delta_m^{\boldsymbol{\alpha}^Y}$. However, we present the reasoning in terms of $\tilde{\delta}_m^{\boldsymbol{\alpha}^Y}$ to allow a direct generalization to the case where $\alpha_m^Y = 0$.

Assume that the true outcome model is a normal linear model of the form (1) and first consider the case $\tilde{\delta}_m^{\boldsymbol{\alpha}^Y} = 0$ for a given model $\boldsymbol{\alpha}^Y$. Then covariate U_m is conditionally independent of Y given the (other) covariates included in model $\boldsymbol{\alpha}^Y$. Hence U_m should be left out of $\boldsymbol{\alpha}^Y$ on the basis of Corollary 2.1. It is thus desirable that $Q_{\boldsymbol{\alpha}^Y}(\alpha_m^Y = 1|\alpha_m^X = 1) \to 0$ as $n \to \infty$, which happens if $\hat{\omega}_m^{\boldsymbol{\alpha}^Y} = \omega \times \left(\hat{\tilde{\delta}}_m^{\boldsymbol{\alpha}^Y} s_{U_m}/s_Y\right)^2 \to 0$ as $n \to \infty$.

Consider the case $\tilde{\delta}_m^{\boldsymbol{\alpha}^Y} \neq 0$. According to Proposition 2.1, it is now desirable that $Q_{\boldsymbol{\alpha}^Y}(\alpha_m^Y = 1|\alpha_m^X = 1) \to 1$ as $n \to \infty$, since this would allow for covariates causing less confounding to be forced in the outcome model as n grows. Thus, we need $\hat{\omega}_m^{\boldsymbol{\alpha}^Y} \to \infty$ as $n \to \infty$ if $\tilde{\delta}_m^{\boldsymbol{\alpha}^Y} \neq 0$.

If $\tilde{\delta}_m^{\boldsymbol{\alpha}^Y} = 0$ then $\hat{\tilde{\delta}}_m^{\boldsymbol{\alpha}^Y} \xrightarrow{P} 0$ and thus, for any finite constant value of ω, $\hat{\omega}_m^{\boldsymbol{\alpha}^Y} \xrightarrow{P} 0$, where $\xrightarrow{P}$ means convergence in probability. However, if $\tilde{\delta}_m^{\boldsymbol{\alpha}^Y} \neq 0$, we need to choose ω as a function of sample size n in order to ensure that $\hat{\omega}_m^{\boldsymbol{\alpha}^Y} \to \infty$ as $n \to \infty$. We consider rates of convergence to find an appropriate function of n.

Recall that $\hat{\omega}_m^{\boldsymbol{\alpha}^Y}$ is a function of the MLE $\hat{\tilde{\delta}}_m^{\boldsymbol{\alpha}^Y}$ (Section 2.3). Under mild regularity conditions, it follows from the results in Yuan and Chan [15] that $\hat{\tilde{\delta}}_m^{\boldsymbol{\alpha}^Y} s_{U_m}/s_Y \xrightarrow{P} \tilde{\delta}_m^{\boldsymbol{\alpha}^Y} \sigma_{U_m}/\sigma_Y$ at rate $O_p(1/\sqrt{n})$, where O_p is the usual big-O_p notation (Agresti [16], p. 588). Thus $\left(\hat{\tilde{\delta}}_m^{\boldsymbol{\alpha}^Y} s_{U_m}/s_Y\right)^2 \xrightarrow{P} \left(\tilde{\delta}_m^{\boldsymbol{\alpha}^Y} \sigma_{U_m}/\sigma_Y\right)^2$ at rate $O_p(1/n)$.

By taking $\omega = cn^b$, with $0 < b < 1$, where c is a user-fixed constant that does not depend on sample size, we obtain $\hat{\omega}_m^{\boldsymbol{\alpha}^Y} \to \infty$ (at rate n^b) if $\tilde{\delta}_m^{\boldsymbol{\alpha}^Y} \neq 0$ and $\hat{\omega}_m^{\boldsymbol{\alpha}^Y} \xrightarrow{P} 0$ (with convergence rate $O_p(1/n^{1-b})$) if $\tilde{\delta}_m^{\boldsymbol{\alpha}^Y} = 0$, as desired. The value $b = 1/2$ appears to make a good compromise between the two desired convergence behaviors. The simulation study presented in Section 4 shows that BCEE performs well for $\omega = c\sqrt{n}$ with $100 \leq c \leq 1000$. We also see that larger values of c yield less bias and more variance in the estimator of the causal effect, and conversely for smaller values of c. Appendix C illustrates how $Q_{\boldsymbol{\alpha}^Y}(\alpha_m^Y = 1|\alpha_m^X = 1)$ behaves for different values of c in some simple settings.

3.2 Implementing BCEE

In this section, we first consider a naive implementation of BCEE that closely follows our presentation of the algorithm in Section 2.3. Then we describe a modified implementation that accounts for using the MLE $\hat{\tilde{\delta}}_m^{\boldsymbol{\alpha}^Y}$ in $P^B(\boldsymbol{\alpha}^Y)$.

We perform three steps to sample one draw from the posterior distribution of the average causal exposure effect $P(\beta|Y)$. Several such draws are taken to obtain approximations to quantities of interest, such as the posterior mean and variance of β. The steps of the sampling procedure are:

S1. Draw $\boldsymbol{\alpha}^X$ from the posterior distribution of the exposure model $P(\boldsymbol{\alpha}^X|X) \propto P(X|\boldsymbol{\alpha}^X)$, using $\exp[-0.5BIC(\boldsymbol{\alpha}^X)]$ to approximate $P(X|\boldsymbol{\alpha}^X)$;

S2. Draw $\boldsymbol{\alpha}^Y$ from the conditional posterior distribution $P(\boldsymbol{\alpha}^Y|\boldsymbol{\alpha}^X, Y) \propto P^B(\boldsymbol{\alpha}^Y|\boldsymbol{\alpha}^X)P(Y|\boldsymbol{\alpha}^Y)$, where the regression coefficients $\tilde{\delta}_m^{\boldsymbol{\alpha}^Y}$ are estimated by their MLEs and $P(Y|\boldsymbol{\alpha}^Y)$ is approximated by $\exp[-0.5BIC(\boldsymbol{\alpha}^Y)]$;

S3. Draw β from the conditional posterior distribution $P(\beta^{\boldsymbol{\alpha}^Y}|\boldsymbol{\alpha}^Y, Y)$, which we approximate by its limit normal distribution $N\left(\hat{\beta}^{\boldsymbol{\alpha}^Y}, \widehat{SE}(\hat{\beta}^{\boldsymbol{\alpha}^Y})\right)$ [17, 18], where $\hat{\beta}^{\boldsymbol{\alpha}^Y}$ is the maximum likelihood estimator of $\beta^{\boldsymbol{\alpha}^Y}$ and $\widehat{SE}(\hat{\beta}^{\boldsymbol{\alpha}^Y})$ is its estimated standard error.

The sampling for the first two steps is done using Markov chain Monte Carlo model composition (MC^3) [19]. We refer to this naive implementation of BCEE as N-BCEE.

Because N-BCEE does not take into account the uncertainty related to the estimation of the regression coefficients $\tilde{\delta}_m^{\boldsymbol{\alpha}^Y}$ in $P^B(\boldsymbol{\alpha}^Y)$, we anticipate that the confidence (credible) interval for β will be too narrow. Our insight relies on the Empirical Bayes literature, where it has been extensively shown that data-dependent prior distributions lead to confidence intervals that tend to be "too short, inappropriately centered, or both" [20]. Also, narrow confidence intervals for β are observed in simulations presented in Section 4. Although

many solutions to this problem have been proposed (see Carlin and Louis [21] for a short discussion), most cannot be realistically applied to BCEE due to the complexity of the algorithm. Therefore, we propose the following simple ad hoc solution, which happens to be notably faster than N-BCEE. We refer to this modified implementation of BCEE as A-BCEE.

A-BCEE is the same as N-BCEE except for step S2. Recall that this step is directed at sampling from the conditional posterior distribution $P(\boldsymbol{\alpha}^{\boldsymbol{Y}}|\boldsymbol{\alpha}^{\boldsymbol{X}}, Y)$ using MC^3. This MC^3 scheme requires calculating a Metropolis-Hasting ratio (RP) which involves the ratio of the (conditional) prior probabilities of the proposed outcome model, $\boldsymbol{\alpha}_{\boldsymbol{1}}^{\boldsymbol{Y}}$, to the current outcome model, $\boldsymbol{\alpha}_{\boldsymbol{2}}^{\boldsymbol{Y}}$:

$$RP = \frac{P^B(\boldsymbol{\alpha}_{\boldsymbol{1}}^{\boldsymbol{Y}}|\boldsymbol{\alpha}^{\boldsymbol{X}})}{P^B(\boldsymbol{\alpha}_{\boldsymbol{2}}^{\boldsymbol{Y}}|\boldsymbol{\alpha}^{\boldsymbol{X}})} = \frac{\prod_{m=1}^{M} Q_{\boldsymbol{\alpha}_{\boldsymbol{1}}^{\boldsymbol{Y}}}(\alpha_m^Y|\alpha_m^X)/C}{\prod_{m=1}^{M} Q_{\boldsymbol{\alpha}_{\boldsymbol{2}}^{\boldsymbol{Y}}}(\alpha_m^Y|\alpha_m^X)/C} = \prod_{m=1}^{M} \frac{Q_{\boldsymbol{\alpha}_{\boldsymbol{1}}^{\boldsymbol{Y}}}(\alpha_m^Y|\alpha_m^X)}{Q_{\boldsymbol{\alpha}_{\boldsymbol{2}}^{\boldsymbol{Y}}}(\alpha_m^Y|\alpha_m^X)}, \tag{9}$$

where C is a normalizing constant such that $P^B(\boldsymbol{\alpha}^{\boldsymbol{Y}}|\boldsymbol{\alpha}^{\boldsymbol{X}}) = \prod_{m=1}^{M} Q_{\boldsymbol{\alpha}^{\boldsymbol{Y}}}(\alpha_m^Y|\alpha_m^X)/C$. In RP, $\boldsymbol{\alpha}_{\boldsymbol{1}}^{\boldsymbol{Y}}$ and $\boldsymbol{\alpha}_{\boldsymbol{2}}^{\boldsymbol{Y}}$ are two neighbor outcome models that differ only by their inclusion of a single covariate $U_{m'}$. A-BCEE utilizes the following simplification for RP:

$$RP \approx \frac{Q_{\boldsymbol{\alpha}_{\boldsymbol{1}}^{\boldsymbol{Y}}}(\alpha_{m'}^Y|\alpha_{m'}^X)}{Q_{\boldsymbol{\alpha}_{\boldsymbol{2}}^{\boldsymbol{Y}}}(\alpha_{m'}^Y|\alpha_{m'}^X)}. \tag{10}$$

The heuristic for suggesting this approximation is that the individual ratio that is the most likely to significantly differ from 1 in eq. (9) is the one associated to covariate $U_{m'}$, that is $Q_{\boldsymbol{\alpha}_{\boldsymbol{1}}^{\boldsymbol{Y}}}(\alpha_{m'}^Y|\alpha_{m'}^X)/Q_{\boldsymbol{\alpha}_{\boldsymbol{2}}^{\boldsymbol{Y}}}(\alpha_{m'}^Y|\alpha_{m'}^X)$. In fact, unless the covariates $\boldsymbol{U}$ are very strongly correlated with each other, we expect the $\hat{\delta}_m^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}$s ($m \neq m'$) to be of the same magnitude between two neighboring models. Note that we also expect many terms in the RP product to be exactly equal to 1 since an individual ratio equals 1 when its corresponding covariate is not included in the exposure model $\left(Q_{\boldsymbol{\alpha}_{\boldsymbol{1}}^{\boldsymbol{Y}}}(\alpha_m^Y|\alpha_m^X = 0)/Q_{\boldsymbol{\alpha}_{\boldsymbol{2}}^{\boldsymbol{Y}}}(\alpha_m^Y|\alpha_m^X = 0) = 1\right)$. Simulations were performed to verify the validity of approximation (10) (results not presented).

Using simplified RP (10), it becomes an easy task to incorporate the variability associated with the estimation of the $\tilde{\delta}^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}$s. We assume that $\tilde{\delta}_{m'}^{\boldsymbol{\alpha}^{\boldsymbol{Y}}} \sim N\left(\hat{\delta}_{m'}^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}, \widehat{SE}\left(\hat{\delta}_{m'}^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}\right)\right)$, where $\widehat{SE}\left(\hat{\delta}_{m'}^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}\right)$ is the estimated standard error of $\hat{\delta}_{m'}^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}$. In summary, in step S2 of the sampling procedure of A-BCEE we simply draw $\tilde{\delta}_{m'}^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}$ from $N\left(\hat{\delta}_{m'}^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}, \widehat{SE}\left(\hat{\delta}_{m'}^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}\right)\right)$ and use it in approximation (10). We remark that this strategy is akin to specifying an empirical Bayes type of hyperprior for $\tilde{\delta}^{\boldsymbol{\alpha}^{\boldsymbol{Y}}}$.

The finite sample properties of N-BCEE and A-BCEE are studied and compared in some simulation scenarios presented in the next section. We also consider nonparametric bootstrap [22] in a few simple and small scale simulations as an alternative to A-BCEE to correct confidence intervals. Note that, due to computing time, this bootstrapped BCEE (B-BCEE) approach is considerably less practical than A-BCEE to evaluate in simulations and to apply to real data sets of moderate to large sizes.

4 Simulation studies

In this section, we study the finite sample properties of BCEE in various simulation scenarios. The first primary objective of the simulations is to compare BCEE to standard or related methods that are used to estimate total average causal effects of exposure. The second primary objective is to study the sensitivity of BCEE to the choice of its user-selected hyperparameter ω. In Appendix D, we study two other secondary objectives relating to the large, whilst finite, properties of BCEE and to the performance of B-BCEE.

To achieve the two main objectives, we examine 24 different simulation scenarios obtained by considering three factors: data-generating process (DGP1, DGP2, DGP3 and DGP4), sample size (200, 600 and

1,000) and true causal effect of exposure ($\beta = 0.1$ or $\beta = 0$). The four data-generating processes are described below.

The first data-generating process (DGP1) satisfies the following relationships:

$$U_3 = U_2 + \epsilon_3$$

$$U_5 = U_4 + \epsilon_5$$

$$X = U_1 + U_2 + U_4 + \epsilon_X$$

$$Y = U_3 + 0.1U_4 + U_5 + \beta X + \epsilon_Y,$$

with $U_1, U_2, U_4, \epsilon_3, \epsilon_5, \epsilon_X, \epsilon_Y \sim N(0,1)$ all independent. The set of available covariates is $\boldsymbol{U} = \{U_1, U_2, \ldots, U_5\}$.

The second data-generating process (DGP2) involves a larger number of covariates than DGP1 and features an indirect effect of X on Y:

$$U_1 = U_4 + \epsilon_1$$

$$U_2 = U_4 + \epsilon_2$$

$$U_3 = U_4 + \epsilon_3$$

$$U_5 = U_1 + \epsilon_5$$

$$X = U_1 + U_2 + U_3 + \epsilon_X$$

$$U_6 = 0.5X + U_3 + \epsilon_6$$

$$Y = 0.1U_4 + 0.1U_5 + \beta U_6 + 0.5\beta X + \epsilon_Y,$$

where $U_4, \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_5, \epsilon_X, \epsilon_6, \epsilon_Y \sim N(0,1)$ all independent. The set of available covariates is $\boldsymbol{U} = \{U_1, U_2, \ldots, U_5, U_7, \ldots, U_{15}\}$, where $U_7, \ldots, U_{15}$ are all independent $N(0,1)$. We exclude U_6 from the set of potential confounding covariates since one must not adjust for descendants of the exposure X to identify the total average causal effect. Here the total effect of X on Y (direct effect plus indirect effect through U_6) is $0.5\beta + 0.5\beta = \beta$. For simulation purposes, we consider the model $\boldsymbol{\alpha}^{\boldsymbol{Y}} = (0, 0, 1, 1, 1, 0, ..., 0)$ as the "true" outcome model.

The third data-generating process (DGP3) is similar to the first simulation example in Wang et al. [6] but includes only 18 additional (noise) covariates (instead of 49):

$$X = 0.7U_1 + \sqrt{(1 - 0.7^2)}\epsilon_X$$

$$Y = 0.1U_1 + 0.1U_2 + \beta X + \epsilon_Y,$$

where $U_1, U_2, \epsilon_X, \epsilon_Y \sim N(0,1)$ all independent. The set of available covariates is $\boldsymbol{U} = \{U_1, U_2, \ldots, U_{20}\}$, where $U_3, \ldots, U_{20}$ are also independent $N(0,1)$.

The fourth data-generating process (DGP4) is inspired by a DAG presented in Morgan and Winship [23], Figure 1.1, page 25:

$$X = 0.1U_1 + 0.1U_2 + 0.1U_3 + \epsilon_X$$

$$U_6 = U_3 + \epsilon_6$$

$$Y = 0.1U_4 + 0.5U_5 + 0.5U_6 + \beta X + \epsilon_Y,$$

where $\epsilon_X, \epsilon_6, \epsilon_Y \sim N(0,1)$ all independent. Covariates U_1, U_2, U_3, U_4, U_5 are also $N(0,1)$ and are all independent except U_1, U_2 and U_1, U_4 for which we have $Cov(U_1, U_2) = 0.7$ and $Cov(U_1, U_4) = 0.7$. Notice that U_1 is a collider between U_2 and U_4 and thus $Cov(U_2, U_4) = 0$.

For each of the 24 simulation scenarios, we randomly generated 500 datasets. We estimated the average causal effect of exposure using 8 different procedures: (1) the true outcome model, (2) the fully adjusted model, (3) Bayesian model averaging (BMA) with a uniform prior distribution on the outcome model, (4) Bayesian adjustment for confounding (BAC) with ω chosen with cross-validation criterion $C_V^m(\omega)$ proposed in Lefebvre et al. [7], (5) BAC with $\omega = \infty$, (6) Two-stage Bayesian adjustment for confounding (TBAC) with $\omega = \infty$, (7) N-BCEE, and (8) A-BCEE. For both N-BCEE and A-BCEE, we used $\omega = c\sqrt{n}$ and considered $c = 100$, $c = 500$ and $c = 1000$. For each scenario and each method of estimation, we computed the average causal effect estimate (*Mean*), the average standard error estimate ($\overline{SEE}$), the standard deviation of the estimates (*SDE*), the root mean squared error ($\sqrt{MSE}$) and the coverage probability of 95% confidence intervals (*CP*). All 95% confidence intervals were computed using the normal approximation $\hat{\beta} \pm 1.96SEE$. Tables 3, 4, 5 and 6 summarize the results for $\beta = 0.1$. The marginal posterior probability of inclusion of

Table 3: Comparison of estimates of β obtained from the true outcome model, the fully adjusted model, BMA, BAC, TBAC, N-BCEE, and A-BCEE for 500 Monte Carlo replicates of the first data-generating process (DGP1).

n	*Method*	*Mean*	$\overline{SEE}$	*SDE*	$\sqrt{MSE}$	*CP*
200	True model	0.100	0.045	0.047	0.047	94
200	Fully adjusted model	0.098	0.072	0.074	0.074	94
200	BMA	0.113	0.047	0.047	0.048	95
200	BAC ($C_V^m(\omega)$)	0.104	0.055	0.064	0.064	92
200	BAC ($\omega = \infty$)	0.098	0.072	0.074	0.074	94
200	TBAC ($\omega = \infty$)	0.098	0.072	0.074	0.074	94
200	N-BCEE ($c = 100$)	0.108	0.051	0.055	0.056	93
200	N-BCEE ($c = 500$)	0.104	0.055	0.062	0.062	92
200	N-BCEE ($c = 1000$)	0.102	0.057	0.065	0.065	93
200	A-BCEE ($c = 100$)	0.107	0.055	0.054	0.054	95
200	A-BCEE ($c = 500$)	0.104	0.061	0.060	0.060	96
200	A-BCEE ($c = 1000$)	0.103	0.063	0.063	0.063	96
600	True model	0.100	0.026	0.025	0.025	96
600	Fully adjusted model	0.100	0.041	0.039	0.039	97
600	BMA	0.111	0.027	0.027	0.029	94
600	BAC ($C_V^m(\omega)$)	0.105	0.031	0.035	0.035	95
600	BAC ($\omega = \infty$)	0.100	0.041	0.039	0.039	97
600	TBAC ($\omega = \infty$)	0.100	0.041	0.039	0.039	96
600	N-BCEE ($c = 100$)	0.108	0.029	0.031	0.031	93
600	N-BCEE ($c = 500$)	0.106	0.030	0.033	0.034	93
600	N-BCEE ($c = 1000$)	0.105	0.031	0.034	0.034	93
600	A-BCEE ($c = 100$)	0.108	0.030	0.030	0.031	95
600	A-BCEE ($c = 500$)	0.105	0.033	0.032	0.032	97
600	A-BCEE ($c = 1000$)	0.105	0.035	0.033	0.033	97
1,000	True model	0.101	0.020	0.020	0.020	95
1,000	Fully adjusted model	0.100	0.032	0.033	0.033	94
1,000	BMA	0.111	0.021	0.022	0.025	92
1,000	BAC ($C_V^m(\omega)$)	0.102	0.026	0.030	0.030	93
1,000	BAC ($\omega = \infty$)	0.100	0.032	0.033	0.033	94
1,000	TBAC ($\omega = \infty$)	0.100	0.032	0.033	0.033	94
1,000	N-BCEE ($c = 100$)	0.107	0.022	0.024	0.025	94
1,000	N-BCEE ($c = 500$)	0.105	0.023	0.026	0.026	94
1,000	N-BCEE ($c = 1000$)	0.104	0.024	0.026	0.027	94
1,000	A-BCEE ($c = 100$)	0.107	0.023	0.024	0.025	95
1,000	A-BCEE ($c = 500$)	0.105	0.026	0.026	0.026	96
1,000	A-BCEE ($c = 1000$)	0.104	0.027	0.027	0.027	96

Note: *Mean* is the mean estimated value of β where the true value is 0.1, $\overline{SEE}$ is the mean standard error estimate, *SDE* is the standard deviation of the estimates of β, $\sqrt{MSE}$ is the squared-root of the mean squared error, *CP* is the coverage probability in % of 95% confidence intervals.

Table 4: Comparison of estimates of β obtained from the true outcome model, the fully adjusted model, BMA, BAC, TBAC, N-BCEE, and A-BCEE for 500 Monte Carlo replicates of the second data-generating process (DGP2).

n	*Method*	*Mean*	$\overline{SEE}$	*SDE*	$\sqrt{MSE}$	*CP*
200	True model	0.102	0.046	0.045	0.045	96
200	Fully adjusted model	0.104	0.075	0.078	0.078	94
200	BMA	0.148	0.044	0.046	0.067	68
200	BAC ($C_V^m(\omega)$)	0.118	0.052	0.075	0.077	76
200	BAC ($\omega = \infty$)	0.103	0.073	0.077	0.077	95
200	TBAC ($\omega = \infty$)	0.103	0.073	0.076	0.076	95
200	N-BCEE ($c = 100$)	0.120	0.053	0.068	0.071	83
200	N-BCEE ($c = 500$)	0.110	0.058	0.073	0.074	86
200	N-BCEE ($c = 1000$)	0.107	0.060	0.074	0.074	88
200	A-BCEE ($c = 100$)	0.120	0.062	0.066	0.069	92
200	A-BCEE ($c = 500$)	0.112	0.067	0.071	0.072	95
200	A-BCEE ($c = 1000$)	0.110	0.068	0.072	0.073	95
600	True model	0.100	0.026	0.026	0.026	96
600	Fully adjusted model	0.102	0.042	0.041	0.041	95
600	BMA	0.133	0.030	0.032	0.046	70
600	BAC ($C_V^m(\omega)$)	0.106	0.036	0.042	0.042	85
600	BAC ($\omega = \infty$)	0.102	0.041	0.041	0.041	96
600	TBAC ($\omega = \infty$)	0.102	0.041	0.041	0.041	96
600	N-BCEE ($c = 100$)	0.114	0.032	0.037	0.040	86
600	N-BCEE ($c = 500$)	0.108	0.034	0.039	0.039	91
600	N-BCEE ($c = 1000$)	0.106	0.035	0.039	0.040	91
600	A-BCEE ($c = 100$)	0.114	0.036	0.037	0.039	92
600	A-BCEE ($c = 500$)	0.109	0.038	0.038	0.039	94
600	A-BCEE ($c = 1000$)	0.107	0.039	0.039	0.039	94
1,000	True model	0.100	0.020	0.021	0.021	95
1,000	Fully adjusted model	0.100	0.032	0.032	0.032	95
1,000	BMA	0.121	0.024	0.027	0.034	80
1,000	BAC ($C_V^m(\omega)$)	0.100	0.029	0.031	0.031	92
1,000	BAC ($\omega = \infty$)	0.099	0.032	0.032	0.031	95
1,000	TBAC ($\omega = \infty$)	0.099	0.032	0.032	0.031	95
1,000	N-BCEE ($c = 100$)	0.107	0.025	0.029	0.029	90
1,000	N-BCEE ($c = 500$)	0.103	0.026	0.029	0.029	90
1,000	N-BCEE ($c = 1000$)	0.102	0.026	0.030	0.030	91
1,000	A-BCEE ($c = 100$)	0.108	0.028	0.028	0.029	93
1,000	A-BCEE ($c = 500$)	0.104	0.029	0.029	0.029	94
1,000	A-BCEE ($c = 1000$)	0.103	0.030	0.029	0.030	95

Note: *Mean* is the mean estimated value of β where the true value is 0.1, $\overline{SEE}$ is the mean standard error estimate, *SDE* is the standard deviation of the estimates of β, $\sqrt{MSE}$ is the squared-root of the mean squared error, *CP* is the coverage probability in % of 95% confidence intervals.

each potential confounding covariate can be found in Tables 11 to 14 in Appendix E. The results for $\beta = 0$ are similar (not presented).

We start by discussing the results pertaining to non-BCEE methods for estimating the average causal effect of exposure. Then, we discuss the results for BCEE and contrast them to the former results.

As expected, Bayesian model averaging (BMA) can perform very poorly to estimate the average causal effect. More precisely, the simulation results show that the bias can be substantial when the most important confounding covariates are only slightly associated with the outcome (DGP2 and DGP3). For instance, in DGP2, U_3 and U_4 are important confounding covariates often excluded by BMA (see Table 12 in Appendix E). Similarly, in DGP3, U_1 is often excluded by BMA (see Table 13). This situation also yields confidence

Table 5: Comparison of estimates of β obtained from the true outcome model, the fully adjusted model, BMA, BAC, TBAC, N-BCEE, and A-BCEE for 500 Monte Carlo replicates of the third data-generating process (DGP3).

n	*Method*	*Mean*	$\overline{SEE}$	*SDE*	$\sqrt{MSE}$	*CP*
200	True model	0.103	0.100	0.100	0.100	95
200	Fully adjusted model	0.101	0.105	0.104	0.104	96
200	BMA	0.149	0.085	0.086	0.099	89
200	BAC ($C_V^m(\omega)$)	0.116	0.087	0.103	0.104	90
200	BAC ($\omega = \infty$)	0.101	0.100	0.101	0.101	95
200	TBAC ($\omega = \infty$)	0.102	0.101	0.101	0.101	95
200	N-BCEE ($c = 100$)	0.113	0.093	0.100	0.101	93
200	N-BCEE ($c = 500$)	0.106	0.096	0.101	0.101	94
200	N-BCEE ($c = 1000$)	0.104	0.097	0.101	0.101	94
200	A-BCEE ($c = 100$)	0.116	0.096	0.098	0.099	95
200	A-BCEE ($c = 500$)	0.109	0.098	0.099	0.100	95
200	A-BCEE ($c = 1000$)	0.108	0.099	0.100	0.100	95
600	True model	0.098	0.057	0.060	0.060	96
600	Fully adjusted model	0.098	0.058	0.061	0.061	96
600	BMA	0.138	0.054	0.061	0.072	80
600	BAC ($C_V^m(\omega)$)	0.104	0.054	0.065	0.065	87
600	BAC ($\omega = \infty$)	0.097	0.058	0.060	0.060	96
600	TBAC ($\omega = \infty$)	0.097	0.057	0.060	0.060	95
600	N-BCEE ($c = 100$)	0.108	0.056	0.064	0.064	88
600	N-BCEE ($c = 500$)	0.101	0.056	0.062	0.062	92
600	N-BCEE ($c = 1000$)	0.100	0.056	0.061	0.061	92
600	A-BCEE ($c = 100$)	0.111	0.057	0.063	0.064	90
600	A-BCEE ($c = 500$)	0.104	0.057	0.062	0.062	92
600	A-BCEE ($c = 1000$)	0.103	0.057	0.062	0.062	94
1,000	True model	0.098	0.044	0.043	0.043	96
1,000	Fully adjusted model	0.098	0.045	0.043	0.043	95
1,000	BMA	0.130	0.045	0.050	0.058	79
1,000	BAC ($C_V^m(\omega)$)	0.102	0.043	0.046	0.046	91
1,000	BAC ($\omega = \infty$)	0.098	0.045	0.043	0.043	96
1,000	TBAC ($\omega = \infty$)	0.098	0.044	0.043	0.043	96
1,000	N-BCEE ($c = 100$)	0.106	0.044	0.048	0.048	91
1,000	N-BCEE ($c = 500$)	0.101	0.044	0.045	0.045	93
1,000	N-BCEE ($c = 1000$)	0.100	0.044	0.044	0.044	94
1,000	A-BCEE ($c = 100$)	0.108	0.045	0.048	0.048	92
1,000	A-BCEE ($c = 500$)	0.103	0.045	0.046	0.046	94
1,000	A-BCEE ($c = 1000$)	0.102	0.045	0.045	0.045	94

Note: *Mean* is the mean estimated value of β where the true value is 0.1, $\overline{SEE}$ is the mean standard error estimate, *SDE* is the standard deviation of the estimates of β, $\sqrt{MSE}$ is the squared-root of the mean squared error, *CP* is the coverage probability in % of 95% confidence intervals.

intervals with poor coverage probabilities. Although increasing sample size seems to reduce the bias, the coverage probability remains mostly unchanged. In situations where the most important confounding covariates are strongly associated with the outcome (DGP1 and DGP4), BMA performs very well both in terms of mean squared error (MSE) and coverage probability.

The simulation results also support the claim that BAC and TBAC with $\omega = \infty$ do not yield a notable reduction in the variance of the estimated causal effect as compared to the fully adjusted model. This is partly due to the fact that BAC and TBAC tend to include more covariates than needed to achieve unbiasedness (see Appendix E). Moreover, using BAC with cross-validation criterion $C_V^m(\omega)$ gives relatively poor results. Even though this method sometimes gives smaller MSE than BAC with $\omega = \infty$, the estimated standard error remarkably underestimate the true standard error (the standard deviation of the estimates of β).

Table 6: Comparison of estimates of β obtained from the true outcome model, the fully adjusted model, BMA, BAC, TBAC, N-BCEE, and A-BCEE for 500 Monte Carlo replicates of the fourth data-generating process (DGP4).

n	*Method*	*Mean*	$\overline{SEE}$	*SDE*	$\sqrt{MSE}$	*CP*
200	True model	0.103	0.054	0.052	0.052	96
200	Fully adjusted model	0.105	0.072	0.068	0.068	95
200	BMA	0.119	0.060	0.054	0.057	96
200	BAC ($C_V^m(\omega)$)	0.110	0.061	0.062	0.063	95
200	BAC ($\omega = \infty$)	0.103	0.072	0.068	0.068	95
200	TBAC ($\omega = \infty$)	0.105	0.071	0.067	0.067	96
200	N-BCEE ($c = 100$)	0.108	0.061	0.063	0.064	93
200	N-BCEE ($c = 500$)	0.106	0.064	0.066	0.066	94
200	N-BCEE ($c = 1000$)	0.105	0.065	0.066	0.066	95
200	A-BCEE ($c = 100$)	0.110	0.066	0.062	0.062	96
200	A-BCEE ($c = 500$)	0.107	0.068	0.064	0.064	96
200	A-BCEE ($c = 1000$)	0.107	0.068	0.065	0.065	96
600	True model	0.099	0.031	0.031	0.031	95
600	Fully adjusted model	0.097	0.041	0.043	0.043	95
600	BMA	0.110	0.036	0.036	0.038	92
600	BAC ($C_V^m(\omega)$)	0.100	0.037	0.042	0.042	92
600	BAC ($\omega = \infty$)	0.096	0.041	0.043	0.043	95
600	TBAC ($\omega = \infty$)	0.096	0.041	0.042	0.043	94
600	N-BCEE ($c = 100$)	0.102	0.036	0.040	0.040	92
600	N-BCEE ($c = 500$)	0.099	0.037	0.041	0.041	92
600	N-BCEE ($c = 1000$)	0.098	0.037	0.042	0.042	92
600	A-BCEE ($c = 100$)	0.102	0.038	0.040	0.040	94
600	A-BCEE ($c = 500$)	0.100	0.039	0.041	0.041	94
600	A-BCEE ($c = 1000$)	0.099	0.040	0.041	0.041	94
1,000	True model	0.099	0.024	0.024	0.024	96
1,000	Fully adjusted model	0.099	0.032	0.032	0.032	95
1,000	BMA	0.107	0.028	0.029	0.030	92
1,000	BAC ($C_V^m(\omega)$)	0.100	0.029	0.032	0.032	91
1,000	BAC ($\omega = \infty$)	0.098	0.032	0.032	0.032	94
1,000	TBAC ($\omega = \infty$)	0.098	0.032	0.032	0.032	93
1,000	N-BCEE ($c = 100$)	0.102	0.028	0.030	0.030	92
1,000	N-BCEE ($c = 500$)	0.100	0.028	0.031	0.031	92
1,000	N-BCEE ($c = 1000$)	0.100	0.029	0.032	0.031	92
1,000	A-BCEE ($c = 100$)	0.102	0.030	0.031	0.031	94
1,000	A-BCEE ($c = 500$)	0.101	0.030	0.031	0.031	94
1,000	A-BCEE ($c = 1000$)	0.100	0.031	0.032	0.032	94

Note: *Mean* is the mean estimated value of β where the true value is 0.1, $\overline{SEE}$ is the mean standard error estimate, *SDE* is the standard deviation of the estimates of β, $\sqrt{MSE}$ is the squared-root of the mean squared error, *CP* is the coverage probability in % of 95% confidence intervals.

One possible explanation for this underestimation is that BAC with $C_V^m(\omega)$ neglects the uncertainty associated with the choice of the hyperparameter ω.

The simulation results show that the choice of using $\omega = c\sqrt{n}$, $c \in [100, 1000]$, for A-BCEE and N-BCEE is reasonable. The results do not appear too sensitive to the choice of c in this interval. The simulation results also confirm that N-BCEE can yield lower than expected coverage probabilities. This seems to be particularly true in complex scenarios that contain many covariates, such as DGP2, DGP3 and DGP4.

Despite sometimes producing slightly biased estimates, A-BCEE performs at least as well as BAC and TBAC with $\omega = \infty$ in terms of MSE. The bias is small enough that in all simulation scenarios we considered, A-BCEE (with any c) yields appropriate coverage probability. In general, A-BCEE gives less weight to variables only associated with the exposure than BAC and TBAC (see Appendix E). In DGP1,

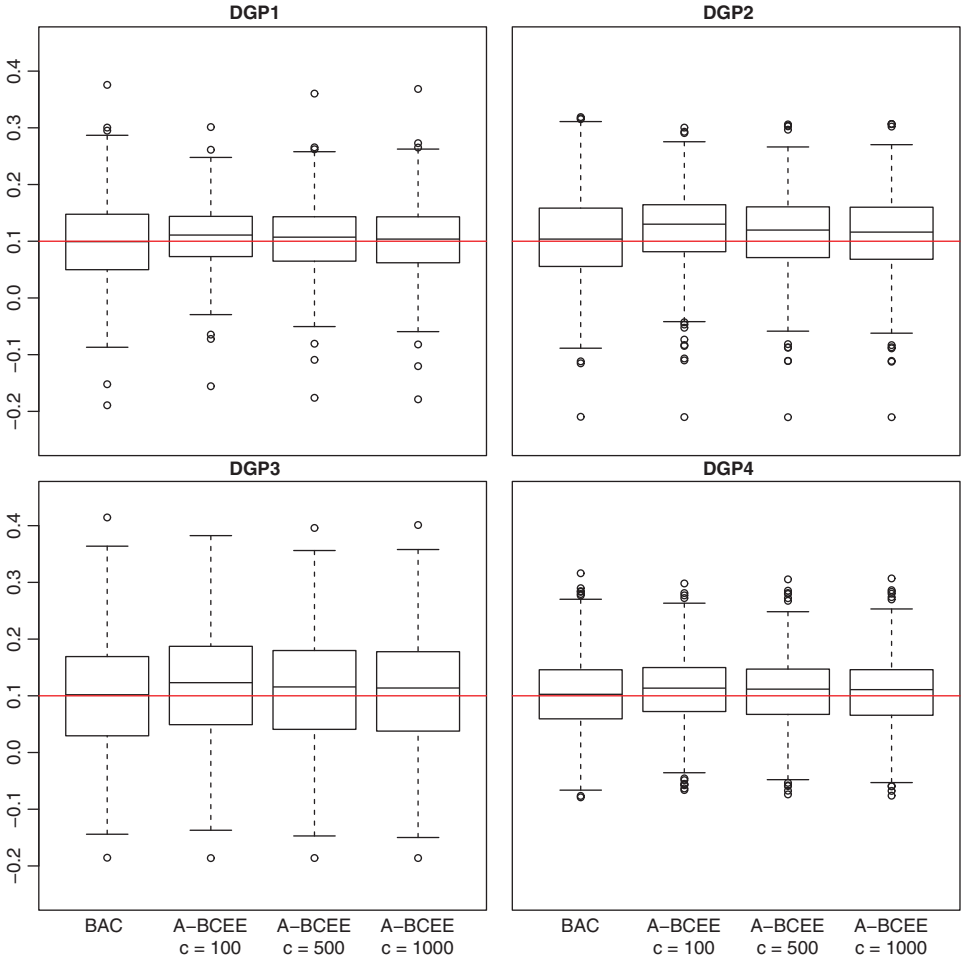

Figure 1: Comparison of the distribution of $\hat{\beta}$ obtained from BAC ($\omega = \infty$) and A-BCEE ($c = 100$, 500, and 1,000) for all four data-generating processes and a sample size $n = 200$. The red line corresponds to the true value $\beta = 0.1$.

A-BCEE outperforms BAC and TBAC with $\omega = \infty$ in terms of MSE. In DGP2 and DGP4, A-BCEE has smaller MSE than BAC and TBAC although comparatively to a lesser extent. Results are quite similar between BAC, TBAC and A-BCEE in DGP3. Note that in DGP3, the true model and the fully adjusted model have the same MSE. There is thus no possible gain in using another model than the fully adjusted model. Figure 1 illustrates the distribution of $\hat{\beta}$ obtained by using A-BCEE and BAC with $\omega = \infty$ for all four data-generating processes with $n = 200$ (analogous figures are displayed in Appendix F with $n = 600$ and $n = 1000$). This figure shows how estimates obtained with A-BCEE, despite being slightly biased, are more concentrated around the true value β than estimates obtained with BAC. Moreover, Figure 1 illustrates the bias-variance tradeoff associated with the choice of c in A-BCEE: smaller values of c, as compared to larger values of c, favor a reduced variance in the estimator of the causal effect at the cost of an increase in bias.

On the basis of these results, we hypothesized that BCEE would perform best when (1) there are some direct causes of the exposure that are strongly associated with the exposure, and (2) there exists variables that can d-separate those direct causes from the outcome. In such situations, we expect BCEE to favor models excluding those direct causes and including the d-separating variables. To verify this, we simulated data according to a fifth data-generating scenario (DGP5) which meets these two conditions. The equations for DGP5 are:

$$U_5 = U_1 + U_2 + U_3 + U_4 + \epsilon_5$$

$$X = U_1 + U_2 + U_3 + U_4 + \epsilon_X$$

$$Y = U_5 + \beta X + \epsilon_Y,$$

where $\epsilon_5, \epsilon_X, \epsilon_Y \sim$ N(0,1), all independent. In this example, BCEE's prior distribution, $P^B(\boldsymbol{\alpha}^Y)$, is devised to give non negligible prior weight to the two following sufficient outcome models: (i) the one including $\{U_1, U_2, U_3, U_4\}$, and (ii) the one including only $\{U_5\}$. However, because the marginal likelihood of the model (ii) should dominate the one of model (i) for large n, we expect the second outcome model to receive increased posterior weight as n grows. To reduce computational burden, we only considered $\beta = 0.1$ and did not estimate β with N-BCEE. The results are presented in Table 7. Those results show how under such ideal conditions, the MSE obtained by using A-BCEE is much smaller than the one obtained using the fully adjusted outcome model, BAC or TBAC. In fact, A-BCEE's MSE is similar to the MSE of the true outcome model. Moreover, Table 15 in Appendix E reveals that models including U_5, but excluding U_1, U_2, U_3 and U_4 are favored by A-BCEE, particularly for the larger sample sizes. Indeed, the marginal posterior probabilities of covariates U_1 to U_4 decrease with sample size while the posterior probability of U_5 remains at 1 for all sample sizes considered. This is as opposed to BAC and TBAC where the full model (including U_1 to U_5) receives a posterior probability of 1 at all sample sizes.

Table 7: Comparison of estimates of β obtained from the true outcome model, the fully adjusted model, BMA, BAC, TBAC and A-BCEE for 500 Monte Carlo replicates of the fifth data-generating process (DGP5).

n	Method	Estimate	$\overline{SEE}$	SDE	$\sqrt{MSE}$	CP
200	True model	0.103	0.053	0.054	0.054	92
200	Fully adjusted model	0.102	0.072	0.076	0.076	94
200	BMA	0.103	0.054	0.055	0.055	93
200	BAC ($C_V^m(\omega)$)	0.102	0.059	0.066	0.066	92
200	BAC ($\omega = \infty$)	0.102	0.072	0.076	0.076	94
200	TBAC ($\omega = \infty$)	0.102	0.072	0.076	0.076	95
200	A-BCEE ($c = 100$)	0.103	0.055	0.056	0.056	93
200	A-BCEE ($c = 500$)	0.103	0.059	0.059	0.059	94
200	A-BCEE ($c = 1000$)	0.102	0.061	0.061	0.061	95
600	True model	0.099	0.031	0.029	0.029	96
600	Fully adjusted model	0.097	0.041	0.040	0.040	96
600	BMA	0.099	0.031	0.029	0.029	96
600	BAC ($C_V^m(\omega)$)	0.097	0.034	0.036	0.036	95
600	BAC ($\omega = \infty$)	0.097	0.041	0.040	0.040	96
600	TBAC ($\omega = \infty$)	0.097	0.041	0.040	0.040	96
600	A-BCEE ($c = 100$)	0.098	0.031	0.030	0.030	96
600	A-BCEE ($c = 500$)	0.098	0.033	0.030	0.030	97
600	A-BCEE ($c = 1000$)	0.098	0.034	0.031	0.031	97
1,000	True model	0.100	0.024	0.023	0.023	95
1,000	Fully adjusted model	0.100	0.032	0.031	0.031	94
1,000	BMA	0.100	0.024	0.023	0.023	96
1,000	BAC ($C_V^m(\omega)$)	0.101	0.027	0.027	0.027	95
1,000	BAC ($\omega = \infty$)	0.100	0.032	0.031	0.031	94
1,000	TBAC ($\omega = \infty$)	0.100	0.032	0.031	0.031	95
1,000	A-BCEE ($c = 100$)	0.100	0.024	0.023	0.023	96
1,000	A-BCEE ($c = 500$)	0.100	0.025	0.023	0.023	96
1,000	A-BCEE ($c = 1000$)	0.100	0.025	0.023	0.023	96

Note: *Mean* is the mean estimated value of β where the true value is 0.1, $\overline{SEE}$ is the mean standard error estimate, *SDE* is the standard deviation of the estimates of β, $\sqrt{MSE}$ is the squared-root of the mean squared error, *CP* is the coverage probability in % of 95% confidence intervals.

5 Application: estimation of the causal effect of perceived mathematical competence on grades in mathematics

In this section we use A-BCEE to estimate the causal effect of perceived competence in mathematics (measured on a scale from 1 to 7) on self-reported grades (in %) in mathematics. We consider longitudinal data obtained from 1,430 students during their first three years of highschool. Participants lived in various regions throughout Quebec, Canada. The data were collected by postal questionnaires every year for a period of three years (time 1, time 2 and time 3). Further details can be found in Guay et al. [24].

We used measures of perceived competence in mathematics at time 2 as the exposure and grades in mathematics at time 3 as the outcome to estimate the causal effect of interest. Recall that A-BCEE requires specifying a set of potential confounding covariates that includes all direct causes of the exposure and none of its descendants. Moreover, it is beneficial that this set also includes strong predictors of the outcome. We took advantage of the longitudinal feature of the data to build the set of potential confounding covariates. Because a cause always precedes its effect in time, we constructed the set of potential confounding covariates by including variables at time 1 that were potential direct causes of perceived-competence at time 2. We also included variables at time 2 that were thought to be strong predictors of grades in mathematics at time 3.

We selected the following 26 covariates: gender, highest level of education reached by the mother, highest level of education reached by the father, perceived competence in mathematics (at time 1), perceived autonomy support from the mother, perceived autonomy support from the father, perceived autonomy support from the mathematics teacher, perceived autonomy support from friends at school, self-reported mathematics' grades, intrinsic motivation in mathematics, identified motivation in mathematics, introjected motivation in mathematics, externally regulated motivation in mathematics, victimization and sense of belonging to school. All variables except the first four were considered both at times 1 and 2.

Before applying A-BCEE on these data, we obtained some descriptive statistics. We drew scatter plots of the outcome versus the exposure and versus each potential confounding covariate to roughly verify the linearity assumption and to check for outliers. For the same reasons, we drew scatter plots of the exposure versus each potential confounding covariate. We also noticed that only 46.5% of the participants have complete information for all the selected covariates. The variables measured at time 1 have generally few missing cases (between 1.8% and 8.3%), but the variables measured at times 2 and 3 have a larger degree of missingness (between 26.4% and 36.4%). We performed multiple imputation [25] to account for the missing data, using 50 imputed datasets to ensure the power falloff is negligible [26].

We estimated the causal effect of perceived competence on grades in mathematics using the fully adjusted outcome model, A-BCEE with $\omega = c\sqrt{n}$ ($c = 100$, 500, 1,000), BAC, and TBAC (with $\omega = \infty$). Results are summarized in Table 8. The computational burden of BCEE on these data is manageable and comparable to the one of TBAC, although quite heavier than the one of BAC when using the BACprior package [27]. The approximate running times of A-BCEE, BAC, and TBAC on one imputed dataset are respectively, 22.5, 1.2, and 21.2 min on a PC with 2.4 GHz and 8 Gb RAM.

Because Step S1 of A-BCEE aims to find the direct causes of the exposure, it is reasonable to only allow covariates measured before the exposure to be selected in this step. Hence, we ran the A-BCEE algorithm a second time, but this time excluding the possibility that covariates measured at time 2 enter the exposure model. We denote this implementation of A-BCEE as A-BCEE* in Table 8.

Table 8 shows that the results from A-BCEE and A-BCEE* are very similar. This is not surprising since the marginal posterior probability of inclusion of covariates do not differ much between A-BCEE and A-BCEE* (not shown). Using A-BCEE instead of the fully adjusted model slightly decreases the standard error of estimate, between 0.3% and 3.5%, which translates in a small decrease of the 95% confidence intervals'

Table 8: Comparison of the estimated causal effect of perceived mathematical competence in mathematics on self-reported mathematics' grades.

Method	*Estimate*	*SEE*	*CI*
Fully adjusted model	0.693	0.460	(−0.208, 1.594)
BAC ($\omega = \infty$)	0.729	0.462	(−0.178, 1.635)
TBAC ($\omega = \infty$)	0.778	0.465	(−0.133, 1.690)
A-BCEE ($c = 100$)	0.807	0.451	(−0.076, 1.691)
A-BCEE ($c = 500$)	0.790	0.456	(−0.105, 1.685)
A-BCEE ($c = 1000$)	0.786	0.459	(−0.113, 1.685)
A-BCEE* ($c = 100$)	0.823	0.445	(−0.049, 1.696)
A-BCEE* ($c = 500$)	0.808	0.444	(−0.062, 1.679)
A-BCEE* ($c = 1000$)	0.803	0.444	(−0.066, 1.673)

Note: *Estimate* is the estimated causal effect, *SEE* is the standard error estimate, *CI* is a 95% confidence interval for the causal effect.

width. Moreover the standard errors of estimate for BAC and TBAC are slightly larger than the one for the fully adjusted model in this illustration. Although the point estimates appear to vary substantially between methods, the differences are small relative to the magnitude of the estimated standard errors. We conclude that perceived competence in mathematics at one point in time likely has little or no causal effect on self-reported grades in mathematics a year later.

6 Discussion

We have introduced the Bayesian causal effect estimation (BCEE) algorithm to estimate causal exposure effects in observational studies. This novel data-driven approach avoids the need to rely on the specification of a causal graph and aims to control the variability of the estimator of the exposure effect. BCEE employs a prior distribution that is motivated by a theoretical proposition embedded in the graphical framework to causal inference. We also proposed a practical implementation of BCEE, A-BCEE, that accounts for the fact that this prior distribution uses information from the data. Using simulation studies, we found that A-BCEE generally achieves at least some reduction of the MSE of the causal effect estimator as compared to the MSE generated by a fully-adjusted model approach or by other data-driven approaches to causal inference, such as BAC and TBAC, thus resulting in estimates that are overall closer to the true value. In some circumstances, the reduction of the MSE can be substantial. Moreover, confidence intervals with appropriate coverage probabilities were obtained. Hence, we believe that BCEE is a promising algorithm to perform causal inference.

Some current limitations of BCEE could be addressed in future research. The generalization to non continuous exposure variable (e.g. binary) is straightforward. Recall that the first step of BCEE aims at identifying the direct causes of the exposure. As in the normal case we have considered, classical Bayesian procedures asymptotically select the true exposure model with probability 1 when assuming X belongs to an exponential family (e.g. Bernoulli) and that an adequate parametric model is considered [12]. The generalization of BCEE to other types of outcome variables is less straightforward. One could specify a generalized linear model for the outcome of the form $g(\mathbb{E}[Y_i|X_i, \boldsymbol{U_i}]) = \delta_0 + \beta X_i + \sum_{m=1}^{M} \delta_m U_{im}$. However, unless g is the identity or the log link, such models are generally not collapsible for β over covariate U_m [28]. In other words, the true value of β, and thus its interpretation, depends on whether U_m is included or not in the outcome model, even when U_m is a not confounding covariate. In such circumstances, averaging the estimated value of β over different outcome models would not be advisable.

We think that BCEE can be particularly helpful to those working in fields where current subject-matter knowledge is sparse. To facilitate usage of the BCEE algorithm, we provide an R package named BCEE (available at http://cran.r-project.org).

Acknowledgements: Dr. Talbot has been supported by doctoral scholarships from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Fonds de recherche du Québec: Nature et Technologies. Dr. Lefebvre's is a Chercheur-Boursier from the Fonds de recherche du Québec – Santé and is also supported by NSERC. Dr. Atherton's research is supported by NSERC.

Funding: Government of Canada – Natural Sciences and Engineering Research Council of Canada, Fonds de Recherche du Québec – Santé, Fonds de Recherche du Québec – Nature et Technologies.

Appendix

A Proofs

A.1 Proof of Proposition 2.1

Proof. First, we know from Pearl [1] section 3.3.1 that a set $\boldsymbol{Z}$ is sufficient for identifying the causal effect of an exposure X on an outcome Y if (i) no descendants of X are in $\boldsymbol{Z}$ and (ii) $\boldsymbol{Z}$ blocks all back-door paths between X and Y. According to condition 1 we assume that there are no descendants of X in $\boldsymbol{Z}$. Suppose that G admits some back-door paths. All back-door paths are such that the second variable appearing in the path is a direct cause of X; the back-door paths thus have the form $X \leftarrow D_j \cdots \rightarrow Y$.

Suppose that a direct cause D_j is included in $\boldsymbol{Z}$. Then D_j (and therefore the set $\boldsymbol{Z}$) blocks all back-door paths of the form $X \leftarrow D_j \cdots \rightarrow Y$. Indeed no variable in $\boldsymbol{Z} \backslash D_j$ can reopen a path $X \leftarrow D_j \cdots \rightarrow Y$ once closed by D_j. Therefore, all back-door paths admitting a direct cause in $\boldsymbol{Z}$ are blocked by $\boldsymbol{Z}$.

It remains to show that all back-door paths for which the second variable in the path is not a direct cause included in $\boldsymbol{Z}$ are closed when condition 2b in the proposition holds. Consider $D_j \notin \boldsymbol{Z}$. Now assume that Y and D_j are d-separated by $\{X \cup \boldsymbol{Z}\}$. By the definition of d-separation, this means that every path connecting D_j to Y is blocked by $\{X \cup \boldsymbol{Z}\}$. Recall that all back-door paths associated with this D_j are of the form $X \leftarrow D_j \cdots \rightarrow Y$. Because by (2b) D_j and Y are d-separated by $\{X \cup \boldsymbol{Z}\}$ and since each subpath $D_j \cdots \rightarrow Y$ in these back-door paths does not contain the variable X, these subpaths are blocked by $\boldsymbol{Z}$. This reasoning is applied to each $D_j \notin \boldsymbol{Z}$ separately.

The proof is complete by the back-door criterion as we realize that all back-door paths, whether their D_j is contained in $\boldsymbol{Z}$ or not, are blocked by $\boldsymbol{Z}$. □

A.2 Proof of Corollary 2.1

Proof. 1. Suppose that G admits some back-door paths of the form $X \leftarrow D_j \cdots \rightarrow Y$. If D_j and Y are d-separated by $\{X \cup \boldsymbol{Z}'\}$, then by definition of d-separation all paths between D_j and Y are blocked by $\{X \cup \boldsymbol{Z}'\}$. Using the same argument as the one used in the third paragraph of the proof of Proposition 2.1, it follows that all back-door paths $X \leftarrow D_j \cdots \rightarrow Y$ are blocked by $\boldsymbol{Z}'$.

2. To prove that $\boldsymbol{Z}'$ is sufficient for estimating the causal effect of X on Y, we show that all back-door paths between X and Y are blocked by $\boldsymbol{Z}'$.

First, we consider the back-door paths that admit D_j as second variable. From point 1) of the corollary, we already know that these back-door paths are blocked.

Next, we divide the back-door paths that do not admit D_j as second variable into two categories: (1) the paths whose second variable is a $D_{j'} \in \boldsymbol{Z}'$, $j' \neq j$, and (2) the paths whose second variable is a $D_{j'} \notin \boldsymbol{Z}'$. For 1), following the same argument as in the second paragraph of the proof of Proposition 2.1, we know that all back-door paths whose second variable is a $D_{j'} \in \boldsymbol{Z}'$ are blocked.

The case where the second variable is a $D_{j'} \notin \boldsymbol{Z}'$ is more involved. Here, note that $D_{j'}$ is not in $\boldsymbol{Z}$ either since $\boldsymbol{Z}' = \boldsymbol{Z} \backslash D_j$. The fact that $\boldsymbol{Z}$ is sufficient to identify the average causal effect according to Proposition 2.1 implies that $D_{j'}$ and Y are d-separated by $\{X \cup \boldsymbol{Z}\}$. Therefore, every path between $D_{j'}$ and Y is blocked by $\{X \cup \boldsymbol{Z}\}$. For those paths that do not include D_j, it is easy to see that they are also blocked by $\{X \cup \boldsymbol{Z}'\}$. For those paths that include D_j, that is, paths of the form $D_{j'} \cdots D_j \cdots \rightarrow Y$, we know from point 1. that they are blocked in the subpaths $D_j \cdots \rightarrow Y$ by $\{X \cup \boldsymbol{Z}'\}$. Thus, every path between $D_{j'}$ and Y is blocked by $\{X \cup \boldsymbol{Z}'\}$, whether or not it includes D_j. Using the same arguments as the ones used in the third paragraph of the proof of Proposition 2.1, it follows that all back-door paths $X \leftarrow D_{j'} \cdots \rightarrow Y$ are blocked by $\{X \cup \boldsymbol{Z}'\}$. The whole reasoning is applied for each possible $D_{j'}$, according to their inclusion or exclusion in $\boldsymbol{Z}'$.

Hence, all back-door paths between X and Y in G are blocked by $\{X \cup \boldsymbol{Z}'\}$. Also, because $\boldsymbol{Z}$ is sufficient to identify the average causal effect according to Proposition 2.1, $\boldsymbol{Z}$ does not include any descendants of X and therefore $\boldsymbol{Z}'$ does not either. According to the back-door criterion, $\boldsymbol{Z}'$ is thus sufficient to identify the average causal effect and the proof is complete. □

B General conditions for the equivalence of zero regression coefficient and conditional independence

We show that the independence of Y and U_k conditional on X and $U_1, \ldots, U_{k-1}, U_{k+1}, \ldots, U_M$ is equivalent to having regression parameter δ_k associated to U_k in the linear regression of Y on X and $\boldsymbol{U}$ equal to zero under less stringent assumptions than multivariate normality for the covariates X and $\boldsymbol{U}$.

Consider the same normal linear model as in eq. (1)

$$Y_i = \delta_0 + \beta X_i + \sum_{m=1}^{M} \delta_m U_{im} + \epsilon_i,$$

where $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$. We assume that this model is correctly specified, that is, the data for Y is generated according to eq. (1) with possibly some regression coefficients set to 0. However, we make no assumptions about the distribution of variables X and $\boldsymbol{U}$. To simplify the notation, we denote $\{U_1, \ldots, U_{k-1}, U_{k+1}, \ldots U_M\}$ by $\boldsymbol{U} \backslash U_k$. We consider the case where U_k is a continuous variable. Similar arguments can be used when U_k is discrete or has a mixture distribution. Using a conditional normal distribution for Y, we have

$$f(y|x, \boldsymbol{u}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2} \left[y - \left(\delta_0 + \beta x + \sum_{m=1}^{M} \delta_m u_m \right) \right]^2 \right\}$$

and the conditional distribution of $Y|X, \boldsymbol{U} \backslash U_k$ can be calculated as

$$f(y|x, \boldsymbol{u} \backslash u_k) = \int_{-\infty}^{\infty} f(u_k|x, \boldsymbol{u} \backslash u_k) f(y|x, \boldsymbol{u}) du_k. \tag{11}$$

If $\delta_k = 0$

$$f(y|x, \boldsymbol{u}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2} \left[y - \left(\delta_0 + \beta x + \sum_{m \neq k} \delta_m u_m \right) \right]^2 \right\},$$

and the expression (11) for $f(y|x, \boldsymbol{u} \backslash u_k)$ becomes

$$\int_{-\infty}^{\infty} f(u_k|x,\boldsymbol{u}\backslash u_k)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}\left[y-\left(\delta_0+\beta x+\sum_{m\neq k}\delta_m u_m\right)\right]^2\right\}du_k$$

$$= f(y|x,\boldsymbol{u})\int_{-\infty}^{\infty} f(u_k|x,\boldsymbol{u}\backslash u_k)du_k,$$

which equals $f(y|x,\boldsymbol{u})$.

Thus if $\delta_k = 0$ in eq. (1) then $Y \perp\!\!\!\perp U_k|X,\boldsymbol{U}\backslash U_k$. Also, it is obvious that if $Y \perp\!\!\!\perp U_k|X,\boldsymbol{U}\backslash U_k$, then $\delta_k = 0$. Therefore, assuming model (1) is correctly specified we have that $Y \perp\!\!\!\perp U_k|X,\boldsymbol{U}\backslash U_k$ if and only if $\delta_k = 0$. Recall that no assumptions were made concerning the distribution of X and $\boldsymbol{U}\backslash U_k$.

C The behavior of $Q_{\boldsymbol{\alpha}^Y}$

In Figure 2 we examine how the term $Q_{\boldsymbol{\alpha}^Y}(\alpha_m^Y = 1|\alpha_m^X = 1)$ in the definition of $P^B(\boldsymbol{\alpha}^Y)$ behaves as a function of the constant c, the sample size n and the standardized parameter $\tilde{\delta}_m^{\alpha^Y} s_{U_m}/s_Y$. Specifically, we take

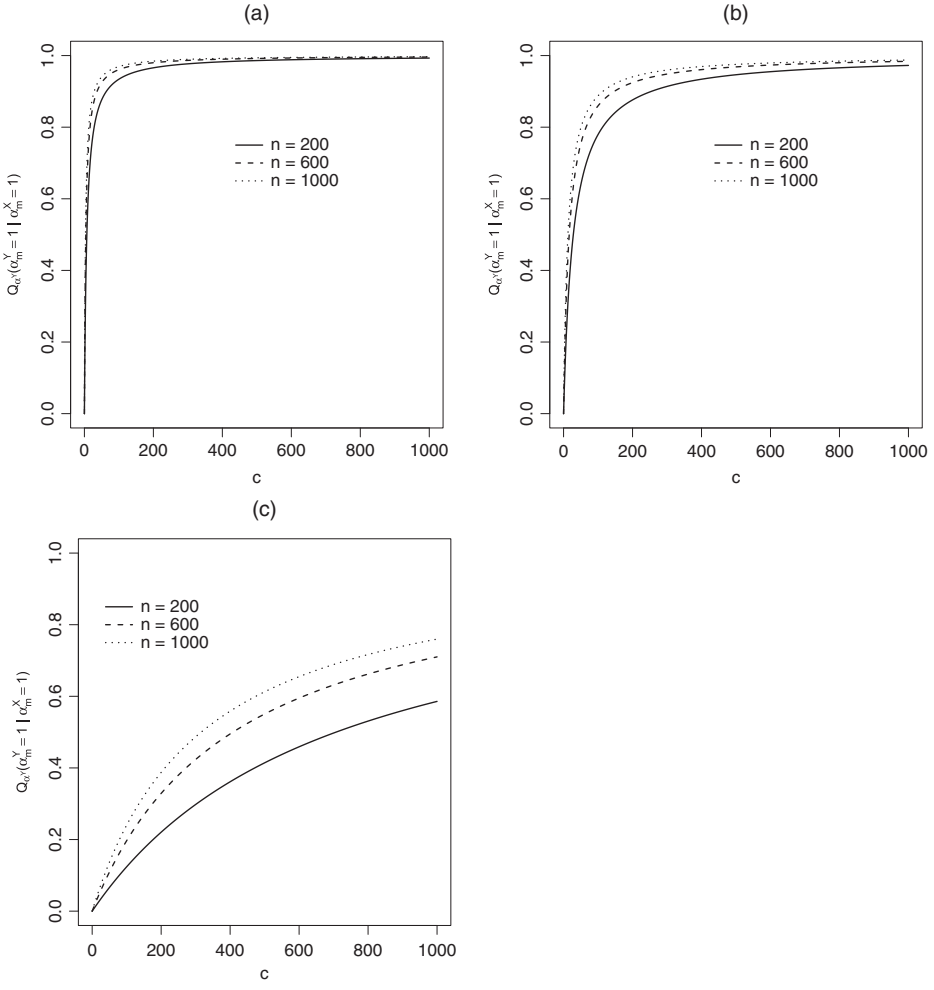

Figure 2: $Q_{\boldsymbol{\alpha}^Y}\left(\alpha_m^X = 1|\alpha_m^X = 1\right)$ with $\omega = c\sqrt{n}$ as a function of $c \in [0,1000]$ for $n = 200, 600, 1000$ and $\tilde{\delta}_m^{\alpha^Y} s_{U_m}/s_Y = 0.1$ (a), $\tilde{\delta}_m^{\alpha^Y} s_{U_m}/s_Y = 0.05$ (b) and $\tilde{\delta}_m^{\alpha^Y} s_{U_m}/s_Y = 0.01$ (c).

$\omega = c\sqrt{n}$, as suggested in Section 3.2, and plot the $Q_{\boldsymbol{\alpha}^Y}(\alpha_m^Y = 1|\alpha_m^X = 1)$ values as a function of $c \in [0, 1000]$ for fixed values of n ($n = 200, 600, 1000$) and $\tilde{\delta}_m^{\alpha^Y} s_{U_m}/s_Y$ ($\tilde{\delta}_m^{\alpha^Y} s_{U_m}/s_Y = 0.1, 0.05, 0.01$).

In Figure 2(A), we see that, for all sample sizes considered, $Q_{\boldsymbol{\alpha}^Y}(\alpha^Y = 1|\alpha^X = 1)$ rapidly increase from 0 to the limit 1 as c goes from 0 to 1,000. This behavior is desirable since a standardized regression parameter of 0.1 is non negligible. A similar pattern is seen in Figure 2(B), although the progression of $Q_{\boldsymbol{\alpha}^Y}(\alpha^Y = 1|\alpha^X = 1)$ from 0 to 1 is slightly less rapid. In Figure 2(C), the progression of $Q_{\boldsymbol{\alpha}^Y}(\alpha^Y = 1|\alpha^X = 1)$ is much slower, especially for the smaller sample size. This behavior is desirable as well since an effect size of 0.01 would usually be considered as negligible.

D Additional simulations

We now address the secondary objectives with additional simulations. The first secondary goal is to study the large, whilst finite, sample properties of BCEE. To do this, we examine four different simulation scenarios obtained by considering the four data-generating processes (DGP1, DGP2, DGP3 and DGP4) with a sample size of 10,000. Once again, for each scenario, we randomly generated 500 datasets. We estimated the average causal effect of exposure using A-BCEE and N-BCEE with $\omega = c\sqrt{n}$. Because the sample size is large and the computational burden is heavy, we considered only one value of c ($c = 500$). The results are shown in Table 9. These simulations suggest that A-BCEE and N-BCEE with $\omega = c\sqrt{n}$ unbiasedly estimate the causal effect of exposure when n is large and when BCEE's working assumptions hold (i.e., $\boldsymbol{U}$ includes all direct causes of X and the normal linear model is a correct specification for both X and Y).

Table 9: Estimates of β for N-BCEE and A-BCEE with a sample size of $n = 10{,}000$ for 500 Monte Carlo replicates of each data-generating process.

DGP	*Method*	*Mean*	$\overline{SEE}$	*SDE*	$\sqrt{MSE}$	*CP*
1	N-BCEE ($c = 500$)	0.100	0.0068	0.0067	0.0067	97
1	A-BCEE ($c = 500$)	0.100	0.0069	0.0066	0.0066	97
2	N-BCEE ($c = 500$)	0.100	0.0074	0.0075	0.0075	96
2	A-BCEE ($c = 500$)	0.100	0.0079	0.0075	0.0075	98
3	N-BCEE ($c = 500$)	0.100	0.0140	0.0141	0.0141	95
3	A-BCEE ($c = 500$)	0.100	0.0140	0.0141	0.0141	95
4	N-BCEE ($c = 500$)	0.099	0.0083	0.0084	0.0084	96
4	A-BCEE ($c = 500$)	0.099	0.0089	0.0086	0.0086	97

Note: *DGP* is the data-generating process, *Mean* is the mean estimated value of β where the true value is 0.1, $\overline{SEE}$ is the mean standard error estimate, *SDE* is the standard deviation of the estimates of β, $\sqrt{MSE}$ is the squared-root of the mean squared error, *CP* is the coverage probability in % of 95% confidence intervals.

The second secondary objective is to study the performance of B-BCEE to correct the confidence intervals of N-BCEE. Since this bootstrapped implementation is very computationally intensive, we only considered two simulation scenarios: DGP1 and DGP4 with a sample size of 200. In this case, only 100 datasets were generated for each scenario. We estimated the causal effect of exposure using the fully adjusted model, the true outcome model, BMA, BAC, TBAC, A-BCEE, N-BCEE and B-BCEE. For A-BCEE, N-BCEE and B-BCEE we took $\omega = c\sqrt{n}$ with $c = 500$. For B-BCEE we performed 200 bootstrap resamplings and considered an estimate with and without a bias correction. The results are presented in Table 10. We find that the non-parametric bootstrap implementation of BCEE yields correct estimates of the standard error of estimate and correct coverage probabilities. However, B-BCEE does not seem to be as efficient nor as practical as A-BCEE.

Table 10: Comparison of estimates of β obtained from the true model, the fully adjusted model, BMA, BAC, TBAC, N-BCEE, A-BCEE and B-BCEE for the first and fourth data-generating processes (DGP1 and DGP4). Sample size is $n = 200$, 100 datasets were generated for each data-generating process. For B-BCEE, 200 bootstrap resamplings were performed.

DGP	*Method*	*Mean*	$\overline{SEE}$	*SDE*	$\sqrt{MSE}$	*CP*
1	True model	0.105	0.045	0.053	0.053	93
1	Fully adjusted model	0.104	0.072	0.075	0.075	94
1	BMA	0.121	0.048	0.050	0.054	93
1	BAC ($\omega = \infty$)	0.104	0.072	0.075	0.075	94
1	TBAC ($\omega = \infty$)	0.104	0.072	0.075	0.074	93
1	N-BCEE ($c = 500$)	0.112	0.056	0.063	0.064	92
1	A-BCEE ($c = 500$)	0.111	0.062	0.061	0.062	96
1	B-BCEE ($c = 500$, no bias corr.)	0.112	0.067	0.063	0.064	96
1	B-BCEE ($c = 500$, w/bias corr.)	0.107	0.067	0.066	0.066	96
4	True model	0.111	0.063	0.063	0.063	96
4	Fully adjusted model	0.106	0.072	0.064	0.064	96
4	BMA	0.120	0.060	0.051	0.055	96
4	BAC ($\omega = \infty$)	0.105	0.072	0.064	0.064	96
4	TBAC ($\omega = \infty$)	0.106	0.071	0.064	0.063	96
4	N-BCEE ($c = 500$)	0.108	0.064	0.061	0.061	97
4	A-BCEE ($c = 500$)	0.109	0.068	0.060	0.060	96
4	B-BCEE ($c = 500$, no bias corr.)	0.108	0.071	0.061	0.061	97
4	B-BCEE ($c = 500$, w/bias corr.)	0.107	0.071	0.062	0.062	98

Note: *DGP* is the data-generating process, *Mean* is the mean estimated value of β where the true value is 0.1, $\overline{SEE}$ is the mean standard error estimate, *SDE* is the standard deviation of the estimates of β, $\sqrt{MSE}$ is the squared-root of the mean squared error, *CP* is the coverage probability in % of 95% confidence intervals.

E Marginal posterior probabilities of inclusion of potential confounding covariates

Table 11: Marginal posterior probability of inclusion of potential confounding covariate U_m, $m = 1, \ldots, 5$, for BMA, BAC, TBAC, N-BCEE, and A-BCEE for 500 Monte Carlo replicates of the first data-generating process (DGP1). The covariates included in the true outcome model are $\{U_3, U_4, U_5\}$.

n	*Method*	U_1	U_2	U_3	U_4	U_5
200	BMA	0.11	0.11	1.00	0.18	1.00
200	BAC ($C_V^m(\omega)$)	0.35	0.35	1.00	0.41	1.00
200	BAC ($\omega = \infty$)	1.00	1.00	1.00	1.00	1.00
200	TBAC ($\omega = \infty$)	1.00	1.00	1.00	1.00	1.00
200	N-BCEE ($c = 100$)	0.19	0.24	1.00	0.37	1.00
200	N-BCEE ($c = 500$)	0.36	0.41	1.00	0.54	1.00
200	N-BCEE ($c = 1000$)	0.44	0.49	1.00	0.61	1.00
200	A-BCEE ($c = 100$)	0.29	0.35	1.00	0.44	1.00
200	A-BCEE ($c = 500$)	0.51	0.56	1.00	0.63	1.00
200	A-BCEE ($c = 1000$)	0.60	0.64	1.00	0.70	1.00
600	BMA	0.06	0.06	1.00	0.24	1.00
600	BAC ($C_V^m(\omega)$)	0.32	0.33	1.00	0.47	1.00
600	BAC ($\omega = \infty$)	1.00	1.00	1.00	1.00	1.00
600	TBAC ($\omega = \infty$)	1.00	1.00	1.00	1.00	1.00
600	N-BCEE ($c = 100$)	0.11	0.15	1.00	0.44	1.00

(*continued*)

Table 11: (*continued*)

n	*Method*	U_1	U_2	U_3	U_4	U_5
600	N-BCEE ($c = 500$)	0.22	0.30	1.00	0.60	1.00
600	N-BCEE ($c = 1000$)	0.28	0.37	1.00	0.66	1.00
600	A-BCEE ($c = 100$)	0.15	0.21	1.00	0.45	1.00
600	A-BCEE ($c = 500$)	0.34	0.42	1.00	0.63	1.00
600	A-BCEE ($c = 1000$)	0.44	0.51	1.00	0.70	1.00
1,000	BMA	0.05	0.04	1.00	0.33	1.00
1,000	BAC ($C_V^m(\omega)$)	0.38	0.37	1.00	0.61	1.00
1,000	BAC ($\omega = \infty$)	1.00	1.00	1.00	1.00	1.00
1,000	TBAC ($\omega = \infty$)	1.00	1.00	1.00	1.00	1.00
1,000	N-BCEE ($c = 100$)	0.09	0.11	1.00	0.55	1.00
1,000	N-BCEE ($c = 500$)	0.19	0.22	1.00	0.69	1.00
1,000	N-BCEE ($c = 1000$)	0.25	0.28	1.00	0.74	1.00
1,000	A-BCEE ($c = 100$)	0.12	0.15	1.00	0.54	1.00
1,000	A-BCEE ($c = 500$)	0.30	0.34	1.00	0.70	1.00
1,000	A-BCEE ($c = 1000$)	0.39	0.44	1.00	0.75	1.00

Table 12: Marginal posterior probability of inclusion of potential confounding covariate U_m, $m = 1, ..., 5, 7, 8$, for BMA, BAC, TBAC, N-BCEE, and A-BCEE for 500 Monte Carlo replicates of the second data-generating process (DGP2). The covariates included in the true outcome model are $\{U_3, U_4, U_5\}$.

n	*Method*	U_1	U_2	U_3	U_4	U_5	U_7	U_8
200	BMA	0.14	0.12	0.20	0.18	0.31	0.10	0.10
200	BAC ($C_V^m(\omega)$)	0.44	0.41	0.48	0.18	0.32	0.12	0.11
200	BAC ($\omega = \infty$)	1.00	1.00	1.00	0.22	0.40	0.15	0.15
200	TBAC ($\omega = \infty$)	1.00	1.00	1.00	0.18	0.29	0.14	0.14
200	N-BCEE ($c = 100$)	0.45	0.39	0.55	0.26	0.34	0.14	0.14
200	N-BCEE ($c = 500$)	0.64	0.58	0.72	0.28	0.35	0.17	0.17
200	N-BCEE ($c = 1000$)	0.71	0.66	0.78	0.29	0.36	0.18	0.18
200	A-BCEE ($c = 100$)	0.64	0.56	0.66	0.19	0.30	0.13	0.13
200	A-BCEE ($c = 500$)	0.79	0.73	0.81	0.19	0.30	0.14	0.14
200	A-BCEE ($c = 1000$)	0.84	0.79	0.85	0.19	0.30	0.14	0.14
600	BMA	0.12	0.09	0.39	0.25	0.64	0.08	0.07
600	BAC ($C_V^m(\omega)$)	0.63	0.59	0.75	0.22	0.62	0.09	0.07
600	BAC ($\omega = \infty$)	1.00	1.00	1.00	0.21	0.57	0.07	0.06
600	TBAC ($\omega = \infty$)	1.00	1.00	1.00	0.19	0.53	0.09	0.08
600	N-BCEE ($c = 100$)	0.37	0.26	0.73	0.28	0.65	0.09	0.08
600	N-BCEE ($c = 500$)	0.56	0.45	0.84	0.29	0.63	0.11	0.09
600	N-BCEE ($c = 1000$)	0.64	0.54	0.88	0.29	0.61	0.12	0.10
600	A-BCEE ($c = 100$)	0.56	0.43	0.76	0.22	0.59	0.08	0.07
600	A-BCEE ($c = 500$)	0.74	0.63	0.86	0.21	0.56	0.09	0.08
600	A-BCEE ($c = 1000$)	0.80	0.70	0.90	0.20	0.56	0.09	0.08
1,000	BMA	0.12	0.08	0.55	0.33	0.82	0.06	0.06
1,000	BAC ($C_V^m(\omega)$)	0.69	0.66	0.86	0.28	0.75	0.06	0.06
1,000	BAC ($\omega = \infty$)	1.00	1.00	1.00	0.25	0.71	0.05	0.05
1,000	TBAC ($\omega = \infty$)	1.00	1.00	1.00	0.25	0.69	0.07	0.07
1,000	N-BCEE ($c = 100$)	0.34	0.23	0.83	0.34	0.78	0.07	0.07
1,000	N-BCEE ($c = 500$)	0.50	0.39	0.90	0.34	0.76	0.08	0.08
1,000	N-BCEE ($c = 1000$)	0.58	0.47	0.92	0.34	0.75	0.08	0.08
1,000	A-BCEE ($c = 100$)	0.50	0.37	0.84	0.28	0.75	0.06	0.07
1,000	A-BCEE ($c = 500$)	0.69	0.57	0.91	0.27	0.72	0.07	0.07
1,000	A-BCEE ($c = 1000$)	0.75	0.65	0.93	0.26	0.71	0.07	0.07

Table 13: Marginal posterior probability of inclusion of potential confounding covariate U_m, $m = 1, \ldots, 4$, for BMA, BAC, TBAC, N-BCEE, and A-BCEE for 500 Monte Carlo replicates of the third data-generating process (DGP3). The covariates included in the true outcome model are $\{U_1, U_2\}$.

n	*Method*	U_1	U_2	U_3	U_4
200	BMA	0.17	0.26	0.08	0.09
200	BAC ($C_V^m(\omega)$)	0.48	0.29	0.09	0.10
200	BAC ($\omega = \infty$)	1.00	0.28	0.08	0.10
200	TBAC ($\omega = \infty$)	1.00	0.30	0.14	0.15
200	N-BCEE ($c = 100$)	0.57	0.32	0.14	0.16
200	N-BCEE ($c = 500$)	0.75	0.34	0.17	0.19
200	N-BCEE ($c = 1000$)	0.80	0.35	0.18	0.20
200	A-BCEE ($c = 100$)	0.62	0.30	0.13	0.14
200	A-BCEE ($c = 500$)	0.77	0.30	0.13	0.14
200	A-BCEE ($c = 1000$)	0.83	0.30	0.13	0.15
600	BMA	0.29	0.50	0.07	0.05
600	BAC ($C_V^m(\omega)$)	0.75	0.53	0.07	0.06
600	BAC ($\omega = \infty$)	1.00	0.52	0.07	0.05
600	TBAC ($\omega = \infty$)	1.00	0.51	0.09	0.08
600	N-BCEE ($c = 100$)	0.68	0.52	0.10	0.08
600	N-BCEE ($c = 500$)	0.82	0.53	0.11	0.10
600	N-BCEE ($c = 1000$)	0.87	0.53	0.12	0.10
600	A-BCEE ($c = 100$)	0.68	0.51	0.09	0.07
600	A-BCEE ($c = 500$)	0.81	0.51	0.09	0.08
600	A-BCEE ($c = 1000$)	0.85	0.51	0.09	0.08
1,000	BMA	0.41	0.68	0.05	0.05
1,000	BAC ($C_V^m(\omega)$)	0.86	0.70	0.06	0.05
1,000	BAC ($\omega = \infty$)	0.99	0.69	0.06	0.05
1,000	TBAC ($\omega = \infty$)	1.00	0.68	0.07	0.07
1,000	N-BCEE ($c = 100$)	0.76	0.69	0.07	0.07
1,000	N-BCEE ($c = 500$)	0.88	0.69	0.08	0.08
1,000	N-BCEE ($c = 1000$)	0.91	0.69	0.09	0.09
1,000	A-BCEE ($c = 100$)	0.75	0.68	0.07	0.07
1,000	A-BCEE ($c = 500$)	0.85	0.68	0.07	0.07
1,000	A-BCEE ($c = 1000$)	0.89	0.68	0.07	0.07

Table 14: Marginal posterior probability of inclusion of potential confounding covariate U_m, $m = 1, \ldots, 6$, for BMA, BAC, TBAC, N-BCEE, and A-BCEE for 500 Monte Carlo replicates of the fourth data-generating process (DGP4). The covariates included in the true outcome model are $\{U_4, U_5, U_6\}$.

n	*Method*	U_1	U_2	U_3	U_4	U_5	U_6
200	BMA	0.15	0.14	0.13	0.22	1.00	1.00
200	BAC ($C_V^m(\omega)$)	0.32	0.15	0.31	0.22	1.00	1.00
200	BAC ($\omega = \infty$)	0.97	0.23	0.98	0.24	1.00	1.00
200	TBAC ($\omega = \infty$)	0.87	0.23	0.99	0.25	1.00	1.00
200	N-BCEE ($c = 100$)	0.36	0.17	0.36	0.29	1.00	1.00
200	N-BCEE ($c = 500$)	0.53	0.22	0.57	0.32	1.00	1.00
200	N-BCEE ($c = 1000$)	0.60	0.25	0.66	0.33	1.00	1.00
200	A-BCEE ($c = 100$)	0.50	0.19	0.48	0.23	1.00	1.00
200	A-BCEE ($c = 500$)	0.65	0.21	0.66	0.24	1.00	1.00
200	A-BCEE ($c = 1000$)	0.70	0.21	0.73	0.24	1.00	1.00
600	BMA	0.14	0.10	0.08	0.35	1.00	1.00
600	BAC ($C_V^m(\omega)$)	0.44	0.16	0.41	0.29	1.00	1.00
600	BAC ($\omega = \infty$)	0.99	0.24	1.00	0.27	1.00	1.00

(*continued*)

Table 14: (*continued*)

n	*Method*	U_1	U_2	U_3	U_4	U_5	U_6
600	TBAC ($\omega = \infty$)	0.96	0.23	1.00	0.26	1.00	1.00
600	N-BCEE ($c = 100$)	0.33	0.12	0.22	0.41	1.00	1.00
600	N-BCEE ($c = 500$)	0.50	0.18	0.41	0.41	1.00	1.00
600	N-BCEE ($c = 1000$)	0.58	0.21	0.50	0.41	1.00	1.00
600	A-BCEE ($c = 100$)	0.48	0.16	0.32	0.31	1.00	1.00
600	A-BCEE ($c = 500$)	0.67	0.19	0.53	0.29	1.00	1.00
600	A-BCEE ($c = 1000$)	0.73	0.20	0.61	0.28	1.00	1.00
1,000	BMA	0.13	0.09	0.07	0.52	1.00	1.00
1,000	BAC ($C_V^m(\omega)$)	0.48	0.18	0.42	0.42	1.00	1.00
1,000	BAC ($\omega = \infty$)	0.99	0.30	1.00	0.35	1.00	1.00
1,000	TBAC ($\omega = \infty$)	0.99	0.30	1.00	0.34	1.00	1.00
1,000	N-BCEE ($c = 100$)	0.30	0.12	0.19	0.57	1.00	1.00
1,000	N-BCEE ($c = 500$)	0.45	0.19	0.36	0.56	1.00	1.00
1,000	N-BCEE ($c = 1000$)	0.53	0.22	0.44	0.54	1.00	1.00
1,000	A-BCEE ($c = 100$)	0.47	0.18	0.26	0.44	1.00	1.00
1,000	A-BCEE ($c = 500$)	0.67	0.23	0.47	0.40	1.00	1.00
1,000	A-BCEE ($c = 1000$)	0.73	0.25	0.56	0.38	1.00	1.00

Table 15: Marginal posterior probability of inclusion of potential confounding covariate U_m, $m = 1, \ldots, 5$, for BMA, BAC, TBAC, and A-BCEE for 500 Monte Carlo replicates of the fifth data-generating process (DGP5). The true outcome model includes only U_5.

n	*Method*	U_1	U_2	U_3	U_4	U_5
200	BMA	0.11	0.12	0.12	0.11	1.00
200	BAC ($C_V^m(\omega)$)	0.42	0.42	0.42	0.42	1.00
200	BAC ($\omega = \infty$)	1.00	1.00	1.00	1.00	1.00
200	TBAC ($\omega = \infty$)	1.00	1.00	1.00	1.00	1.00
200	A-BCEE ($c = 100$)	0.21	0.22	0.22	0.21	1.00
200	A-BCEE ($c = 500$)	0.45	0.45	0.46	0.45	1.00
200	A-BCEE ($c = 1000$)	0.55	0.55	0.56	0.55	1.00
600	BMA	0.08	0.07	0.08	0.08	1.00
600	BAC ($C_V^m(\omega)$)	0.42	0.41	0.42	0.42	1.00
600	BAC ($\omega = \infty$)	1.00	1.00	1.00	1.00	1.00
600	TBAC ($\omega = \infty$)	1.00	1.00	1.00	1.00	1.00
600	A-BCEE ($c = 100$)	0.12	0.11	0.12	0.12	1.00
600	A-BCEE ($c = 500$)	0.30	0.29	0.31	0.31	1.00
600	A-BCEE ($c = 1000$)	0.40	0.40	0.41	0.41	1.00
1,000	BMA	0.06	0.07	0.06	0.06	1.00
1,000	BAC ($C_V^m(\omega)$)	0.41	0.41	0.41	0.40	1.00
1,000	BAC ($\omega = \infty$)	1.00	1.00	1.00	1.00	1.00
1,000	TBAC ($\omega = \infty$)	1.00	1.00	1.00	1.00	1.00
1,000	A-BCEE ($c = 100$)	0.08	0.09	0.09	0.08	1.00
1,000	A-BCEE ($c = 500$)	0.22	0.23	0.23	0.22	1.00
1,000	A-BCEE ($c = 1000$)	0.32	0.33	0.33	0.32	1.00

F Comparison of the distribution of $\hat{\beta}$ obtained from A-BCEE and BAC

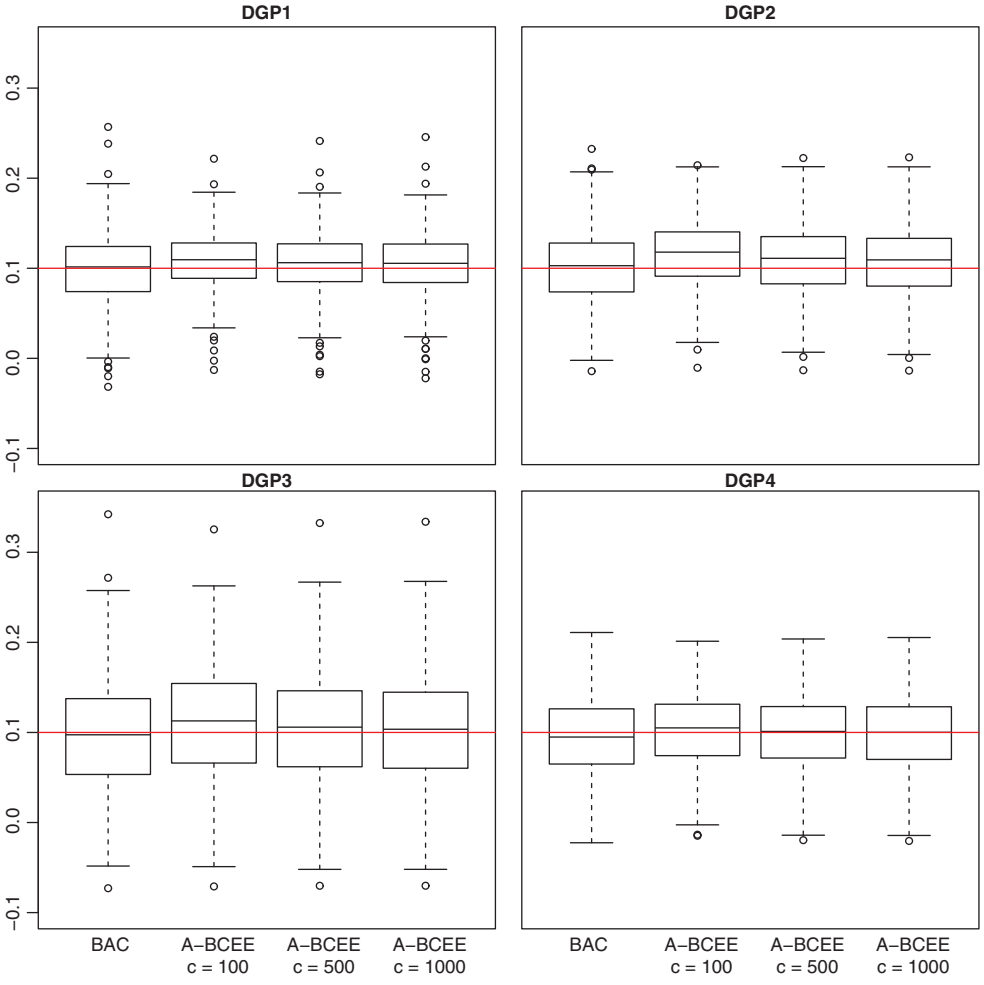

Figure 3: Comparison of the distribution of $\hat{\beta}$ obtained from BAC ($\omega = \infty$) and A-BCEE ($c = 100$, 500, and 1,000) for all four data-generating processes and a sample size $n = 600$. The red line corresponds to the true value $\beta = 0.1$.

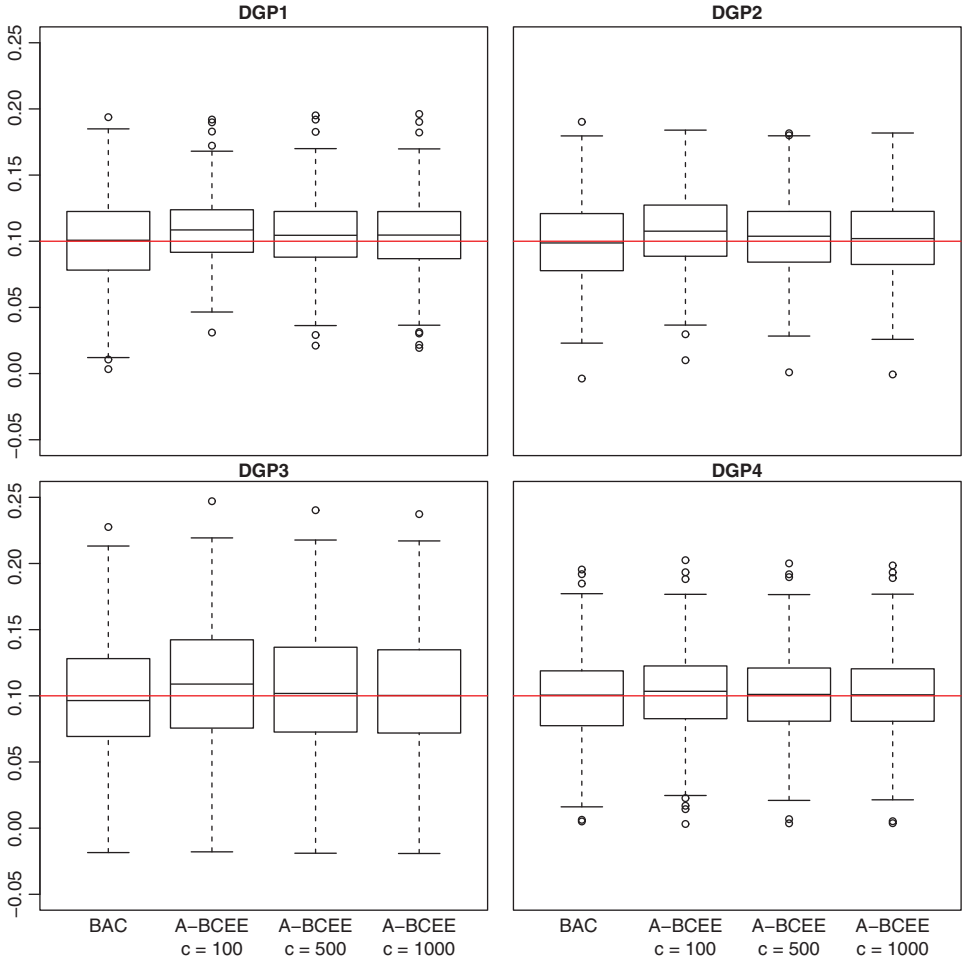

Figure 4: Comparison of the distribution of $\hat{\beta}$ obtained from BAC ($\omega = \infty$) and A-BCEE ($c = 100$, 500, and 1,000) for all four data-generating processes and a sample size $n = 1000$. The red line corresponds to the true value $\beta = 0.1$.

References

1. Pearl J. Causality: models, reasoning, and inference, 2nd ed. New York: Cambridge University Press, 2009.
2. Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: a tutorial. Stat Sci 1999;14:382–401.
3. Raftery AE, Madigan D, Hoeting JA. Bayesian model averaging for linear regression models. J Am Stat Assoc 1997;92:179–91.
4. Raftery AE, Zheng Y. Discussion: performance of Bayesian model averaging. J Am Stat Assoc 2003;98:931–8.
5. Crainiceanu CM, Dominici F, Parmigiani G. Adjustment uncertainty in effect estimation. Biometrika 2008;95:635–51.
6. Wang C, Parmigiani G, Dominici F. Bayesian effect estimation accounting for adjustment uncertainty. Biometrics 2012;68:661–71.
7. Lefebvre G, Atherton J, Talbot D. The effect of the prior distribution in the Bayesian adjustment for confounding algorithm. Comput Stat Data Anal 2014;70:227–40.
8. Vansteelandt S. Discussions. Biometrics 2012;68:675–8.
9. Wang C, Parmigiani G, Dominici F. Rejoinder: Bayesian effect estimation accounting for adjustment uncertainty. Biometrics 2012;68:680–6.
10. VanderWeele TJ, Shpitser I. A new criterion for confounder selection. Biometrics 2011;67:1406–13.
11. Baba K, Shibata R, Sibuya M. Partial correlation and conditional correlation as measures of conditional independence. Aust NZ J Stat 2004;46:657–64.

12. Haughton DMA. On the choice of a model to fit data from an exponential family. Ann Stat 1988;16:342–55.
13. Wasserman L. Bayesian model selection and model averaging. J Math Psychol 2000;44:92–107.
14. Clyde M. Subjective and objective Bayesian statistics, 2nd ed. S. James Press, New Jersey: Wiley-Interscience, 2003
15. Yuan K-H, Chan W. Biases and standard errors of standardized regression coefficients. Psychometrika 2011;76:670–90.
16. Agresti A. Categorical data analysis, 3rd ed. New Jersey: John Wiley & Sons, 2014.
17. Dawid AP. On the limiting normality of posterior distributions. Proc Cambridge Philos Soc 1970;67:625–33.
18. Walker AM. On the asymptotic behavior of posterior distributions. J R Stat Soc Ser B 1969;31:80–8.
19. Madigan D, York J, Allard D. Bayesian graphical models for discrete data. Int Stat Rev 1995;63:215–32.
20. Carlin BP, Gelfand AE. Approaches for empirical Bayes confidence intervals. J Am Stat Assoc 1990;85:105–14.
21. Carlin BP, Louis TA. Bayes and empirical Bayes methods for data analysis, 2nd ed. Boca Raton: Chapman & Hall/CRC, 2000.
22. Laird NM, Louis TA. Empirical Bayes confidence intervals based on bootstrap samples. J Am Stat Assoc 1987;82:739–50.
23. Morgan SL, Winship C. Counterfactuals and causal inference: methods and principles for social research. New York: Cambridge University Press, 2007
24. Guay F, Larose S, Ratelle C, Sénécal C, Vallerand RJ, Vitaro F (2011): Mes amis, mes parents et mes professeurs: Une analyses comparée de leurs effets respectifs sur la motivation, la réussite, l'orientation et la persévérance scolaires. Technical Report 2007-PE-118485, Fonds de recherche du Québec - Société et culture, Québec.
25. Little RJ, Rubin DB. Statistical analysis with missing data, 2nd ed. New Jersey: John Wiley & Sons, 2002.
26. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prev Sci 2007;8:206–13.
27. Talbot D, Lefebvre G, Atherton J (2014): BACprior: Choice of the Hyperparameter Omega in the Bayesian Adjustment for Confounding (BAC) Algorithm. Available at: http://CRAN.R-project.org/package = BACprior, R package version 2.0.
28. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. Stat Sci 1999;14:29–46.