Bayesian Modeling of Treatment Response Data from Eligibility Designs

Siddhartha Chib
John M. Olin School of Business
Campus Box 1133
Washington University in St. Louis
1 Brookings Dr., St. Louis, MO 63130
email: chib@wustl.edu

Liana Jacobi
Department of Economics
Campus Box 1208
Washington University in St. Louis
1 Brookings Dr., St. Louis, MO 63130
email: jacobi@economics.wustl.edu

September 2005

Abstract

This paper is concerned with the problem of isolating the effect of a randomly assigned binary treatment on an outcome when the actual treatment intake in the treatment arm is not necessarily the same as the assignment; in the control arm, however, assignment and intake are the same. In general, the analysis of data from such designs is not straightforward because of likely observed and unobserved confounders. In one approach for dealing with this problem, developed by Sommer and Zeger (1991), the unobserved confounder is modeled as a discrete indicator variable representing compliance with the assignment. In this paper, this approach is first reformulated in Bayesian terms, without involving the counterfactual outcomes in the modeling and estimation. The approach is then contrasted with a new approach in which the confounder is modeled as a continuous variable without reference to compliance. This opens up the possibility of fitting eligibility design data under each framework, leading to a sort of global sensitivity analysis. We also show that the two frameworks can be compared via marginal likelihoods that are estimated by the method of Chib (1995). Our studies show that, when the data come from the discrete confounder model, the continuous confounder model is preferred when the sample size is small but less preferred as the sample size is increased. On the other hand, when the data come from the continuous confounder model, the discrete confounder model is rarely the better supported model.

Keywords: Confounding; Eligibility Designs; Heterogeneity; Instrumental variable; Marginal likelihood; Markov Chain Monte Carlo; Non-Compliance; Non-randomly assigned treatment; Potential outcomes; Treatment Effect.

1 Introduction

This paper is concerned with the problem of isolating the effect of a randomly assigned binary treatment on an outcome of interest when the actual treatment intake in the treatment arm is not necessarily the same as the assignment; in the control arm, however, assignment and intake are the same. The equality of treatment assignment and intake in the control arm is the defining feature of this experimental design and distinguishes it from randomized trials where treatment intake is determined by the subject in both the treatment and the control arms. We refer to the former design as an eligibility design, since actual intake of the treatment rests

on prior eligibility, i.e., randomization into the treatment arm, and the latter design as a non-eligibility design. Eligibility designs often arise in the context of clinical trials and social science experiments. For example, in the study of Sommer and Zeger (1991), Vitamin A supplementation was randomly given to children in rural Indonesia in an effort to determine its effect on child mortality. Children assigned to the treatment arm, however, did not necessarily receive the supplementation.

In general, the analysis of data from eligibility trials is not straightforward. The principal reason for this is that the factors that determine the intake (beyond assignment) are likely to be correlated with the outcome. In such a case, the effect of the treatment on the outcome cannot be obtained, for example, by regressing the outcome on the treatment. If the factors, or confounders, are observed, then it is possible in principle to make adjustments and to obtain the effect of the treatment by various techniques. However, when the confounders are unobserved, as for example due to unmeasured, unmeasurable, or hidden variables, it is not possible to find the treatment effect without auxiliary, untestable assumptions about the unobserved confounders and a model of their effect on the intake and the outcome. The challenge therefore is to isolate assumptions and models that are appropriate and meaningful.

This challenge has been addressed in a number of recent papers. Sommer and Zeger (1991), for example, provide an approach that has been actively pursued in the subsequent literature. In their approach, the unobserved confounder is a discrete quantity representing compliance, where compliance is a subject-specific attribute. In the nomenclature of Imbens and Rubin (1997), non-compliers are of two types, never-takers and always-takers. In these terms, subjects in the control arm are of two types, compliers or never-takers, those in the treatment arm who forgo the treatment are never-takers, whereas those in the treatment arm who take the treatment are compliers or always-takers. The problem is that in the eligibility design always-takers cannot be identified. Provided we are willing to make the (non-testable) assumption that there are no always-takers in the population or the sample it is possible to infer the compliers in the control arm and to compare them with those in the other arm.

The discrete confounder approach for eligibility trials has been further pursued by Albert 2003, Ten Have et al (2003), Frangakis and Rubin (1999), Levy et al (2004), Mealli et al (2004), and Yau and Little (2001), mostly from a non-Bayesian perspective. Generally, these papers deal with binary outcomes and do not include covariates beyond assignment. These points are

worth noting because in this paper we reformulate the discrete confounder approach in fully Bayesian terms for continuous outcomes and allow for a full set of covariates. We develop the appropriate Markov chain Monte Carlo (MCMC) approach for making inferences and test the performance of the methods in simulation experiments. A key aspect of our formulation is that it does not need the missing counterfactuals in either the modeling or subsequent estimation of the model.

The latter reformulation is not the main objective of the paper but is needed as part of our larger objective - to compare the virtues of the discrete confounder approach with a new alternative approach in which the confounder is modeled without reference to compliance. We refer to this alternative approach as the continuous confounder approach.

This continuous confounder approach, which has not been considered before in this context, is a natural outgrowth of our previous work on non-eligibility designs (Chib, 2004 and Chib and Jacobi, 2005). But unlike our previous work, the framework here is different on account of the fact that intake in the control arm is non-stochastic. Although this change may at first glance appear to be minor, the resulting analysis is substantially different. In particular, in the control arm, the modeling only requires the marginal distribution of the outcome. In addition, the fact that subjects who take the treatment are only observed under one level of the assignment makes it difficult to estimate the extent of confounding in the intake state. Taken together, these features necessitate a different prior-posterior analysis and a different MCMC approach for summarizing the posterior distribution. In experiments we show that our approach is efficient in learning about the parameters, and in separating out the treatment effect in the presence of various levels of confounding.

Needless to say, our work opens up the possibility of fitting and comparing eligibility design data under each framework, leading to a sort of global sensitivity analysis that is likely to be quite useful in practice. To pin things down, this support for the models is computed via the Bayes factor, which we show can be computed from the MCMC output by the approach of Chib (1995). We also examine in simulation studies to what extent each framework is supported by the data when the data comes from one or the other model. Our studies show that, when the data come from the discrete confounder model, the continuous confounder model is preferred when the sample size is small but less preferred as the sample size is increased. On the other hand, when the data come from the continuous confounder model, the discrete confounder model

is rarely the better supported model.

The remainder of the paper is organized as follows. In Section 2 we reformulate the discrete confounder approach for eligibility trials in Bayesian terms and supply an inferential methodology for summarizing the posterior distribution and extracting treatment effects. We also test the performance of these methods in simulation experiments. In Section 3 we develop the continuous confounder approach and supply the relevant details of the prior-posterior analysis along with simulation evidence on the performance of our inferential procedures. Section 4 is concerned with the problem of model comparisons through marginal likelihoods and Bayes factors. Section 5 focuses on a real data example in which the question is to find the effect of job-training for the unemployed on subsequent depression scores. We conclude the paper with remarks in Section 6.

2 Preliminaries: Discrete Confounder Approach

2.1 Modeling

We begin our analysis by reformulating the discrete confounder approach in fully Bayesian terms, with allowance for a full set of covariates. We provide this reformulation to set the stage for an alternative approach to confounding that we present in Section 3. Suppose that we have a random sample of n individuals. Let i denote the ith subject in the sample, and let $\mathbf{y} = (y_1, ..., y_n)$ and $\mathbf{x} = (x_1, ..., x_n)$ denote the random sample of n observations on the continuous outcome and the binary treatment intake. Also let $\mathbf{w}_i : p \times 1$ indicate a set of covariates, and let $z_i \in \{0, 1\}$ indicate the random treatment assignment (or alternatively, the instrumental variable). The objective is to model for each individual the joint density of the observed data

$$p(y_i, x_i = j|\mathbf{w}_i, z_i = l) \ , \ l = 0, 1$$

such that $x_i \in (0, 1)$ if $z = 1$ and $x_i = 0$ otherwise. We state the problem in these terms and, unlike the previous literature, avoid the so-called counterfactuals in either the modeling or in the subsequent estimation of the model.

To model the joint density $p(y_i, x_i = j|\mathbf{w}_i, z_i = l)$, let $s_i = k$ indicate a discrete confounder random variable that affects both the intake x_i and the outcome y_i. Now imagine that this discrete confounder takes two possible values that represent subject type, namely $\{c, n\}$, for complier and never-taker, respectively. Formally, subject i is of the type complier if $x_{i0} = 0$

and $x_{i1} = 1$, and of the type never-taker if $x_{i0} = 0$ and $x_{i1} = 0$, where x_{il} is the (potential) treatment under the assignment $z_i = l$. It is important to note that we are restricted to these two subject-types because under this design it is not possible to identify always-takers (individuals for whom $x_{i0} = 1$ and $x_{i1} = 1$) from compliers. The assumption that always-takers are absent is non-testable.

Then, conditional on $z_i = l$ and $s_i = k$, the joint density of (y_i, x_i) factors as

$$p(y_i, x_i = j|\mathbf{w}_i, z_i = l, s_i = k) = p_j(y_i|\mathbf{w}_i, s_i = k) \Pr(x_i = j|\mathbf{w}_i, z_i = l, s_i = k) \tag{1}$$

where the first expression is the density of y_i conditional on the latent subject type in the jth treatment state and the second expression is the mass-function of $x_i = j$. The marginal density of the outcome does not involve $z_i = l$ on account of the so-called exclusion restriction. Notice, too, that the second term in this expression is either zero or one. This can be seen from Table 1 which gives the distribution of type by treatment arm and intake. For example, if $z_i = 0$ and

	$x_i = 0$	$x_i = 1$
$z_i = 0$	c, n	—
$z_i = 1$	n	c

Table 1: *Distribution of types by treatment arm and intake*

$s_i = c$, then $x_i = 0$, so that $\Pr(x_i = 0|\mathbf{w}_i, z_i = 0, s_i = c) = 1$ and zero otherwise. In addition, if $z_i = 1$ and $s_i = n$, then $x_i = 0$, implying that $\Pr(x_i = 0|\mathbf{w}_i, z_i = 1, s_i = n) = 1$ and zero otherwise. Thus, given $z_i = l$ and $s_i = k$, the intake is fully determined. This implication of the model is evidently strong. It can be avoided, however, if we model the confounder differently, without reference to s_i, as we do in Section 3.

Let $I_{lj} = \{i : z_i = l \text{ and } x_i = j\}$ denote the sample indices of the subjects in each of the three non-empty cells of Table 1. Also let $\Pr(s_i = c|z_i = l) = q_c$ denote the probability of subject type c, which is independent of z_i because of the random assignment of subjects to the treatment arms. It follows now that the joint density of y_i and $x_i = j$ conditional on $z_i = l$ but

marginalized over s_i is given by

$$p(y_i, x_i = j|\mathbf{w}_i, z_i = l) = \begin{cases} q_c\, p_0(y_i|\mathbf{w}_i, s_i = c) + (1 - q_c)\, p_0(y_i|\mathbf{w}_i, s_i = n) & \text{if } i \in I_{00} \\ (1 - q_c)\, p_0(y_i|\mathbf{w}_i, s_i = n) & \text{if } i \in I_{10} \\ q_c\, p_1(y_i|\mathbf{w}_i, s_i = c) & \text{if } i \in I_{11} \end{cases} \tag{2}$$

which does not involve the mass function of the intake due to the discussion surrounding (1). We see that for $i \in I_{00}$, the density is a mixture of the outcome densities $p_0(y_i|\mathbf{w}_i, s_i = k)$, $k = (c, n)$, because subjects in this cell can be one of these two types. The other two cases are justified in a similar way.

Thus, from (2) we see that the modeling of $(y_i, x_i = j)$ requires three type and treatment state specific distributions, $p_0(y_i|\mathbf{w}_i, s_i = n)$, and $p_j(y_i|\mathbf{w}_i, s_i = c)$, for $j = 0, 1$. For specificity, we assume that these are student-t with known degrees of freedom ν and of the form

$$\begin{aligned} p_0(y_i|\mathbf{w}_i, s_i = n) &= t_\nu(y_i|\mathbf{w}_i'\boldsymbol{\beta}_{0,n}, \eta^2_{0,n}) \\ p_j(y_i|\mathbf{w}_i, s_i = c) &= t_\nu(y_i|\mathbf{w}_i'\boldsymbol{\beta}_{j,c}, \eta^2_{j,c}),\ j = 0, 1 \end{aligned} \tag{3}$$

where $\boldsymbol{\beta}_{0,n}$ and $\boldsymbol{\beta}_{j,c}$ are type and treatment state-specific p-dimensional vectors of regression parameters, $\eta^2_{0,n}$ and $\eta^2_{j,c}$ are the corresponding dispersion parameters and $t_\nu(\cdot|\mu, s^2)$ is the student-t density function with ν degrees of freedom, mean μ and variance $\nu s^2/(\nu - 2)$. The student-t assumption is a generalization of the common Gaussian assumption in the literature.

To complete the model specification we formulate a prior density for the model parameters $\boldsymbol{\psi} = (\boldsymbol{\beta}, \boldsymbol{\eta}^2, q_c)$ with $\boldsymbol{\beta} = (\boldsymbol{\beta}_{0,n}, \boldsymbol{\beta}_{0,c}, \boldsymbol{\beta}_{1,c})$ and $\boldsymbol{\eta}^2 = (\eta^2_{0,n}, \eta^2_{0,c}, \eta^2_{1,c})$. Also let $\boldsymbol{\psi}_{j,k} = (\boldsymbol{\beta}_{j,k}, \eta^2_{j,k})$. We assume that the parameters are mutually independent and that the prior density is of the form

$$\pi(\boldsymbol{\psi}) = \pi(q_c) \prod_{j=0,1} \prod_{k \in \mathcal{K}_j} \pi(\boldsymbol{\psi}_{j,k})$$

where $\pi(q_c) = \mathcal{B}(q_c|a_0, b_0)$, a beta density with hyperparameters a_0 and b_0, $\mathcal{K}_0 = \{c, n\}$, $\mathcal{K}_1 = \{c\}$ and

$$\pi(\boldsymbol{\psi}_{j,k}) = \mathcal{N}_p(\boldsymbol{\beta}_{j,k}|\boldsymbol{\beta}_{jk,0}, \mathbf{B}_{jk,0})\mathcal{IG}\left(\eta^2_{j,k}|\frac{n_{jk,0}}{2}, \frac{d_{jk,0}}{2}\right)$$

where $\mathcal{IG}$ is the inverse-gamma density.

2.1.1 Fitting

In the Bayesian approach we take all information about the model parameters $\boldsymbol{\psi} = (\boldsymbol{\beta}, \boldsymbol{\eta}^2, q_c)$ is summarized in the posterior distribution $\pi(\boldsymbol{\psi}|\mathbf{y}, \mathbf{x}, \mathbf{z}, \mathbf{W}) \propto f(\mathbf{y}, \mathbf{x}|\mathbf{z}, \mathbf{W})\pi(\boldsymbol{\psi})$, which is proportional to the product of the likelihood $f(\mathbf{y}, \mathbf{x}|\mathbf{z}, \mathbf{W})$ of the observed data and the prior distribution $\pi(\boldsymbol{\psi})$. The posterior distribution is not tractable because of the mixture distribution in the control arm. We achieve tractability, however, by treating s_i for these subjects as additional unknown parameters. In addition, we rewrite the student-t density $t_\nu(\cdot|\mu, s^2)$ in familiar form as $\mathcal{N}(\cdot|\mu, \lambda_i^{-1}s^2)$, where $\lambda_i \sim \mathcal{G}(\frac{\nu}{2}, \frac{\nu}{2})$, and then augment the parameter space to include λ_i, $i \leq n$.

Under these assumptions and augmentations, $\pi(\boldsymbol{\psi}, \{\lambda_i\}, \{s_i : i \in I_{00}\}|\mathbf{y}, \mathbf{x}, \mathbf{z}, \mathbf{W})$ is the posterior density of interest, which is proportional to

$$\pi(\boldsymbol{\psi}) \prod_{i=1}^{n} \mathcal{G}\left(\lambda_i|\frac{\upsilon}{2}, \frac{\upsilon}{2}\right) \prod_{i\in I_{00}} \sum_{k\in\{c,n\}} q_k \times \mathcal{N}(y_i|\mathbf{w}_i'\boldsymbol{\beta}_{0,k}, \lambda_i^{-1}\eta_{0,k}^2)$$
$$\times \prod_{i\in I_{10}} (1-q_c) \times \mathcal{N}(y_i|\mathbf{w}_i'\boldsymbol{\beta}_{0,n}, \lambda_i^{-1}\eta_{0,n}^2) \prod_{i\in I_{11}} q_c \times \mathcal{N}(y_i|\mathbf{w}_i'\boldsymbol{\beta}_{1,c}, \lambda_i^{-1}\eta_{1,c}^2) \tag{4}$$

where $q_k = \Pr(s_i = k)$.

In the remainder of this section we briefly describe a 4-block MCMC algorithm we have developed to efficiently sample the posterior distribution. In the first step of the algorithm we jointly update the latent subject types s_i, $i \in I_{00}$, and the scale parameters λ_i, $i \leq n$. This joint sampling strategy helps to enhance the mixing of the chain and is executed by first sampling $s_i = c$ with probability

$$\Pr(s_i = c|y_i, x_i, q_c, \boldsymbol{\beta}_{0,c}, \boldsymbol{\beta}_{0,n}, \eta_{0,c}^2, \eta_{0,n}^2) \propto q_c t_\nu(y_i|\mathbf{w}_i'\boldsymbol{\beta}_{0,c}, \eta_{0,c}^2)$$

where the normalizing constant is $[q_c\ t_\nu(y_i|\mathbf{w}_i'\boldsymbol{\beta}_{0,c}, \eta_{0,c}^2) + (1-q_c)\ t_\nu(y_i|\mathbf{w}_i'\boldsymbol{\beta}_{0,n}, \eta_{0,n}^2)]$. Then, given $s_i = k$ $(i \leq n)$, the first step is completed by sampling the latent scale parameters from the gamma density

$$\pi(\lambda_i|y_i, x_i = j, \mathbf{w}_i, s_i = k, \boldsymbol{\beta}, \boldsymbol{\eta}^2) = \mathcal{G}\left(\lambda_i \middle| \frac{\nu+1}{2}, \frac{\nu + (y_i - \mathbf{w}_i'\boldsymbol{\beta}_{j,k})\eta_{j,k}^{-2}(y_i - \mathbf{w}_i'\boldsymbol{\beta}_{j,k}}{2}\right)$$

In the second and third steps, conditional on $\{s_i\}$ and the sub-samples of individuals $i \in I_{jk}$, we sample the regression parameters $\boldsymbol{\beta}_{j,k}$ and the variance parameters $\eta_{j,k}^2$. Specifically, let

$\mathbf{y}_{jk} = \{y_i : i \in I_{jk}\}$, $\mathbf{x}_{jk} = \{x_i : i \in I_{jk}\}$, $\mathbf{W}_{jk} = \{\mathbf{w}_i : i \in I_{jk}\}$ and $\boldsymbol{\lambda}_{jk} = \{\lambda_i : i \in I_{jk}\}$ denote the sub-sample of observations, by intake state j and type k, on the outcome, intake, covariates and latent scale, respectively. Then, our sampling of $\boldsymbol{\beta}_{jk}$ is from

$$\begin{aligned}&\pi(\boldsymbol{\beta}_{jk}|\mathbf{y}_{jk}, \mathbf{x}_{jk}, \mathbf{W}_{jk}, \eta^2_{j,k}, \boldsymbol{\lambda}_{jk}) \\ &\quad = \mathcal{N}(\boldsymbol{\beta}_{jk}|\mathbf{B}_{jk}\{\mathbf{B}^{-1}_{jk,0}\boldsymbol{\beta}_{jk,0} + \sum_{i\in I_{jk}} \lambda_i \mathbf{w}_i' \eta^{-2}_{j,k} y_i\}, \{\mathbf{B}^{-1}_{jk,0} + \sum_{i\in I_{jk}} \lambda_i \mathbf{w}_i' \eta^{-2}_{j,k} \mathbf{w}_i\}^{-1}) \end{aligned} \tag{5}$$

and $\eta^2_{j,k}$ from

$$\pi(\eta^2_{j,k}|\mathbf{y}_{jk}, \mathbf{x}_{jk}, \mathbf{W}_{jk}, \boldsymbol{\beta}_{jk}, \boldsymbol{\lambda}_{jk}) = \mathcal{IG}\left(\eta^2_{j,k} \middle| \frac{n_{jk,0} + n_{j,k}}{2}, \frac{d_{jk,0} + \sum_{i\in I_{jk}} \lambda_i (y_i - \mathbf{w}_i' \boldsymbol{\beta}_{j,k})^2}{2}\right) \tag{6}$$

where n_{jk} denotes the number of individuals in the set I_{jk}.

Finally, in the fourth step, we sample q_c from the beta density

$$\pi(q_c|\mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\beta}, \boldsymbol{\eta}^2, \{\lambda_i\}, \{s_i\}) = \mathcal{B}(q_c|a_0 + \sum_{i\leq n} I[s_i = c], b_0 + \sum_{i\leq n} I[s_i = n]) \tag{7}$$

where $I[s_i = k]$ is the indicator function that takes the value one if $s_i = k$ and zero otherwise.

2.1.2 Inferring Treatment Effects for Compliers

We now discuss a particular approach for ascertaining the causal effect that is appropriate in this context. The general idea, as discussed by Sommer and Zeger (1991), is to compare the potential outcomes for the sub-sample of compliers. In our Bayesian context, a rather informative solution to this quest can be constructed by focusing on the predictive distribution (for compliers) of each potential outcome, marginalized over the covariates and the parameters. (Note that the potential outcomes are involved in this computation but did not arise in the estimation of the model). A number of interesting treatment effects can then be obtained from these predictive distributions, including something we call the predictive average causal effect (PACE) and predictive quantile causal effects.

Specifically, letting $y_{jc,n+1}$ denote the potential outcome for subject $n+1$ who is a complier, we calculate the marginal predictive densities $p(y_{jc,n+1}|\mathbf{y}, \mathbf{x}, \mathbf{W}, \mathbf{z})$ given by

$$\int p(y_{jc,n+1}|\mathbf{w}_{n+1}, \boldsymbol{\beta}_{jc}, \eta^2_{jc}) \pi(\boldsymbol{\beta}_{jc}, \eta^2_{jc}, \{s_i\}, \mathbf{w}_{n+1}|\mathbf{y}, \mathbf{x}, \mathbf{W}, \mathbf{z}) d\boldsymbol{\beta}_{jc} d\eta^2_{jc} d\{s_i\} d\mathbf{w}_{n+1} \tag{8}$$

where $p(y_{jc,n+1}|\mathbf{w}_{n+1}, \boldsymbol{\beta}_{jc}, \eta^2_{jc})$ is $t_\nu(y_{jc,n+1}|\mathbf{w}'_{n+1}\boldsymbol{\beta}_{jc}, \eta^2_{jc})$ and the unknowns are marginalized with respect to the posterior distribution.

We use the method of composition to generate draws from $p(y_{jc,n+1}|\mathbf{y}, \mathbf{x}, \mathbf{W}, \mathbf{z})$. Specifically, we append the following steps at the end of each MCMC iteration. First, we form the $n_c \times p$-dimensional covariate matrix $\mathbf{W}_c^{(g)} = \{\mathbf{w}_i : s_i^{(g)} = c\}$ for the set of compliers in the gth iteration, where $n_c^{(g)}$ is the number of subjects with $s_i^{(g)} = c$, and then sample the covariate vector $\mathbf{w}_{n+1}^{(g)}$ by assigning probability $1/n_c^{(g)}$ to each row of $\mathbf{W}_c^{(g)}$. We then draw $y_{jc,n+1}^{(g)}$ from $t_\nu(y_{jc,n+1}|\mathbf{w}'_{n+1}{}^{(g)}\boldsymbol{\beta}_{jc}^{(g)}, \eta^{2(g)}_{jc})$. The resulting draws $(y_{0c,n+1}^{(1)}, ..., y_{0c,n+1}^{(G)})$ and $(y_{1c,n+1}^{(1)}, ..., y_{1c,n+1}^{(G)})$ are stored and used to compute several treatment effects of interest. For example, the PACE average causal effect

$$\text{PACE} = E(y_{1c,n+1}|\mathbf{y}, \mathbf{x}, \mathbf{W}, \mathbf{z}) - E(y_{0c,n+1}|\mathbf{y}, \mathbf{x}, \mathbf{W}, \mathbf{z}) \tag{9}$$

can be computed from the simulated draws in an obvious manner.

2.1.3 Simulation Study

Before proceeding to our further methodological developments we examine the performance of our fitting approach in a simulation study. We consider a general design with one continuous covarariate w_i that we generate from a $\mathcal{N}(2, 4)$ distribution. The coefficient vectors are taken to be $\boldsymbol{\beta}_{0c} = (1, 2)$, $\boldsymbol{\beta}_{0n} = (-0.5, 1)$ and $\boldsymbol{\beta}_{1c} = (2, 3)$, and the variances are fixed at $\boldsymbol{\eta}^2 = (4.00, 4.00, 4.00)$. Finally, q_c is set in turn to 0.4, 0.6 and 0.8. For each design, we consider sample sizes of 250, 500 and 1,000. In all 20 replications of each design are considered. To get a feel for the data, we report in Table 2 the distribution of compliers and never takers under the different designs. One obvious consequence of the eligibility design, noticeable from Table 2, is

	Sample Distribution of Types								
	$n = 250$			$n = 500$			$n = 1000$		
q_c:	.4	.6	.8	.4	.6	.8	.4	.6	.8
n_{0c}	30	47	60	61	92	123	124	190	251
n_{0n}	150	99	50	202	200	100	606	398	199
n_{1c}	70	104	140	137	208	277	270	412	550

Table 2: *Discrete confounder model: Number of compliers and never takers under different values of q_c and different sample sizes, averaged over 20 replications.*

the smaller number of compliers in the control arm relative to the treatment arm. For example,

when $q_c = 0.4$ and the sample size is 250 we have that $n_{0c} = 30$ and $n_{1c} = 70$.

In each experiment, our prior-posterior analysis is conducted under the prior assumption that $\boldsymbol{\beta}_{jk} \sim \mathcal{N}_p(\mathbf{0}, 25 \times \mathbf{I}_p)$ and $q_c \sim \mathcal{B}(3, 5)$ while the hyper-parameters of the inverse gamma distribution for $\eta^2_{j,k}$ are set to have a prior mean of 2 and standard deviation of 6.

In Table 3 we report the posterior means and standard deviations of $\boldsymbol{\beta}_{0c}$, $\boldsymbol{\beta}_{0n}$, $\boldsymbol{\beta}_{1c}$ and q_c, averaged over the 20 replications. The average inefficiency factors (a measure of the mixing of the MCMC chain) when $q_c = 0.6$ are given in Table 4. In general the results in Tables 3 and 4

q_c	Sample Size	$\hat{\boldsymbol{\beta}}_{0c}$		$\hat{\boldsymbol{\beta}}_{0n}$		$\hat{\boldsymbol{\beta}}_{1c}$		$\hat{q}_c$
.4	$n = 250$	1.28(0.82)	1.96(0.27)	-0.67(0.27)	1.04(0.09)	2.06(0.35)	2.95(0.12)	0.40(0.03)
	$n = 500$	1.10(0.60)	1.92(0.19)	-0.62(0.19)	1.02(0.07)	1.94(0.26)	3.01(0.09)	0.40(0.03)
	$n = 1000$	1.23(0.43)	1.93(0.13)	-0.48(0.14)	0.99(0.05)	1.93(0.18)	3.01(0.07)	0.39(0.02)
.6	$n = 250$	1.22(0.61)	1.97(0.20)	-0.57(0.35)	1.02(0.12)	1.945(0.29)	3.01(0.10)	0.59(0.03)
	$n = 500$	0.88(0.43)	2.06(0.14)	-0.38(0.24)	0.95(0.08)	2.04(0.21)	2.98(0.07)	0.60(0.03)
	$n = 1000$	1.00(0.30)	2.00(0.10)	-0.43(0.17)	1.00(0.06)	2.03(0.15)	2.97(0.05)	0.60(0.02)
.8	$n = 250$	1.04(0.48)	1.94(0.16)	-0.54(0.50)	1.02(0.17)	1.96(0.25)	2.99(0.09)	0.78(0.03)
	$n = 500$	1.00(0.31)	2.04(0.11)	-0.44(0.35)	1.00(0.12)	1.98(0.18)	2.99(0.06)	0.79(0.02)
	$n = 1000$	0.98(0.23)	2.00(0.08)	-0.51(0.24)	1.03(0.08)	2.02(0.13)	3.01(0.05)	0.80(0.02)

Table 3: *Discrete confounder model: Simulation study with $\boldsymbol{\beta}_{0c} = (1, 2)$, $\boldsymbol{\beta}_{0n} = (-0.5, 1)$, $\boldsymbol{\beta}_{1c} = (2, 3)$. Reported are the average of the posterior means and standard deviations from the 20 replications of each design.*

show that the MCMC algorithm performs well. We should note the higher posterior standard errors of $\boldsymbol{\beta}_{0c}$, in particular the intercept. It appears that $\boldsymbol{\beta}_{0c}$ is the more difficult parameter to estimate, especially when n_{0c} is relatively small. On the other hand, $\boldsymbol{\beta}_{0n}$ is more precisely estimated even when n_{0n} is small because we effectively obtain a clean estimate of $\boldsymbol{\beta}_{0n}$ from the subjects in the set I_{10}.

	Inefficiency Factors									
Sample Size	$\boldsymbol{\beta}_{0c}$		$\boldsymbol{\beta}_{0n}$		$\boldsymbol{\beta}_{1c}$		η^2_{0c}	η^2_{0n}	η^2_{1c}	q_c
250	3.54	2.87	2.03	1.76	1.20	1.23	3.92	1.992	1.39	1.81
500	3.23	2.41	2.09	1.77	1.20	1.18	3.60	1.88	1.36	1.83
1000	3.42	2.44	2.01	1.85	1.20	1.16	3.73	1.82	1.35	1.94

Table 4: *Discrete confounder model: Inefficiency factors from the sampled MCMC output for the case $q_c = 0.6$. The results are averaged over 20 replications.*

For a more detailed analysis of the performance of the algorithm we graph in Figure 1 the posterior means of the model parameters across the 20 replications when the data is generated with $q_c = 0.6$. In line with the results above, we observe higher variation in the posterior mean

of $\beta_{0c,0}$ and larger variation in the estimates of the variance parameter η^2_{0c} when n_{0c} is smaller.

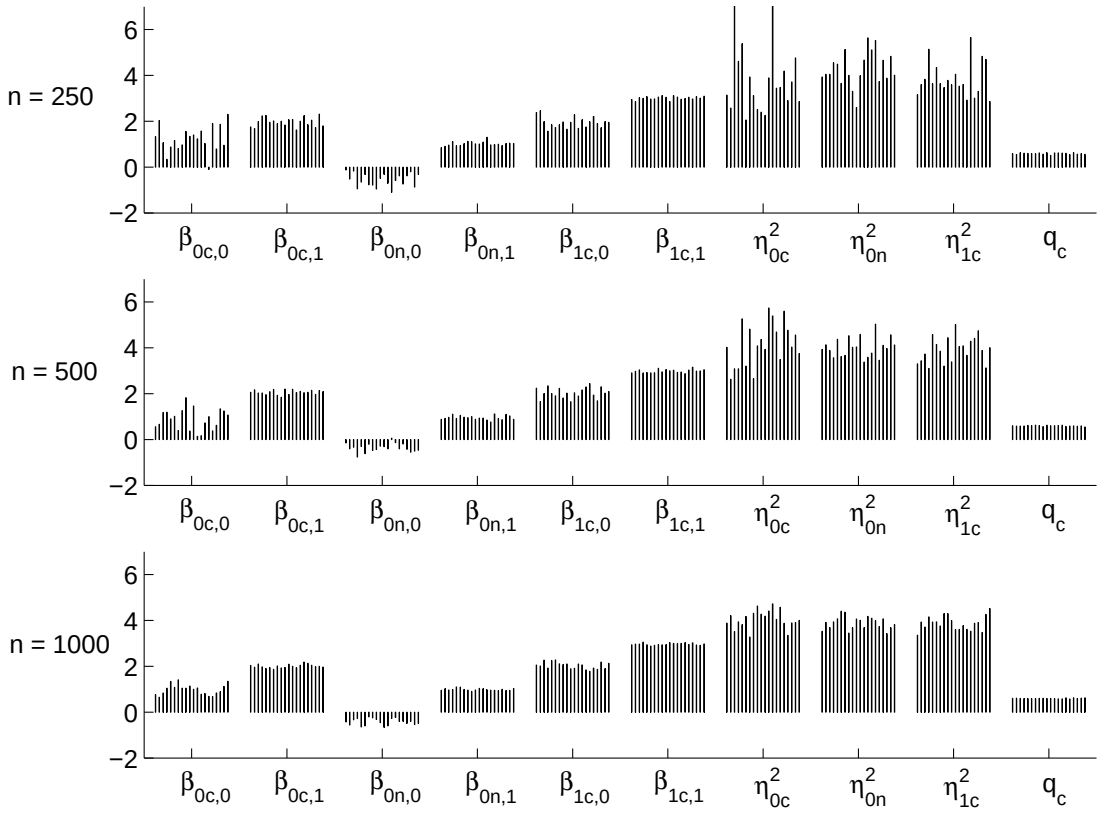
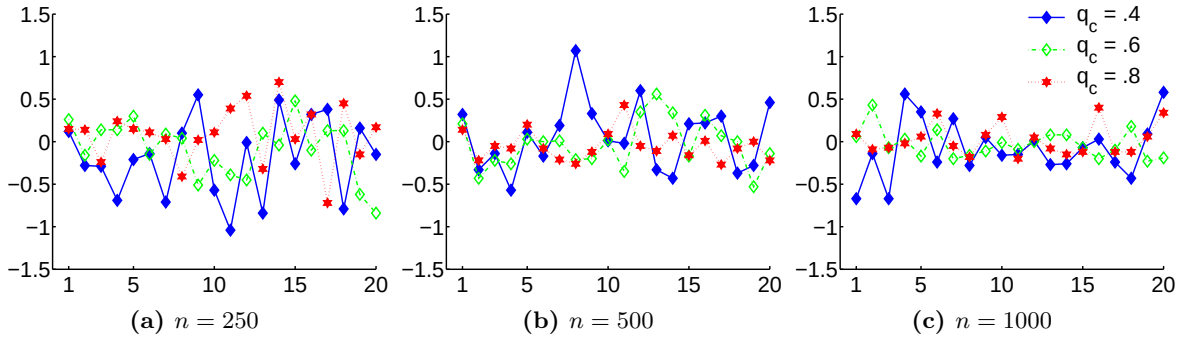


Figure 1: *Discrete confounder model: Simulation study with* $\boldsymbol{\beta}_{0c} = (1, 2)$, $\boldsymbol{\beta}_{0n} = (-0.5, 1)$, $\boldsymbol{\beta}_{1c} = (2, 3)$ *and* $q_c = 0.6$. *Posterior means of the model parameters across the 20 replications.*

Given the above results we would expect that our inference of the average treatment effect for compliers would improve when the sample contains a larger number of compliers. This is



(a) $n = 250$ **(b)** $n = 500$ **(c)** $n = 1000$

Figure 2: *Discrete confounder model: Difference between the estimated PACE and true average treatment effects for compliers across different designs and across the 20 replications.*

borne out in Figure 2 where we plot the difference between the estimated PACE and the true treatment effect for compliers by sample size and proportion of compliers. The figure reveals

that an increase in the sample size improves inferences about the treatment effect but that the extent of improvement depends on the proportion of compliers in the sample. Specifically, an increase in the sample size leads to a smaller improvement in inferences about the treatment effect when the sample contains a smaller proportion of compliers. Conversely, the improvement is more pronounced when the sample has a larger proportion of compliers.

3 Continuous Confounder Approach

3.1 Modeling

We now consider an alternative way of modeling confounding that does not involve the latent subject type s_i. We provide the needed modeling and inferential techniques and discuss the effectiveness of our methods. Then in Section 4 we consider how the discrete and continuous confounder approaches can be compared in terms of marginal likelihoods and Bayes factors.

The modeling we describe is an outgrowth of our previous work on non-eligibility designs (Chib (2004) and Chib and Jacobi (2005)) but unlike our previous work, the framework here takes account of the fact that intake in the control arm is non-stochastic. This change has a substantial effect on the subsequent analysis and on our Markov chain Monte Carlo (MCMC) approach for fitting the model.

As before, the objective is to model for each individual the joint density of the observed data

$$p(y_i, x_i = j|\mathbf{w}_i, z_i = l)$$

for $l = 0, 1$. The first requirement we impose on this joint density is the exclusion restriction that z_i, the instrumental variable, does not affect the implied marginal distribution of y_i. The second requirement we impose is that the treatment intake in the control arm is non-stochastic. Finally, we require that the form of this joint distribution varies by intake. These considerations lead us to the model

$$p(y_i, x_i = j|\mathbf{w}_i, z_i = l) = \begin{cases} p_0(y_i|\mathbf{w}_i) & \text{if } i \in I_0 \\ p_j(y_i, x_i = j|\mathbf{w}_i, z_i = 1),\ j = 0, 1 & \text{if } i \in I_1 \end{cases} \tag{10}$$

where $I_l = \{i : z_i = l\}$ denotes the sample indices of a subject with $z_i = l$. Thus, for subjects in the control arm the model is specified in terms of the marginal density of the outcome $p_0(y_i|\mathbf{w}_i)$.

Naturally, this marginal density is deduced from the joint density $p_0(y_i, x_i = 0|\mathbf{w}_i, z_i = 1)$ of subjects in the treatment arm who forgo the treatment.

The task at hand is to specify the form of $p_j(y_i, x_i = j|\mathbf{w}_i, z_i = 1)$, $j = 0, 1$. Letting y_{ji} indicate the response in treatment state j, we specify these joint densities in terms of the generating process

$$\begin{aligned} y_{ji} &= \mathbf{w}_i\boldsymbol{\beta}_j + \varepsilon_{ji}\ ,\ j = 0, 1 \\ x_i^* &= \mathbf{w}_i'\boldsymbol{\gamma} + u_i \\ x_i &= I\left\{x_i^* > 0\right\} \end{aligned}$$

where x_i^* is a latent variable and $I(\cdot)$ is the indicator function. To model continuous unobserved confounders we assume that (ε_{ji}, u_i) given λ_i have the distribution

$$(\varepsilon_{ji}, u_i)|\lambda_i \sim \mathrm{N}_2\left(\mathbf{0}, \lambda_i^{-1}\boldsymbol{\Omega}_j\right)$$

where

$$\boldsymbol{\Omega}_j = \begin{pmatrix} \eta_j^2 & \omega_j \\ \omega_j & 1 \end{pmatrix},$$

is upto scale the conditional covariance matrix between the y_{ji} and x_i^*. Here λ_i is a positive random-variable that is assumed to be iid gamma $\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ for some known value $\nu > 0$. We could entertain other possible mixing distributions but our choice of the gamma distribution leads to the bivariate student-t distribution which is simple to work with and reasonably flexible.

If we let $\boldsymbol{\beta} = (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\gamma})$, $\sigma_j^2 = \eta_j^2 - \omega_j^2$ (for reasons presented below) and $\boldsymbol{\psi}_j = (\boldsymbol{\beta}_j, \omega_j, \sigma_j^2)$, it follows that the joint density of y_i and x_i^* for subjects in I_1 is

$$p_j^*(y_i, x_i^*|\mathbf{w}_i, z_i = 1, \boldsymbol{\psi}_j, \boldsymbol{\gamma}, \lambda_i) = \mathcal{N}_2(y_i, x_i^*|\mathbf{X}_{j,i}\boldsymbol{\beta}, \lambda_i^{-1}\boldsymbol{\Omega}_j) \tag{11}$$

where

$$\mathbf{X}_{j,i} = \begin{pmatrix} \mathbf{w}_i' \times (1-j) & \mathbf{w}_i' \times j & 0 \\ 0 & 0 & \mathbf{w}_i' \end{pmatrix}$$

Therefore, marginal of λ_i, the joint densities $p_0^*(y_i, x_i^*|\mathbf{w}_i, z_i = 1, \boldsymbol{\psi}_j, \boldsymbol{\gamma})$ and $p_1^*(y_i, x_i^*|\mathbf{w}_i, z_i = 1, \boldsymbol{\psi}_j, \boldsymbol{\gamma})$ are bivariate student-t. From here, on integrating out the latent scale, paying particular

attention to the interval of integration, we get that for $i \in I_1$,

$$p_j(y_i, x_i|\mathbf{w}_i, z_i = 1, \boldsymbol{\psi}_j, \boldsymbol{\gamma}) = p_j(y_i|\mathbf{w}_i, \boldsymbol{\beta}_j, \eta_j^2) \int_{A_j} p_j^*(x_i^*|\mathbf{w}_i, z_i = l, y_i, \boldsymbol{\beta}_j, \eta_j^2, \omega_j) dx_i^*$$
$$= t_\nu(y_i|\mathbf{w}_i'\boldsymbol{\beta}_j, \eta_j^2) T_\nu \left((2j-1) \frac{\mu_{ji}}{h_{ji}\psi_j} \right) \tag{12}$$

where A_j is the set $(-\infty, 0)$ if $j = 0$ and $(0, \infty)$ if $j = 1$, t_ν is the student-t density with ν degrees of freedom, $T_{\nu+1}$ is the cdf of the standard t-density and

$$\mu_{ji} = \mathbf{w}_i'\boldsymbol{\gamma} + \omega_j \eta_j^{-2}(y_i - \mathbf{w}_i'\boldsymbol{\beta}_j)$$
$$h_{ji}^2 = [\nu(\nu+1)][1 - \omega_j^2/\eta_j^2]$$
$$\psi_j^2 = 1 - \omega_j^2/\eta_j^2$$

Accordingly, for subjects in $i \in I_0$ this implies that

$$p_j(y_i, x_i|\mathbf{w}_i, z_i = 0, \boldsymbol{\beta}_0, \eta_0^2) = t_\nu(y_i|\mathbf{w}_i'\boldsymbol{\beta}_0, \eta_0^2) \tag{13}$$

To complete the model specification we now supply a prior density for the model parameters $\boldsymbol{\psi} = (\boldsymbol{\gamma}, \boldsymbol{\psi}_0, \boldsymbol{\psi}_1)$. Specifically, we assume that σ_j^2 is distributed inverse-gamma and $\boldsymbol{\beta}_j$ and ω_j are Gaussian:

$$\pi(\boldsymbol{\psi}_j) = \mathcal{IG}\left(\sigma_j^2 | \frac{n_{j,0}}{2}, \frac{d_{j,0}}{2}\right) \mathcal{N}_p(\boldsymbol{\beta}_j|\boldsymbol{\beta}_{j,0}, \mathbf{B}_{j,0}) \mathcal{N}(\omega_j|m_{j,0}, M_{j,0})$$

where the quantities indexed by zero are the prior hyperparameters, and p is the dimension of $\boldsymbol{\beta}_j$. We assume that $\boldsymbol{\psi}_0$ and $\boldsymbol{\psi}_1$ are apriori independent. Notice that σ_j^2 is the determinant of $\boldsymbol{\Omega}_j$ and our choice of an inverse-gamma distribution for it ensures that $\boldsymbol{\Omega}_j$ is positive definite. We note for future use that these assumptions imply that the prior density of the $p+1$-dimensional vector $\tilde{\boldsymbol{\beta}}_1 = (\boldsymbol{\beta}_1, \omega_1)$ is $\mathcal{N}_{p+1}(\tilde{\boldsymbol{\beta}}_1|\tilde{\boldsymbol{\beta}}_{1,0}, \tilde{\mathbf{B}}_{1,0})$ with mean $\tilde{\boldsymbol{\beta}}_{1,0} = (\boldsymbol{\beta}_{1,0}|m_{1,0})$ and covariance matrix

$$\tilde{\mathbf{B}}_{1,0} = \begin{pmatrix} \mathbf{B}_{1,0} & \mathbf{0} \\ \mathbf{0} & M_{1,0} \end{pmatrix}.$$

We further assume that $\boldsymbol{\gamma}$ is distributed as $\mathcal{N}_p(\boldsymbol{\gamma}|\boldsymbol{\gamma}_0, \mathbf{G}_0)$, independent of $\boldsymbol{\psi}_j$. Putting these

assumptions together, the prior on $\boldsymbol{\psi} = (\boldsymbol{\beta}, \boldsymbol{\psi}_0, \boldsymbol{\psi}_1)$ is of the form

$$\pi(\boldsymbol{\psi}) = \mathcal{N}_p(\boldsymbol{\gamma}|\boldsymbol{\gamma}_0, \mathbf{G}_0) \prod_{j=0}^{1} \mathcal{IG}\left(\sigma_j^2|\frac{n_{j,0}}{2}, \frac{d_{j,0}}{2}\right) \mathcal{N}_p(\boldsymbol{\beta}_j|\boldsymbol{\beta}_{j,o}, \mathbf{B}_{j,0}) \mathcal{N}(\omega_j|m_{j,0}, M_{j,0}) \tag{14}$$

3.1.1 Fitting

We now formulate an efficient approach for summarizing the posterior density of $\boldsymbol{\psi}$. From Bayes theorem this density is given by $\pi(\boldsymbol{\psi}|\mathbf{y}, \mathbf{x}) \propto \pi(\boldsymbol{\psi}) f(\mathbf{y}, \mathbf{x}|\boldsymbol{\psi})$. Because this posterior density is analytically intractable, we utilize MCMC methods along with ideas from Albert and Chib (1993) to summarize it numerically. In particular, we augment the parameter space with the latent treatment variables $\{x_i^*\}$, for $i \in I_1$, and the scale parameters $\{\lambda_i\}$, for all i, and work with the joint density

$$\begin{aligned} &p(y_i, x_i^*, x_i = j|\mathbf{w}_i, z_i = l, \boldsymbol{\beta}, \boldsymbol{\psi}_j, \lambda_i) \\ &\qquad = p_j^*(y_i, x_i^*|\mathbf{w}_i, z_i = l, \boldsymbol{\beta}, \boldsymbol{\psi}_j, \lambda_i) \left\{I(x_i^* < 0)^{1-j} + I(x_i^* > 0)^j\right\} \end{aligned}$$

where $p_j^*(y_i, x_i^*|\mathbf{w}_i, z_i = l, \boldsymbol{\beta}, \boldsymbol{\psi}_j, \lambda_i)$ is the bivariate normal density given in expression (11). Therefore, the posterior distribution of interest $\pi(\boldsymbol{\psi}, \{x_i^*\}, \{\lambda_i\}|\mathbf{y}, \mathbf{x})$ is proportional to:

$$\begin{aligned} &\pi(\boldsymbol{\psi}) \prod_{i=1}^{n} \mathcal{G}\left(\lambda_i|\frac{\upsilon}{2}, \frac{\upsilon}{2}\right) \prod_{i \in I_0} \mathcal{N}(y_i|\mathbf{w}_i'\boldsymbol{\beta}_0, \lambda_i^{-1}, \boldsymbol{\eta}_0^2) \\ &\times \prod_{i \in I_{10}} \mathcal{N}_2(y_i, x_i^*|\mathbf{X}_{0,i}\boldsymbol{\beta}, \lambda_i^{-1}\boldsymbol{\Omega}_0) I(x_i^* < 0) \prod_{i \in I_{11}} \mathcal{N}_2(y_i, x_i^*|\mathbf{X}_{1,i}\boldsymbol{\beta}, \lambda_i^{-1}\boldsymbol{\Omega}_1) I(x_i^* > 0) \end{aligned} \tag{15}$$

Thus, the contribution of the subjects in the control arm ($i \in I_0$) is from a product of univariate normal densities, whereas the contribution of the subjects in the treatment arm ($i \in I_1$) is from a product of bivariate normals, according to I_{10} and I_{11}. The fact that some subjects with intake $x_i = 0$ contribute through the marginal density of the outcome and others contribute through the joint density of the outcome and the intake is a consequence, of course, of the eligibility design.

There are two key points to consider when designing an MCMC algorithm to sample from the posterior density in (15). First, the update of σ_0^2 and ω_0 requires an M-H step since the posterior density of these parameters conditional on the observed data and the remaining model parameters is not tractable due to form of the likelihood function for the subset of subjects

with $x_i = 0$. Second, we must also recognize that the degree of confounding when $x = 1$ is more difficult to estimate because subjects in this case are only observed under one level of the assignment. We address this issue by updating ω_1 and $\boldsymbol{\beta}_1$ together. Note that this problem does not occur for ω_0 and $\boldsymbol{\beta}_0$ since the data in the set I_0 supplies information about $\boldsymbol{\beta}_0$, independent of ω_0.

Following these considerations we propose a five-block algorithm to sample $\boldsymbol{\psi}$, $\mathbf{x}^* = \{x_i^*\}$ and $\boldsymbol{\lambda} = \{\lambda_i\}$. This algorithm bears only distant resemblance to the algorithm described in Chib (2005) for non-eligibility models.

In the first block of the algorithm we jointly sample the latent treatment variables x_i^*, for $i \in I_0$ and the scale parameters λ_i, for $i \leq n$. We begin by sampling x_i marginalized over λ_i from $t_{v+1}(x_i^*|m_{ji}, \psi_{ji}^2)\left\{I(x_i^* < 0)^{1-j} + I(x_i^* > 0)^j\right\}$, where $m_{ji} = \mathbf{w}_i'\boldsymbol{\gamma} + \omega_j\eta_j^{-2}(y_{ji} - \mathbf{w}_i'\boldsymbol{\beta}_j)$ and $\psi_{ji}^2 = 1 - \omega_j/\eta_j^2$. This is followed by the update of λ_i for these subjects ($i \in I_1$) from a gamma density $\mathcal{G}(\lambda_i|\frac{\nu_1}{2}, \frac{d_1}{2})$, where $\nu_1 = \nu_0 + 2$ and $d_1 = \nu_0 + (\boldsymbol{y}_i^* - \mathbf{X}_{1i}\boldsymbol{\beta})'\boldsymbol{\Omega}_j^{-1}(\boldsymbol{y}_i^* - \mathbf{X}_{1i}\boldsymbol{\beta})$. In the final step of the first block we update λ_i for subjects in the control group ($i \in I_0$) from a gamma density, where the parameters ν_1 and d_1 are now given by $\nu_1 = \nu_0 + 1$ and $d_1 = \nu_0 + (y_i - \mathbf{w}_i'\boldsymbol{\beta}_0)\eta_0^{-2}(y_i - \mathbf{w}_i'\boldsymbol{\beta}_0)$.

Next we sample $\boldsymbol{\gamma}$ based on the observations in the sets I_{10} and I_{11} from

$$\pi(\boldsymbol{\gamma}|\mathbf{y}, \mathbf{x}, \mathbf{W}, \{x_i^*\}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \sigma_0^2, \omega_0, \sigma_1^2, \omega_1, \{\lambda_i\}) = \mathcal{N}_p(\boldsymbol{\gamma}|\hat{\boldsymbol{\gamma}}, \mathbf{G}) \tag{16}$$

with mean $\hat{\boldsymbol{\gamma}} = \mathbf{G}\{\mathbf{G}_0^{-1}\boldsymbol{\gamma}_0 + \sum_{i\in I_{10}}\lambda_i\mathbf{w}_i(1 - \omega_0^2\eta_0^{-2})\bar{x}_{i0}^* + \sum_{i\in I_{11}}\lambda_i\mathbf{w}_i(1 - \omega_1^2\eta_1^{-2})\bar{x}_{i1}^*\}$, where $\bar{x}_{ij}^* = (x_i^* - \omega_j\eta_j^{-2}(y_i - \mathbf{w}_i'\boldsymbol{\beta}_j))$, and covariance $\mathbf{G} = \{\mathbf{G}_0^{-1} + \sum_{i\in I_{10}}\lambda_i\mathbf{w}_i(1 - \omega_0^2\eta_0^{-2})\mathbf{w}_i + \sum_{i\in I_{11}}\lambda_i\mathbf{w}_i(1 - \omega_1^2\eta_1^{-2})\mathbf{w}_i\}^{-1}$.

We continue by updating $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$ and ω_1 in one block, sampling $\boldsymbol{\beta}_0$ from

$$\pi(\boldsymbol{\beta}_0|\mathbf{y}, \mathbf{x}, \mathbf{W}, \{x_i^*\}, \boldsymbol{\gamma}, \sigma_0^2, \omega_0, \{\lambda_i\}) = \mathcal{N}_p(\boldsymbol{\beta}_0|\hat{\boldsymbol{\beta}}_0, \mathbf{B}_0) \tag{17}$$

with mean $\hat{\boldsymbol{\beta}}_0 = \mathbf{B}_0\{\mathbf{B}_{0,0}^{-1}\boldsymbol{\beta}_{0,0} + \sum_{i\in I_0}\lambda_i\mathbf{w}_i\eta_0^{-2}y_i + \sum_{i\in I_{10}}\lambda_i\mathbf{w}_i\sigma_0^{-2}(y_i - \omega_0\hat{x}_i^*)\}$ and covariance $\mathbf{B}_0 = \{\mathbf{B}_{1,0}^{-1} + \sum_{i\in I_0}\lambda_i\mathbf{w}_i\eta_0^{-2}\mathbf{w}_i' + \sum_{i\in I_{10}}\lambda_i\mathbf{w}_i\sigma_0^{-2}\mathbf{w}_i'\}^{-1}$, and $\tilde{\boldsymbol{\beta}}_1 = (\boldsymbol{\beta}_1, \omega_1)$ from

$$\begin{aligned}\pi(\tilde{\boldsymbol{\beta}}_1|\mathbf{y}, \mathbf{x}, \mathbf{W}, \{x_i^*\}, \boldsymbol{\gamma}, \sigma_1^2, \{\lambda_i\}) = \\ \mathcal{N}_{p+1}(\tilde{\boldsymbol{\beta}}_1|\tilde{\mathbf{B}}_1\{\tilde{\mathbf{B}}_{1,0}^{-1}\tilde{\boldsymbol{\beta}}_{1,0} + \sum_{i\in I_{11}}\lambda_i\tilde{\mathbf{w}}_i\sigma_1^{-2}y_i\}, \tilde{\mathbf{B}}_1 = \{\tilde{\mathbf{B}}_{1,0}^{-1} + \sum_{i\in I_{11}}\lambda_i\tilde{\mathbf{w}}_i\sigma_1^{-2}\tilde{\mathbf{w}}_i'\}^{-1})\end{aligned} \tag{18}$$

where $\tilde{\mathbf{w}}_i = \mathbf{w}_i \sim u_i$.

In the next block we jointly update $\boldsymbol{\zeta}_0 = (\sigma_0^2, \omega_0)$ in an MH-step. Using the approach in Chib and Greenberg (1995), we generate the proposal values for $\sigma_0^{2'}$ and ω_0' from a tailored student-t density $t_\nu(\mu, V)$, where μ is the (approximate) mode of the likelihood function for the subjects in the sets I_0 and I_{10} and V is the inverse Hessian of this likelihood evaluated at the mode. We accept the proposal values with probability of move $\alpha = \alpha(\boldsymbol{\zeta}_0, \boldsymbol{\zeta}_0^* | \mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\beta}, \sigma_1^2, \omega_1, \mathbf{x}^*, \boldsymbol{\lambda})$ where

$$\alpha = \min \left\{ 1, \frac{\pi(\boldsymbol{\zeta}_0') \prod_{i \in I_0} \mathcal{N}(y_i | \mathbf{w}_i' \boldsymbol{\beta}_0, \lambda_i^{-1} \eta_0^{2'}) \prod_{i \in I_{10}} \mathcal{N}_2(y_i, x_i^* | \mathbf{X}_{0,i} \boldsymbol{\beta}, \lambda_i^{-1} \boldsymbol{\Omega}_0') \times t_\nu(\boldsymbol{\zeta}_0 | \mu, V)}{\pi(\boldsymbol{\zeta}_0) \prod_{i \in I_0} \mathcal{N}(y_i | \mathbf{w}_i' \boldsymbol{\beta}_0, \lambda_i^{-1} \eta_0^2) \prod_{i \in I_{10}} \mathcal{N}_2(y_i, x_i^* | \mathbf{X}_{0,i} \boldsymbol{\beta}, \lambda_i^{-1} \boldsymbol{\Omega}_0) \times t_\nu(\boldsymbol{\zeta}_0' | \mu, V)} \right\}, \tag{19}$$

$\eta_0^{2'} = \sigma_0^{2'} + \omega_0^{2'}$ and $\eta_0^2 = \sigma_0^2 + \omega_0^2$.

Finally, we sample σ_1^2 conditional on ω_1 and the remaining parameters from

$$\pi(\sigma_1^2 | \mathbf{y}, \mathbf{x}, \mathbf{W}, \{x_i^*\}, \boldsymbol{\beta}, \{\lambda_i\}) = \mathcal{IG}\left(\sigma_1^2 \middle| \frac{\nu_{1,0} + n_{11}}{2}, \frac{d_{1,0} + \sum_{i \in I_{11}} \lambda_i (y_i - \mathbf{w}_i' \boldsymbol{\beta}_1 - \omega_1 u_i)^2}{2} \right) \tag{20}$$

where n_{11} is the number of subjects in the set I_{11}.

We run our MCMC sampler for M iterations after an initial burn-in period of m_0 iterations.

3.1.2 Inferring Treatment Effects

For the estimation of treatment effects we follow the predictive approach outlined in Section 2.1.3, but focus now on the marginal densities of the potential outcomes $p(y_{j,n+1} | \mathbf{y}, \mathbf{x}, \mathbf{W}, \mathbf{z})$, $j = 0, 1$, where the subscript $n+1$ refers to a subject randomly drawn from the entire population. We generate draws from these densities by performing the following steps at each iteration $g \leq M$ of the algorithm. First, we randomly sample $\mathbf{w}_{n+1}^{(g)}$ from the full set of covariates. We then draw $y_{j,n+1}^{(g)}$ from $t_\nu(y_{j,n+1}^{(g)} | \mathbf{w}_{n+1}^{(g)'} \boldsymbol{\beta}_j^{(g)}, \eta_j^{2(g)})$, where $(\boldsymbol{\beta}_j^{(g)}, \eta_j^{2(g)})$ are the values of the parameters at the gth iteration of the chain. The simulated samples $(y_{1,n+1}^{(1)}, ..., y_{1,n+1}^{(M)})$ and $(y_{0,n+1}^{(1)}, ..., y_{0,n+1}^{(M)})$ can be used to compute various effects of interest, including the predictive average causal effect.

3.1.3 Simulation Study

To examine the quality of our inferences about the model parameters and the treatment effects we consider a number of simulation designs, with and without covariates, with varying degrees of confounding, and different sample sizes. We first consider a design with one continuous covarariate w_i. In particular, we assume that $w_i \sim \mathcal{N}(2, 4)$ and let $\boldsymbol{\beta}_0 = (1.00, 2.00)$, $\boldsymbol{\beta}_1 =$

$(2.00, 3.00)$ and $\boldsymbol{\gamma} = (-1.00, 1.00)$. In addition, we let $\boldsymbol{\eta}^2 = (4.00, 4.00)$. To investigate the inference under different degrees of confounding we set the covariance parameters ω_j at values between -1.6 and 1.6. Given the values of η_j^2 this implies correlation coefficients ρ_j between -0.8 and 0.8. For each setting of these parameters, we generate 20 replications of sample size 250, 500 and 1,000.

In our fitting we assume that $\boldsymbol{\gamma} \sim \mathcal{N}_p(\mathbf{0}, 25 \times \mathbf{I}_p)$, $\boldsymbol{\beta}_j \sim \mathcal{N}_p(\mathbf{0}, 25 \times \mathbf{I}_p)$ and $\boldsymbol{\omega}_j \sim \mathcal{N}_p(0, 16)$, for $j = 0, 1$. In addition, we specify the hyper-parameters $\nu_{0,j}$ and $d_{0,j}$ of the inverse gamma prior on η_j^2 in such a way that the prior mean and standard deviation are 2 and $\sqrt{20}$ respectively.

Table 5 contains results for $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$, ρ_0 and ρ_1 from the fitting to the different simulated data sets. The first column of the table refers to the specific combination of ρ_0 and ρ_1 and the second

	Sample Size	$\hat{\boldsymbol{\beta}}_0$		$\hat{\boldsymbol{\beta}}_1$		$\hat{\rho}_0$	$\hat{\rho}_1$
$\boldsymbol{\rho} = (-.5, .5)$	$n = 250$	0.95(0.22)	2.00(0.08)	1.96(0.87)	3.00(0.23)	-0.60(0.16)	0.42(0.35)
	$n = 500$	1.00(0.15)	1.97(0.06)	2.17(0.61)	2.97(0.16)	-0.48(0.14)	0.38(0.27)
	$n = 1000$	0.99(0.11)	2.01(0.04)	1.93(0.37)	3.01(0.10)	-0.51(0.10)	0.55(0.15)
$\boldsymbol{\rho} = (.5, -.5)$	$n = 250$	0.93(0.22)	2.00(0.08)	2.28(0.77)	2.95(0.21)	0.47(0.19)	-0.58(0.26)
	$n = 500$	0.97(0.15)	1.98(0.06)	1.85(0.63)	3.04(0.17)	0.45(0.14)	-0.43(0.27)
	$n = 1000$	0.10(.11)	2.01(0.04)	2.00(0.43)	2.99(0.11)	0.49(0.10)	-0.46(0.19)
$\boldsymbol{\rho} = (.5, .5)$	$n = 250$	1.00(0.23)	2.00(0.08)	2.00(0.87)	2.99(0.23)	0.47(0.20)	0.40(0.35)
	$n = 500$	0.97(0.15)	1.98(0.06)	2.16(0.61)	2.98(0.16)	0.45(0.14)	0.38(0.27)
	$n = 1000$	1.00(0.11)	2.01(0.04)	1.93(0.37)	3.01(0.10)	0.49(1.00)	0.55(0.15)
$\boldsymbol{\rho} = (-.5, -.5)$	$n = 250$	0.95(0.22)	2.00(0.08)	1.67(0.88)	3.09(0.23)	-0.60(0.16)	-0.42(0.37)
	$n = 500$	1.00(0.15)	1.97(0.06)	1.86(0.63)	3.03(0.17)	-0.48(0.14)	-0.44(0.27)
	$n = 1000$	0.99(0.11)	2.01(0.04)	1.99(0.42)	2.99(0.11)	-0.51(0.10)	-0.46(0.19)
$\boldsymbol{\rho} = (-.8, .8)$	$n = 250$	1.00(0.10)	2.01(0.04)	2.04(0.29)	2.99(0.08)	-0.79(0.06)	0.79(0.08)
	$n = 500$	0.99(0.21)	2.00(0.08)	2.16(0.66)	2.95(0.18)	-0.81(0.09)	0.70(0.23)
	$n = 1000$	1.01(0.15)	1.97(0.06)	2.19(0.47)	2.97(0.13)	-0.78(0.08)	0.69(0.17)
$\boldsymbol{\rho} = (.8, -.8)$	$n = 250$	0.98(0.22)	2.00(0.08)	1.64(0.67)	3.11(0.19)	0.76(0.12)	-0.71(0.23)
	$n = 500$	0.96(0.15)	1.98(0.06)	1.86(0.45)	3.03(0.13)	0.75(0.09)	-0.75(0.14)
	$n = 1000$	1.01(0.10)	2.01(0.04)	2.04(0.29)	2.99(0.08)	0.79(0.06)	-0.80(0.08)
$\boldsymbol{\rho} = (.8, .8)$	$n = 250$	0.98(0.22)	2.00(0.08)	2.16(0.70)	2.95(0.19)	0.76(0.12)	0.68(0.25)
	$n = 500$	0.96(0.15)	1.98(0.06)	2.18(0.47)	2.97(0.13)	0.75(0.09)	0.69(0.17)
	$n = 1000$	1.00(0.10)	2.01(0.04)	2.04(0.29)	2.99(0.08)	0.79(0.06)	0.79(0.08)
$\boldsymbol{\rho} = (-.8, -.8)$	$n = 250$	0.99(0.21)	2.00(0.08)	1.56(0.67)	3.12(0.19)	-0.81(0.09)	-0.71(0.23)
	$n = 500$	1.01(0.15)	1.97(0.06)	1.86(0.44)	3.03(0.12)	-0.78(0.08)	-0.76(0.14)
	$n = 1000$	1.00(0.11)	2.01(0.04)	2.03(0.29)	2.99(0.09)	-0.79(0.06)	-0.80(0.08)

Table 5: *Continuous confounder model: Posterior means and standard deviations, averaged over 20 replications to data simulated under different values of* $\boldsymbol{\rho}$ *and* n.

column to the sample size. The remaining columns provide the posterior means and standard

deviations averaged over the 20 replications. We find that the parameters are well estimated for the different degrees of confounding and sample size. As expected, the parameters are better estimated across all designs as the sample size is increased. Note, however, that the posterior distribution of $\boldsymbol{\beta}_1$ is more dispersed than that of $\boldsymbol{\beta}_0$. Also, the posterior mean of ρ_1, and to a smaller extent $\boldsymbol{\beta}_1$, are less well estimated than ρ_0 and $\boldsymbol{\beta}_0$, especially when the sample size is small ($n \leq 500$).

These findings reflect the point made earlier in our discussion, that the particular structure of the eligibility design makes it more difficult to separate the effect of $\boldsymbol{\beta}_1$ from the effect of ω_1. This is mirrored in the lower panel of Figure 3 which shows the scatter plot of the draws of ω_1

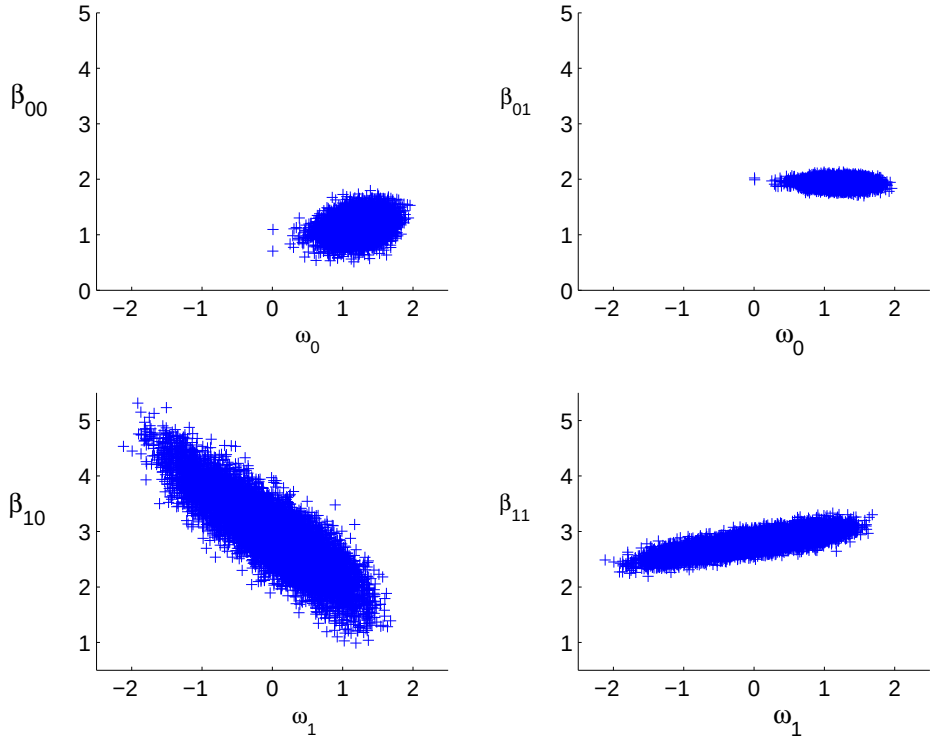


Figure 3: *Continuous confounder model: Scatter plots from the MCMC output of each element of* $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1})$ *and* ω_j*, for the case* $\boldsymbol{\rho} = (0.5, 0.5)$ *and* $n = 500$.

and the elements in $\boldsymbol{\beta}_1$ in the case when $\boldsymbol{\rho} = (0.5, 0.5)$ and $n = 500$. The scatter plots of ω_0 and the elements in $\boldsymbol{\beta}_0$ that are in the upper panel of Figure 3 display less dependence.

Naturally, the dependence in the joint distribution of $\boldsymbol{\beta}_1$ and ω_1 manifests itself in higher inefficiency factors for these parameters, as shown in Table 6 for the designs with $\boldsymbol{\rho} = (0.5, 0.5)$ and $\boldsymbol{\rho} = (0.8, 0.8)$. Notice that the inefficiency factors for $\boldsymbol{\gamma}$ are also relatively higher.

For a more detailed analysis of the performance of our MCMC algorithm we focus on the

	Inefficiency Factors									
Sample Size	$\boldsymbol{\beta}_0$		$\boldsymbol{\beta}_1$		$\boldsymbol{\gamma}$		σ_0^2	σ_1^2	ω_0	ω_1
$n = 250$	3.26	1.80	34.66	28.82	31.28	63.17	9.59	26.88	10.06	48.28
	3.28	2.19	20.78	16.83	20.31	35.76	8.96	32.92	7.62	32.70
$n = 500$	2.49	1.42	31.12	26.09	29.51	57.23	6.90	28.31	7.89	44.68
	2.51	1.88	27.39	23.46	20.81	36.54	10.91	24.01	9.97	34.18
$n = 1000$	2.28	1.43	20.32	15.96	25.68	54.95	7.25	28.99	8.07	31.03
	3.03	1.82	28.54	17.73	34.53	72.83	20.12	52.46	20.20	50.59

Table 6: *Continuous confounder model: Inefficiency factors for designs with* $\boldsymbol{\rho} = (0.5, 0.5)$ *and* $\boldsymbol{\rho} = (0.8, 0.8)$ *(second line), averaged over 20 replications.*

results by replication when the data is generated with $\boldsymbol{\rho} = (0.5, 0.5)$. The three panels in

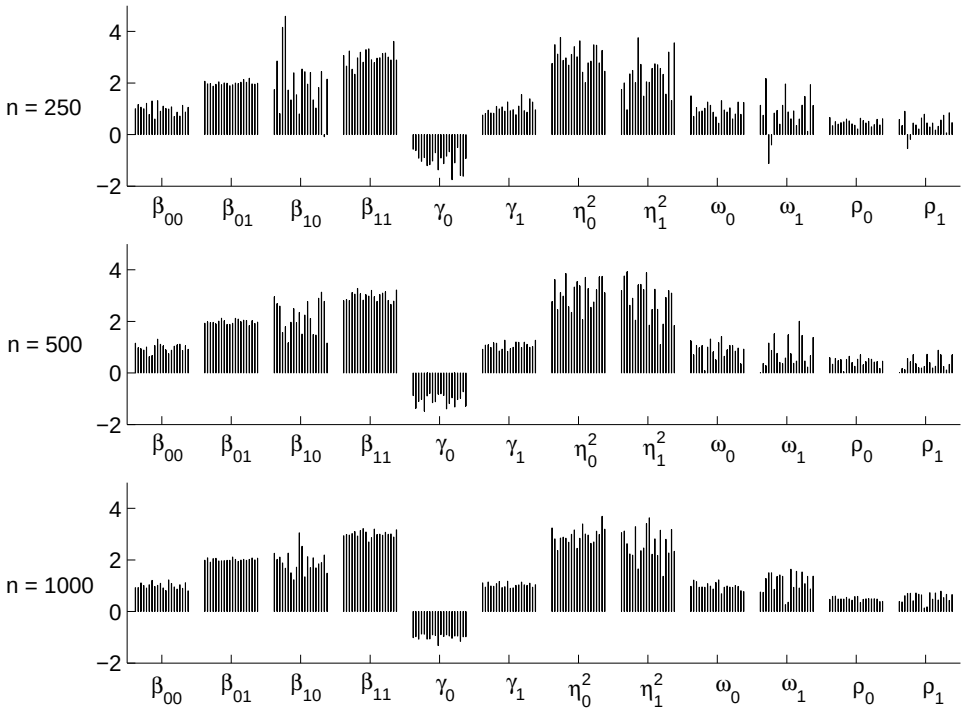


Figure 4: *Posterior means of the model parameters for the 20 replications, by sample size.*

Figure 4 display the posterior means of the model parameters from the 20 replications for each sample size. We see that the variation across replications is generally small and tends to decline considerably with an increase in the sample size.

We next turn to the estimation of the PACE. Our results are given in Table 7 where we report our estimates of this effect, averaged over the 20 replications, for each of the simulation designs in Table 5. To see how well our method performs we compare these results to the true average treatment effect. We compute the true effect as the sample mean from the draws of

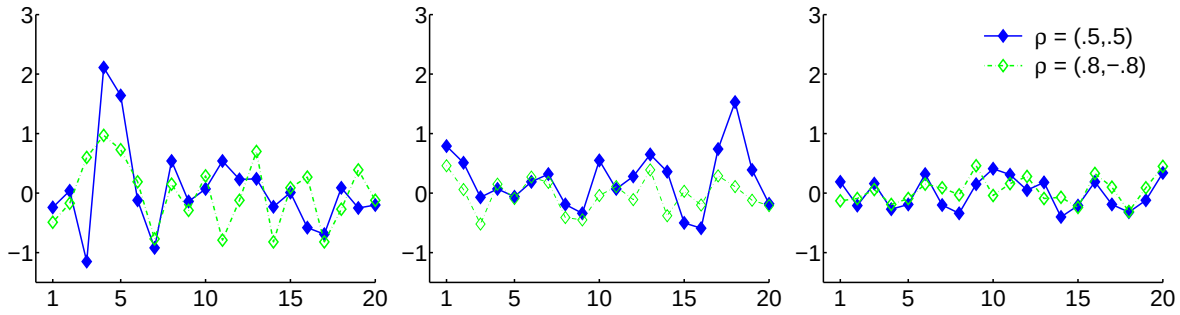
the potential outcomes under both treatment states for all individuals in the sample. Averaging these results over the 20 replications of each design, the true values of the treatment effect are found to be 3.00, 3.01 and 2.93 for $n = 250$, 500 and 1000 respectively. These effects do not vary by $\boldsymbol{\rho}$ because, by definition, the causal effect is computed from the marginal distribution of the potential outcomes. The results in Table 7 indicate that the estimated PACE is close to

	Predictive Average Causal Effect (PACE)							
ρ_0 :	−0.5	0.5	0.5	−0.5	−0.8	0.8	0.8	−0.8
ρ_1 :	0.5	−0.5	0.5	−0.5	0.8	−0.8	0.8	−0.8
$n = 250$	3.07	3.28	3.06	2.96	3.13	2.94	3.13	2.91
$n = 500$	3.16	2.99	3.17	2.98	3.18	2.99	3.19	2.97
$n = 1000$	2.94	2.99	2.93	2.96	3.00	2.98	2.99	2.98

Table 7: *Continuous confounder model: Estimated PACE, averaged over 20 replications.*

the true average treatment effect. Our results also indicate that the estimate PACE is slightly larger than the true effect though the deviations are under 10% even for the smaller sample size. As expected, the estimate improves with the sample size.

To complete the analysis we look at the difference between the estimated PACE and the true average causal treatment effect, by replication. These differences are given in Figure 5, by sample size, for data generated under the designs $\boldsymbol{\rho} = (.5, .5)$ and $\boldsymbol{\rho} = (.8, -.8)$. We observe that



(a) $n = 250$ **(b)** $n = 500$ **(c)** $n = 1000$

Figure 5: *Continuous confounder model: Difference between the estimated PACE and the true causal effect, by sample size, and different degrees of confounding, across replications.*

the true treatment effect is not as well estimated when the sample size is small and the degree of confounding is low. However, our estimates improve considerably with sample size and degree of confounding.

4 Model Comparison

4.1 Estimation of the Marginal Likelihood

The developments in the previous sections open up the possibility of fitting eligibility design data under each framework, with the aim of appraising which model and assumptions are better supported by the data. With this in mind, we compute the support for each model in terms of the Bayes factor, which we compute from the MCMC output by the approach of Chib (1995).

We begin with the basic marginal likelihood identity in Chib (1995) and express the marginal likelihood of the model M in terms of the likelihood function, the joint prior and the posterior as

$$\ln m(\mathbf{y}, \mathbf{x}|M) = \ln f(\mathbf{y}, \mathbf{x}|\mathbf{z}, \mathbf{W}, \boldsymbol{\psi}^*, M) + \ln \pi(\boldsymbol{\psi}^*|M) - \ln \pi(\boldsymbol{\psi}^*|\mathbf{y}, \mathbf{x}, \mathbf{W}, M) \tag{21}$$

each evaluated at $\boldsymbol{\psi}^*$ (the posterior mean of the parameter vector $\boldsymbol{\psi}$). Clearly, the first two terms in this expression can be found directly. For instance, the likelihood ordinate of the discrete confounder model comes from (2) and (3) whereas that of continuous confounder model from expressions (12) and (13). To estimate the posterior ordinate we proceed as follows.

Consider first the case of the discrete confounder model. In this case, the parameter vector is $\boldsymbol{\psi} = (\boldsymbol{\beta}, \boldsymbol{\eta}^2, q_c)$, where $\boldsymbol{\beta} = (\boldsymbol{\beta}_{0c}, \boldsymbol{\beta}_{0n}, \boldsymbol{\beta}_{1c})$ and $\boldsymbol{\eta}^2 = (\eta^2_{0c}, \eta^2_{0n}, \eta^2_{1c})$. To estimate $\pi(\boldsymbol{\psi}^*|\mathbf{y}, \mathbf{x}, \mathbf{W})$ we utilize the decomposition

$$\pi(\boldsymbol{\psi}^*|\mathbf{y}, \mathbf{x}, \mathbf{W}) = \pi(\boldsymbol{\eta}^{2^*}|\mathbf{y}, \mathbf{x}, \mathbf{W})\pi(\boldsymbol{\beta}^*|\mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\eta}^{2^*})\pi(q_c^*|\mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\eta}^{2^*}, \boldsymbol{\beta}^*)$$

where $\boldsymbol{\eta}^2 = (\eta^2_{0c}, \eta^2_{0n}, \eta^2_{1c})$. The marginal posterior ordinate of $\boldsymbol{\eta}^{2^*}$ can be estimated from expression (6) as

$$\pi(\boldsymbol{\eta}^{2^*}|\mathbf{y}, \mathbf{x}, \mathbf{W}) = M^{-1} \sum_{g=1}^{M} \prod_{j=0,1} \prod_{k \in K_j} \pi(\eta^{2(g)}_{j,k}|\mathbf{y}_{jk}, \mathbf{x}_{jk}, \mathbf{W}_{jk}, \boldsymbol{\beta}^{(g)}_{jk}, \boldsymbol{\lambda}^{(g)}_{jk}, q^{c^{(g)}})$$

where $K_0 = \{c, n\}$, $K_1 = c$ and $(\boldsymbol{\beta}^{(g)}_{jk}, \boldsymbol{\lambda}^{(g)}_{jk}, q^{c^{(g)}})$ is the gth draw from the MCMC run. Next, we estimate $\pi(\boldsymbol{\beta}^*|\mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\eta}^{2^*})$ from the output of a reduced run of the MCMC algorithm in which $\boldsymbol{\eta}^2$ is set at $\boldsymbol{\eta}^{2^*}$. This output is used to average the conditional densities in (5) for $j = 0, 1$ and $k \in K_j$. Finally, we get the reduced ordinate $\pi(q_c^*|\mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\eta}^{2^*}, \boldsymbol{\beta}^*)$ by averaging expression (7) over the output of a second reduced run with $(\boldsymbol{\eta}^2, \boldsymbol{\beta}^2)$ fixed at $(\boldsymbol{\eta}^{2^*}, \boldsymbol{\beta}^{2^*})$.

In the case of the continuous confounder model the parameter is $\boldsymbol{\psi} = (\boldsymbol{\gamma}, \boldsymbol{\psi}_0, \boldsymbol{\psi}_1)$, where as before $\boldsymbol{\psi}_j = (\boldsymbol{\beta}_j, \sigma_j^2, \omega_j)$. We decompose the posterior ordinate into a product of marginal and reduced posterior ordinates as

$$\pi(\boldsymbol{\psi}^*|\mathbf{y}, \mathbf{x}, \mathbf{W}) = \pi(\boldsymbol{\zeta}_0^*|\mathbf{y}, \mathbf{x}, \mathbf{W})\pi(\boldsymbol{\gamma}^*|\mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\zeta}_0^*)\pi(\sigma_1^{2^*}|\mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\zeta}_0^*, \boldsymbol{\gamma}^*)\pi(\boldsymbol{\beta}_0^*, \boldsymbol{\beta}_1^*, \omega_1^*|\mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\zeta}_0^*, \boldsymbol{\gamma}^*, \sigma_1^{2^*})$$

where $\boldsymbol{\zeta}_0^* = (\sigma_0^{2^*}, \omega_0^*)$.

In this case we cannot average the conditional ordinate of $\boldsymbol{\zeta}_0$ to obtain $\pi(\boldsymbol{\zeta}_0^*|\mathbf{y}, \mathbf{x}, \mathbf{W})$ because the normalizing constant of the conditional ordinate is not known. Nonetheless, $\pi(\boldsymbol{\zeta}_0^*|\mathbf{y}, \mathbf{x}, \mathbf{W})$ can be estimated by the method of Chib and Jeliazkov (2001). An application of their method shows that

$$\pi(\boldsymbol{\zeta}_0^*|\mathbf{y}, \mathbf{x}, \mathbf{W}) = \frac{E_1[\alpha(\boldsymbol{\zeta}_0, \boldsymbol{\zeta}_0^*|\mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\beta}, \sigma_1^2, \omega_1, \mathbf{x}^*, \boldsymbol{\lambda})q(\boldsymbol{\zeta}_0^*|\mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\beta}, \sigma_1^2, \omega_1, \mathbf{x}^*, \boldsymbol{\lambda})]}{E_2[\alpha(\boldsymbol{\zeta}_0^*, \boldsymbol{\zeta}_0|\mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\beta}, \sigma_1^2, \omega_1, \mathbf{x}^*, \boldsymbol{\lambda})]}$$

where the expectation E_1 in the numerator is taken with respect to the posterior distribution $\pi(\boldsymbol{\zeta}_0, \boldsymbol{\beta}, \sigma_1^2, \omega_1, \mathbf{x}^*, \boldsymbol{\lambda}|\mathbf{y}, \mathbf{x}, \mathbf{W})$ and the expectation E_2 in the denominator is taken with respect to the conditional product measure $\pi(\boldsymbol{\beta}, \sigma_1^2, \omega_1, \mathbf{x}^*, \boldsymbol{\lambda}|\mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\zeta}_0^*)q(\boldsymbol{\zeta}_0|\mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\beta}, \sigma_1^2, \omega_1, \mathbf{x}^*, \boldsymbol{\lambda})$. Thus, to estimate the numerator we use the output of the full MCMC run average expression $\alpha(\boldsymbol{\zeta}_0, \boldsymbol{\zeta}_0^*|\mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\beta}, \sigma_1^2, \omega_1, \mathbf{x}^*, \boldsymbol{\lambda})q(\boldsymbol{\zeta}_0^*|\mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\beta}, \sigma_1^2, \omega_1, \mathbf{x}^*, \boldsymbol{\lambda})$. To estimate the expectation in the denominator we average $\alpha(\boldsymbol{\zeta}_0^*, \boldsymbol{\zeta}_0|\mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\beta}, \sigma_1^2, \omega_1, \mathbf{x}^*, \boldsymbol{\lambda})$, where the draws on the parameters are from a reduced run where $\boldsymbol{\zeta}_0$ is fixed at $\boldsymbol{\zeta}_0^*$, and $\boldsymbol{\zeta}_0$ is drawn from $q(\boldsymbol{\zeta}_0|\mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\beta}, \sigma_1^2, \omega_1, \mathbf{x}^*, \boldsymbol{\lambda})$.

Notice that the output from this reduced run can also be used to obtain the reduced ordinate of $\boldsymbol{\gamma}$ as

$$\pi(\boldsymbol{\gamma}^*|\mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\zeta}_0^*) = M^{-1}\sum_{g=1}^{M}\pi(\boldsymbol{\gamma}^*|\mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\zeta}_0^*, \{x_i^{*(g)}\}, \boldsymbol{\beta}_0^{(g)}, \boldsymbol{\beta}_1^{(g)}, \sigma_1^{2^{(g)}}, \omega_1^{(g)}, \{\lambda_i^{(g)}\})$$

where the summand is the density given in (16).

Next, to estimate $\pi(\sigma_1^{2^*}|\mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\zeta}_0^*, \boldsymbol{\gamma}^*)$ we fix $(\boldsymbol{\zeta}_0, \boldsymbol{\gamma})$ at $(\boldsymbol{\zeta}_0^*, \boldsymbol{\gamma}^*)$ and continue the MCMC iterations. The draws from this reduced run produce the estimate

$$M^{-1}\sum_{g=1}^{M}\pi(\sigma_1^{2^*}|\mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\zeta}_0^*, \boldsymbol{\gamma}^*, \boldsymbol{\beta}_0^{(g)}, \boldsymbol{\beta}_1^{(g)}, \sigma_1^{2^{(g)}}, \omega_1^{(g)}, \{x_i^{*(g)}\}, \{\lambda_i^{(g)}\})$$

where the conditional density of σ_1^2 is given (20).

Finally, we compute the reduced ordinate of $(\boldsymbol{\beta}_0^*, \boldsymbol{\beta}_1^*, \omega_1^*)$ from the output of the final reduced run, now with $(\boldsymbol{\zeta}_0, \boldsymbol{\gamma}, \sigma_1^2)$ fixed at $(\boldsymbol{\zeta}_0^*, \boldsymbol{\gamma}^*, \sigma_1^{2^*})$, to yield

$$M^{-1} \sum_{g=1}^{M} \pi(\boldsymbol{\beta}_0^* | \mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\zeta}_0^*, \boldsymbol{\gamma}^*, \sigma_1^{2^*}, \{x_i^{*(g)}\}, \{\lambda_i^{(g)}\}) \pi(\boldsymbol{\beta}_1^*, \omega_1^* | \mathbf{y}, \mathbf{x}, \mathbf{W}, \boldsymbol{\zeta}_0^*, \boldsymbol{\gamma}^*, \sigma_1^{2^*}, \{x_i^{*(g)}\}, \{\lambda_i^{(g)}\})$$

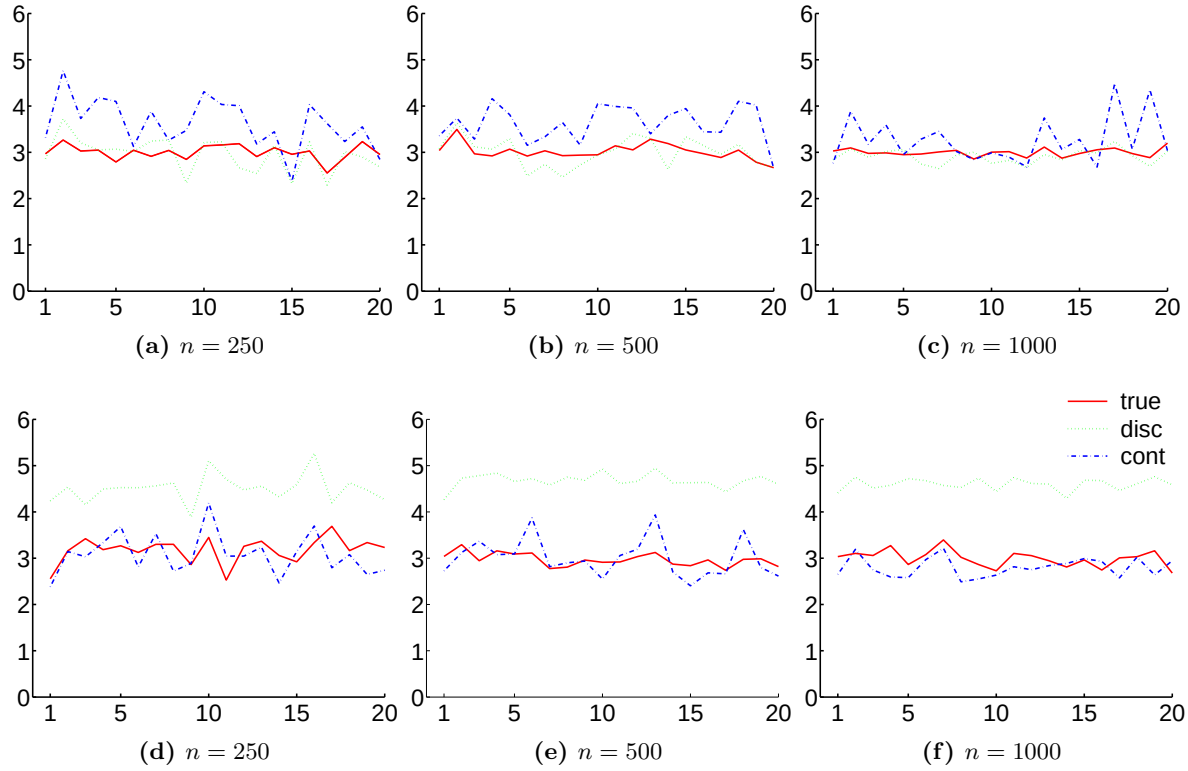
where the first conditional density is given in equation (17) and the second in equation (18).

4.1.1 Simulation Example

We now examine in simulation studies to what extent each of the two frameworks is supported by the data when the data comes from one or the other model. We consider the performance of each modeling approach in terms of inferences about the treatment effect and in terms of the model marginal likelihoods (which we estimate by the methods just outlined).

For consistency, our simulation designs are based on those in Sections 2.1.3 and 3.1.3. In particular, we restrict our attention to models with no covariates. Furthermore, in the discrete confounder model we let $\beta_{0c} = 2$, $\beta_{0n} = -1$ and $\beta_{1c} = 5$ and set the variance parameters η_{jk} at 4. Last we fix the proportion of compliers at 0.8. For the continuous confounder model, we let $\beta_0 = 2$, $\beta_1 = 5$, $\gamma = .5$, $\eta_j^2 = 4$ and fix the vector of correlation coefficients at $\boldsymbol{\rho} = (-0.8, 0.8)$. We generate 20 data sets under each design with sample sizes of 250, 500 and 1000. In the fitting we adopt the prior distributions given in Sections 2.1.3 and 3.1.3.

Our first set of results appear in Figure 6 where, for each of the 20 replications, we give the true average treatment effect (solid line) together with the estimates of this effect from the discrete confounder approach (dotted line) and the continuous confounder approach (dashed line). The upper panel of this figure contains the results when the data is generated from the discrete confounder model and the lower panel when the data comes from the continuous confounder model. Not surprisingly, the best estimates of the treatment effect arise when the correct model is fit to the data. We can also see (on contrasting the graphs in the top and bottom panels) that the fit from the discrete confounder model is especially good when it is used in the correct circumstance though it does much less well when it is applied to the wrong data. This problem persists even when the sample size is large. On the other hand, the continuous confounder model appears to be more resistant to departures from its assumptions, at least as



(a) $n = 250$ **(b)** $n = 500$ **(c)** $n = 1000$

(d) $n = 250$ **(e)** $n = 500$ **(f)** $n = 1000$

Figure 6: *Cross Fitting: True average treatment effects by replication (solid line) and the estimates from the discrete confounder model (dotted line) and the continuous confounder model (dash-dot line) when the data is generated from the discrete confounder model with* $q_c = .8$ *(upper panel) and the continuous confounder model with* $\boldsymbol{\rho} = (-.8, .8)$ *(lower panel), for each of three different sample sizes.*

long as the sample size is moderately large.

We next examine the support for the two approach in terms of the model marginal likelihoods. Our results are summarized in Table 8. In particular, columns 4 and 5 provide the log marginal likelihood estimates averaged over 20 replications. The next two columns give the total number of replications (out of 20) in which a particular model is favored by the data according to the marginal likelihood criterion. These summaries are provided for four alternative designs and three different sample sizes. The first two of the four designs are the designs utilized thus far. In the remaining two designs we examine to what extent inferences depend on the two key parameters in each framework: the proportion of compliers in the discrete confounder model and the degree of confounding in the continuous confounder model. In particular, in our third design we generate data with a lower proportion of compliers and in the last design with a lower degree of confounding.

The results in Table 8 show that in the two initial designs the continuous confounder model is essentially always picked when the data comes from this model. Interestingly, even when the

True Model			Log Marg. Lik. Disc	Log Marg. Lik. Cont	Successes Disc	Successes Cont
Discrete	$q_c = .8$	$n = 250$	-663.11	-660.73	0	20
		$n = 500$	-1306.71	-1305.55	7	13
		$n = 1000$	-2586.81	-2573.34	14	6
Continuous	$\boldsymbol{\rho} = (-.8, .8)$	$n = 250$	-611.03	-608.30	1	19
		$n = 500$	-1207.46	-1188.77	1	19
		$n = 1000$	-2404.00	-2395.92	0	20
Discrete	$q_c = .4$	$n = 250$	-693.08	-693.77	14	6
		$n = 500$	-1371.96	-1374.18	19	1
		$n = 1000$	-2720.58	-2722.77	16	4
Continuous	$\boldsymbol{\rho} = (-.5, .5)$	$n = 250$	-635.70	-634.58	3	17
		$n = 500$	-1246.17	-1241.03	4	16
		$n = 1000$	-2475.21	-2470.65	6	14

Table 8: *Results from the cross-fitting: Columns 3 and 5 give the number of replications (out of the total of 20 replications) in which this model wins according to the marginal likelihood criterion. Columns 4 and 6 show the estimate of the log marginal likelihood averaged over 20 replications.*

data originates from the discrete confounder model, the continuous confounder model is favored in the majority of cases, except with a sample size of $n = 1000$ when the discrete confounder model is picked in 14 out of the 20 replications.

Moreover, based on the lower half of Table 8, we see that the support for the continuous confounder model over the discrete confounder model is not affected by a lowering of the degree of confounding. On the other hand, when the data comes from a discrete confounder design with a smaller, more balanced proportion of compliers, the discrete confounder model becomes the favored model.

Thus, these experiments point to the conclusion that the continuous confounder model is almost always the better supported model when the data comes from that model, regardless of the sample size or the extent of confounding, and that when the data originate from the discrete confounder model, the discrete confounder model is favored over the continuous model only when the sample size is large and the proportion of compliers and never-takers in the sample is more or less equal.

5 Job Training and Depression

We now illustrate the twin approaches for dealing with unobserved confounders on a real data set from the JOBS II Intervention Project at the University of Michigan (see Vinokur et al 1995).

From this trial, a study sample of 502 "high risk" subjects who had become unemployed in the preceding 13 weeks and were searching for new employment were selected and 335 subjects were randomized into a training program that consisted of five half-day sessions designed to enhance job search skills, self esteem and sense of control. The goal of the study was to see if such training causally affected the change in a depression score from baseline to six months after the training, with a negative outcome representing an improvement in mental health.

In Table 9 we present a summary of the study data in terms of the sample mean and sample standard deviation for the outcome y, the assignment z, the intake x and the nine covariates. The summary stratified according to the intake is in the last two columns of the table. Overall,

Variable	Explanation		Means (Standard Deviations)			
			All		$x=0$	$x=1$
y	change in depression score		-0.43	(0.78)	-0.39	-0.51
z	assignment	Control: 167 Treatment: 335	0.67			
x	treatment intake	Control: 0 Treatment: 183	0.36			
	Covariates					
Basedep	baseline depression score		2.45	(0.30)	2.46	2.42
Baserisk	baseline risk score		1.68	(0.21)	1.69	1.67
Age	age in years		36.58	(9.96)	34.81	39.68
Motivate	motivation to attend		5.33	(0.81)	5.24	5.50
Educ	school grade completed		13.37	(2.02)	13.13	13.77
Assert	assertiveness		3.07	(0.92)	3.13	2.96
Single	dummy		0.62		0.61	0.65
Econ	economic hardship		3.61	(0.87)	3.62	3.60
Nonwhite	dummy		0.19		0.21	0.15

Table 9: *Sample summary statistics of the study data from the JOBS II Intervention Project*

the sample mean of the outcome is -0.43 whereas by intake the mean is -0.39 for those who had $x = 0$ and -0.51 for those with $x = 1$. This unpolished comparison points to (but not does not of course establish) an improvement in mental health due to the training because some of this improvement is perhaps due to unobserved confounders (only 183 of the 335 subjects in the treatment arm chose to participate in the training) and observed confounders (the distribution of the covariates by intake being somewhat different - for example, program participants were older, more educated, less assertive and more likely white).

In previous analyses of these data, Jo (2002) and Skrondahl and Rabe-Hesketh (2004) utilized the assumption of a discrete compliance confounder and estimated the complier average causal effect of the training. We now re-examine these data, focusing on what was previously not possible, namely, the comparison of the discrete and continuous modeling approaches in order

to isolate the sensitivity of the inferential predictions to these key modeling assumptions.

To start with we analyze these data from the discrete confounder perspective. Letting y_i denote the change in the depression score over the 6 month period, we assume that

$$p_0(y_i|\mathbf{w}_i, s_i = n) = t_{15}(y_i|\mathbf{w}_i'\boldsymbol{\beta}_{0,n}, \eta^2_{0,n})$$
$$p_j(y_i|\mathbf{w}_i, s_i = c) = t_{15}(y_i|\mathbf{w}_i'\boldsymbol{\beta}_{j,c}, \eta^2_{j,c}),\ j = 0, 1 \tag{22}$$

where, following Skrondahl and Rabe-Hesketh (2004), $\mathbf{w}_i' = [1, basedep, baserisk]$. We further assume that apriori $\boldsymbol{\beta}_{j,k} \sim \mathcal{N}_3(\mathbf{0}, 25 \times \mathbf{I}_3)$, that $\eta^2_{j,k}$ is distributed as inverse gamma with a prior mean of 1 and standard deviation of 2, and that $q_c \sim \mathcal{B}(2, 2)$ with a mean of 0.5 and standard deviation of 0.2. A summary of the results from this fitting is given in Table 10. The first three

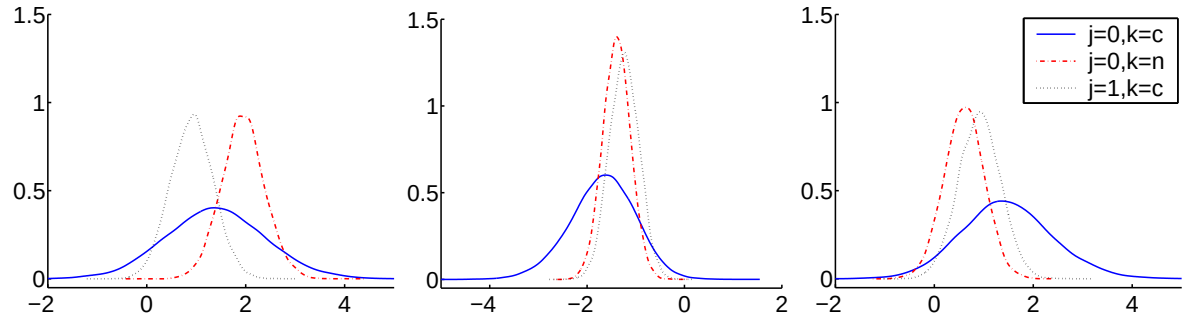
	Discrete C. Model			Continuous C. Model	
			Outcome model		
	$\boldsymbol{\beta}_{0c}$	$\boldsymbol{\beta}_{0,n}$	$\boldsymbol{\beta}_{1,c}$	$\boldsymbol{\beta}_0$	$\boldsymbol{\beta}_1$
Intercept	1.39(1.00)	1.93(0.44)	0.92(0.43)	1.82(0.35)	0.69(0.43)
Basedep	-1.68(0.67)	-1.39(0.29)	-1.24(0.30)	-1.46(0.22)	-1.20(0.30)
Baserisk	1.45(0.93)	0.61(0.41)	0.92(0.42)	0.82(0.32)	0.86(0.43)
	η^2_{0c}	$\eta^2_{0,n}$	$\eta^2_{1,c}$	η^2_0	η^2_1
	0.53(0.14)	0.43(0.06)	0.42(0.05)	0.47(0.04)	0.33(0.08)
			Intake model		
Intercept				0.28(0.05)	
Age - 20				0.04(0.01)	
Motivate				0.41(0.10)	
Educ				0.18(0.04)	
Assert				-0.22(0.09)	
Single				0.35(0.16)	
Econ				-0.09(0.10)	
Nonwhite				-0.29(0.19)	
		Compliance		*Confounding*	
		q_c		ρ_0	ρ_1
		0.54(0.03)		0.17(0.13)	0.49(0.21)
			Log Marginal Likelihoods		
		-804.86		-791.03	

Table 10: *Estimation results from the discrete and the continuous confounder models. The results are based on 10,000 runs of the MCMC algorithms (1,000 burn-in cycles). Table entries are the posterior means (standard deviation in parentheses), along with the log marginal likelihoods.*

columns of the table, under the heading "outcome model," contain the posterior means and standard deviations of $\boldsymbol{\beta}_{0,c}$, $\boldsymbol{\beta}_{0,n}$ and $\boldsymbol{\beta}_{1,c}$. In conformity with the evidence provided in Section

2.1.3, one notices that the posterior standard deviation of each element of $\boldsymbol{\beta}_{0,c}$ is nearly two times that of the corresponding element in $\boldsymbol{\beta}_{0,n}$ and $\boldsymbol{\beta}_{1,c}$, a consequence of the fact that the proportion of compliers is not large (the posterior mean of q_c is .54 with a posterior standard deviation of .03). Unfortunately, this relative paucity of compliers is not compensated by a larger sample size (the sample size in the control arm is 167).

One also notices a certain overlap in the marginal posterior distributions of $\boldsymbol{\beta}_{0,c}$, $\boldsymbol{\beta}_{0,n}$. This overlap, if pronounced, would be an informal challenge to the assumption of two types and to the presence of confounding through this route. With this in mind, we report in Figure 7 the kernel-smoothed posterior density plots for each component of $\boldsymbol{\beta}_{0c}$ (solid line), $\boldsymbol{\beta}_{0n}$ (dashed line) and $\boldsymbol{\beta}_{1c}$ (dotted line). Although the marginal posterior densities of $\boldsymbol{\beta}_{0c}$ and $\boldsymbol{\beta}_{0n}$ are different,



(a) $\beta_{jk,\text{intercept}}$ **(b)** $\beta_{jk,\text{basedep}}$ **(c)** $\beta_{jk,\text{baserisk}}$

Figure 7: *Posterior density plots of* $\boldsymbol{\beta}_{0c}$, $\boldsymbol{\beta}_{0n}$ *and* $\boldsymbol{\beta}_{1c}$.

the overlap is visible.

We conclude this analysis by noting that the imprecision in the estimation of $\boldsymbol{\beta}_{0c}$ carries over into inferences about the treatment effect. Following the approach outlined in Section 2.1.2, we compute the marginal predictive densities of each potential outcome, confined to compliers, from which we compute the predictive quantile effects, and the PACE. The results are given in Table 5. We see that the estimated quantile treatment effects of the training vary from -0.12 to -0.55,

	Treatment Effects Estimates					
		Quantile Effects				
Model	PACE	0.05	0.25	0.5	0.75	0.95
Discrete C. Model	-0.31	-0.12	-0.23	-0.31	-0.39	-0.55

Table 11: *Estimates of the predictive average and quantile causal effects for the discrete confounder model. The results are based on 10,000 runs of the MCMC algorithms (1,000 burn-in cycles).*

whereas the estimate of the PACE is -0.31. Quite remarkably the latter effect is identical to the complier average causal effect found previously by Skrondahl and Rabe-Hesketh (2004) from a

non-predictive perspective and a different estimation framework.

Turning to the continuous confounder model we assume that the marginal distribution of the outcome in intake state j is

$$p_j(y_i|\mathbf{w}_i, \boldsymbol{\beta}_j, \eta_j^2) = t_{15}(y_i|\mathbf{w}_i'\boldsymbol{\beta}_j, \eta_j^2)$$

where, as before, the change in depression score depends on the covariates "basedep" and "baserisk." In addition, the marginal distribution of the intake is specified in terms of the process

$$x_i = I[\tilde{\mathbf{w}}_i'\boldsymbol{\gamma} + u_i > 0] \ , \ u_i \sim t_{15}(0, 1),$$

where the covariates $\tilde{\mathbf{w}}_i$ are $[1, age - 20, motivate, educ, assert, single, econ, nonwhite]$, in keeping with a model suggested by Skrondahl and Rabe-Hesketh (2004) for compliance. Finally, the errors (ε_j, u_i) are modeled jointly in terms of a student-t density with 15 degrees of freedom with covariance parameter ω_j. In our fitting we assume that $\boldsymbol{\beta}_j \sim \mathcal{N}_3(\mathbf{0}, 25 \times \mathbf{I}_3)$, $\boldsymbol{\gamma} \sim \mathcal{N}_8(\mathbf{0}, 25 \times \mathbf{I}_8)$, $\boldsymbol{\omega}_j \sim \mathcal{N}(0, 4)$, for $j = 0, 1$ and let the hyperparameters $\nu_{0,j}$ and $d_{0,j}$ in the prior of η_j^2 be such that the implied mean and standard deviation are 1 and 2, respectively.

A summary of the results from this fitting appear in Table 10. One thing to notice is that the posterior standard deviations of the elements of $\boldsymbol{\beta}_0$ are lower than those of $\boldsymbol{\beta}_{0,c}$ and $\boldsymbol{\beta}_{0,n}$, while those of $\boldsymbol{\beta}_1$ are about the same size as those of $\boldsymbol{\beta}_{1,c}$. In agreement with the findings of Section 3.1.3, we see that the $\boldsymbol{\beta}_1$ is less precisely estimated than $\boldsymbol{\beta}_0$, however, it appears that for these data the effect of the covariates on the outcome is not obscured by the effect of the confounders (that the confounders are present is revealed in the posterior distribution of the correlation coefficients ρ_0 and ρ_1 whose means are 0.17 and 0.49, and standard deviations are 0.13 and 0.21, respectively). To show this, we have plotted in Figure 8 the bivariate scatter plots of ω_j against each element of $\boldsymbol{\beta}_j$. Although it is clear from this graph that the joint distribution of ω_1 and $\boldsymbol{\beta}_1$ is more dispersed than that of ω_0 and $\boldsymbol{\beta}_0$, there is reassuringly little evidence of extreme pairwise dependence. In fact, these parameters are effectively aposteriori uncorrelated. That such limited dependence between ω_j and $\boldsymbol{\beta}_j$ can be a feature of the posterior distribution in such models is quite interesting.

Given these positive findings, the estimates of the treatment effect under this model take on some added importance. These causal effects, after accounting for the unobserved confounders,

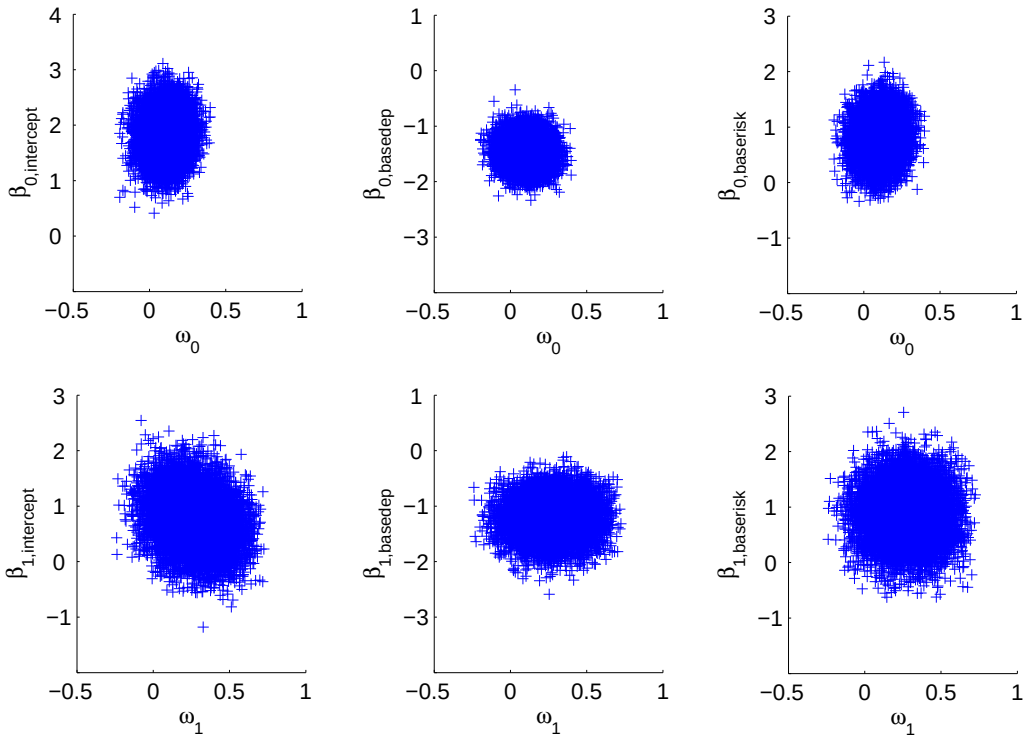


Figure 8: *Scatter plots of the draws of ω_j and the elements in $\boldsymbol{\beta}_j$.*

are reported in Table 12. We see from the table that the estimate of the PACE is -0.45, a larger effect than that from the discrete confounder model, and that the various quantile effects are

	Treatment Effects Estimates					
		Quantile Effects				
Model	PACE	0.05	0.25	0.5	0.75	0.95
Continuous C. Model	-0.45	-0.48	-0.44	-0.44	-0.46	-0.45

Table 12: *Estimates of predictive average and quantile causal effects for the continuous confounder model. The results are based on 10,000 runs of the MCMC algorithms (1,000 burn-in cycles).*

almost identical.

In summary, the evidence from the separate fitting of these two models appears to informally support the view that the continuous confounder framework is more appropriate for these data. For substantiation of this view we calculate the log marginal likelihood of each model by the method described in Section 4. We find that the log marginal likelihood of the discrete confounder model and the continuous confounder models are -804.86 and -791.03, respectively, an indication of strong support on Jeffreys' scale in favor of the continuous confounder model.

6 Conclusion

There are many problems in practice, with unobserved confounders, which share the structure of the eligibility design that we have studied in this paper. Our work is broadly relevant to such problems. Particularly useful for future work in this area is our reformulation of the discrete confounder approach and the development of the continuous confounder approach. In practice both approaches can now be applied and compared via marginal likelihoods, Bayes factors, and inferential predictions in a way that was not possible before.

Our extensive simulation experiments elucidate the strengths and weaknesses of each approach. In particular, we find that the continuous confounder approach offers a viable approach for modeling confounding, one that appears more resistant to model perturbations than the discrete confounder approach. We also find that the discrete confounder model despite its apparent simplicity is not necessarily easy to fit, especially when the sample size is small and/or the proportion of compliers is high. In these circumstance, our studies show that even when the data originate from the discrete confounder model, it is generally not favored over the continuous confounder model. On the other hand, when the data arise from the continuous confounder model, the continuous confounder model is almost always the better supported model, regardless of the sample size or the extent of confounding.

We conclude by mentioning that the analysis described in this paper extends readily to other settings. These extensions, for example involving clustered outcomes and binary responses, are ongoing and will be reported elsewhere.

References

[1] Albert, J. , 2002, "Estimating efficacy in clinical trials with clustered binary responses," Statistics in Medicine 21, 649-661.

[2] Albert, J., Chib, S., 1993, "Bayesian Analysis of Binary and Polychotomous Response Data," Journal of the American Statistical Association 88, 669-779.

[3] Bilodeau, M., Brenner, D., 1999, *Theory of Multivariate Statistics.* Springer, New York.

[4] Chib, S., 1995, "Marginal Likelihood from the Gibbs Output," Journal of the American Statistical Association 90, 1313-1321.

[5] Chib, S., 2001, "Markov Chain Monte Carlo Methods: Computation and Inference," in *Handbook of Econometrics* volume 5, (eds J.J. Heckman and E. Leamer), North Holland, Amsterdam, 3569-3649.

[6] Chib, S., 2004, "Analysis of Treatment Response Data without the Joint Distribution of Counterfactuals," manuscript, John M Olin School of Business, Washington University, St. Louis.

[7] Chib, S., Greenberg, E., 1995, "Understanding the Metropolis-Hastings algorithm," American Statistician 49, 327-35.

[8] Frangakis, C.F., Rubin, D. B., 1999, "Adressing complications of intention- to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes," Biometrika 86, 365-379.

[9] Imbens G.W., Rubin D.B., 1997, "Bayesian Inference for causal effects in randomized with noncompliance," The Annals of Statistics, 25, 305-327.

[10] Jo, B. 2002, "Estimation of intervention effects with noncompliance: Alternative model specifications," Journal of Educational and Behavioral Statistics, 27, 385-409.

[11] Levy, D. E., O'Malley, J. A., Normand, S. T., 2004, "Covariate adjustment in clinical trials with non-ignorable missing data and non-compliance," Statistics in Medicine 23, 2319-2339.

[12] Mealli, F., Imbens, G. W., Ferro, S., Biggeri, A., 2004, "Analyzing a randomized trial on breat self-examination with non-compliance and missing outcomes," Biostatistics 5, 207-222.

[13] Shaffer M.L., Chinchilli V.M., 2004, "Bayesian inference for randomized clinical trials with treatment failures," Statistics in Medicine, 23, 1215-1228.

[14] Sommer A., Zeger, S., 1991, "On estimating efficacy in clinical trials," Statistics in Medicine 10,45-52.

[15] Skrondahl A., Rabe-Hesketh, S., 2004, *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Model.* Chapman & Hall/CRC, New York.

[16] Ten Have T.R., Joffe M., Cary, M., 2003, "Causal logistic models for non-compliance under ranomized treatment with univariate binary response," Statistics in Medicine, 22, 1255-1283.

[17] Vinokur, A. D., Price, R. H. and Schul, Y., 1995, " Impact of JOBS intervention on unemployed workers varying in risk for depression," American Journal of Community Psychology 19, 543-562.

[18] Yau, L., Little, R., 2001, "Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed," Journal of the American Statistical Association, 96, 1232-1244.