Scientific Research Publishing

Open Journal of Statistics, 2023, 13, 1-15
https://www.scirp.org/journal/ojs
ISSN Online: 2161-7198
ISSN Print: 2161-718X

Performance of Bayesian Propensity Score Adjustment for Estimating Causal Effects in Small Clinical Trials

Airi Takagi[1], Takuhiro Yamaguchi[2]

[1]Clinical Research Data Center, Tohoku University Hospital, Sendai, Japan
[2]Department of Biostatistics, Tohoku University Graduate School of Medicine, Sendai, Japan
Email: takagi.airi@crietodc.hosp.tohoku.ac.jp

How to cite this paper: Takagi, A. and Yamaguchi, T. (2023) Performance of Bayesian Propensity Score Adjustment for Estimating Causal Effects in Small Clinical Trials. *Open Journal of Statistics*, **13**, 1-15.
https://doi.org/10.4236/ojs.2023.131001

Received: December 26, 2022
Accepted: February 17, 2023
Published: February 20, 2023

Abstract

Propensity score (PS) adjustment can control confounding effects and reduce bias when estimating treatment effects in non-randomized trials or observational studies. PS methods are becoming increasingly used to estimate causal effects, including when the sample size is small compared to the number of confounders. With numerous confounders, quasi-complete separation can easily occur in logistic regression used for estimating the PS, but this has not been addressed. We focused on a Bayesian PS method to address the limitations of quasi-complete separation faced by small trials. Bayesian methods are useful because they estimate the PS and causal effects simultaneously while considering the uncertainty of the PS by modelling it as a latent variable. In this study, we conducted simulations to evaluate the performance of Bayesian simultaneous PS estimation by considering the specification of prior distributions for model comparison. We propose a method to improve predictive performance with discrete outcomes in small trials. We found that the specification of prior distributions assigned to logistic regression coefficients was more important in the second step than in the first step, even when there was a quasi-complete separation in the first step. Assigning Cauchy (0, 2.5) to coefficients improved the predictive performance for estimating causal effects and improving the balancing properties of the confounder.

Keywords

Bayesian Estimation, Causal Inference, Propensity Score, Quasi-Complete Separation, Prior Distribution

1. Introduction

Adjustment by propensity score (PS) is a valuable method for comparing treat-

ment effects in non-randomized trials or observational studies. Rosenbaum and Rubin proposed a PS adjustment method to provide an unbiased estimate of the average treatment effect [1]. PS is the treatment assignment probability, calculated from the measured confounding variables using logistic regression. Subjects with the same PS can be considered comparable or exchangeable. Thus, adjustment with PS provides an unbiased comparison between groups.

The conventional frequentist PS adjustment method involves two sequential steps: estimating PS and estimating causal effects. In the first step, PS is the point-estimated probability of treatment assignment. In the second step, causal effects are estimated using the PS calculated in the first step. However, this conventional two-step sequential method does not consider the impact of uncertainty in the PS on the performance of causal effect estimation [2] [3] [4]. Conventional sequential estimations that use regression adjustment with the PS under the correct specification of the causal model in the second step result in a smaller RMSE than other methods, such as stratification, IPTW, and doubly robust estimation [4]. For comparison with the Bayesian simultaneous method, we use the maximum likelihood estimation of the regression model as our conventional two-step method.

The Bayesian method, in which the first and second steps are estimated simultaneously, has been proposed [5] [6]. There has been also a two-step Bayesian estimation method in which the impact of uncertainty in the PS on the performance of causal effect estimation is evaluated [7]. However, the most useful feature of Bayesian estimation is a simultaneous estimation. The simultaneous Bayesian method accounts for the uncertainty in PS estimation by treating the PS as a latent variable during the estimation of causal effects. When using the maximum likelihood estimation of the regression model, if there are insufficient participants for the number of confounders, quasi-complete separation can easily occur in logistic regression used for estimating the PS. Although the Bayesian simultaneous PS method presents a potential solution for the issue of quasi-complete separation faced by small trials, research is lacking in this area. We focus on the Bayesian method to address these issues and compare the performance of Bayesian simultaneous estimation with that of conventional maximum likelihood estimation. In Bayesian estimation, the priors are important. The utility of priors for providing interpretable shrinkage and conducting causal sensitivity analyses has been discussed [8]. Some prior distributions for Bayesian logistic regression have been proposed as useful methods for dealing with separation in discrete outcomes [9] [10]. Simulations were conducted to evaluate the performance of the Bayesian simultaneous estimation assigned to each prior distribution. Furthermore, in Bayesian simultaneous estimation, the first step is affected by the second step through the simultaneous model fitting of PS and causal effects. The balancing properties of PS are important for unbiased comparisons between the treated and untreated groups. Therefore, we also evaluate covariate balancing properties between treatment and control groups [11] [12] [13] [14].

In Section 2, we describe the conventional frequentist approach and Bayesian simultaneous estimation and outline the methods and prior distributions applied to the data. In Section 3, we present a simulation comparing Bayesian simultaneous and conventional frequentist methods. Data are generated from optional distributions and models in Part 1, and conditions similar to reports of postoperative statin trials comparing the development of acute kidney injury (AKI) between treatment groups in Part 2 [15]. In Section 4, the results of these simulations are reported. In Section 5, we summarize our findings and discuss how the predictive performance of the Bayesian simultaneous method can be improved in small trials.

2. Methodology

2.1. Propensity Score Adjustment

PS is defined as the probability of treatment assignment given measured covariates if the treatment assignment is strongly ignorable and there are no unmeasured confounders. In other words, the PS is estimated under the condition that the treatment assignment cannot be a result of confounding and all confounders have been observed. When estimating PS, a multidimensional vector of measured confounders can be reduced to a one-dimensional vector by using logistic regression. Comparing treated and untreated subjects with the same PS provided an unbiased estimate of the treatment effect. This is similar to the random assignment of the treatment status in clinical trials.

The PS adjustment for a binary outcome is described below with treatment variable $\boldsymbol{X}$, outcome variable $\boldsymbol{Y}$, a vector of p covariates plus intercept $\boldsymbol{C}$ (1, C_1, …, C_p), and regression coefficients vector γ:

$$\operatorname{logit}\left[\Pr\left(\boldsymbol{X}=1\mid\boldsymbol{C}\right)\right]=\gamma\boldsymbol{C} \tag{1}$$

and

$$\operatorname{logit}\left[\Pr\left(\boldsymbol{Y}=1\mid\boldsymbol{X},\boldsymbol{C}\right)\right]=\alpha\boldsymbol{X}+\beta g\left(\boldsymbol{C},\gamma\right). \tag{2}$$

First, the PS is estimated using a logistic regression model that regresses the treatment assignment on observed confounders, as described in Equation (1). Subsequently, the treatment effect is regressed on the treatment assignment and PS, as described in Equation (2).

Equation (1) defines the association between $\boldsymbol{X}$ given $\boldsymbol{C}$, and Equation (2) defines the association between $\boldsymbol{Y}$ and $\boldsymbol{X}$ given $\boldsymbol{C}$. $g(\boldsymbol{C}, \gamma)$ in Equation (2) is the PS. In the simultaneous Bayesian method, posterior estimates of the parameters calculated from Bayesian logistic regression were obtained using the models given in Equations (1) and (2). The PS is modeled as a latent variable and the marginal posterior for the treatment effect is integrated over this variable. In this study, we sample from posterior distributions using the Metropolis-Hastings MCMC algorithm. Updating γ from its conditional posterior distribution corresponds to the simultaneous updating of PS, α, and β in the outcome model.

2.2. Separation and Prior Information

2.2.1. Quasi-Complete Separation

Quasi-complete separation occurs when one or more covariates almost perfectly predict the outcome variable in models for binary outcomes.

Consider logistic regression for PS estimation with the dependent variable X_i and a vector of covariates C_i with coefficient vector γ:

$$\Pr\left(X_i = 1 \mid C_i, \gamma\right) = \frac{1}{1+\exp\left(-C_i\gamma\right)}.$$

The log-likelihood for this model is:

$$\ln L\left(\gamma \mid \boldsymbol{X}\right) = \sum_{i=1}^{N}\left\{X_i \ln\left[\frac{1}{1+\exp\left(-C_i\gamma\right)}\right] + \left(1-X_i\right)\ln\left[1-\frac{1}{1+\exp\left(-C_i\gamma\right)}\right]\right\}.$$

In quasi-complete separation, the maximum likelihood estimate γ is large. Separation is related to little or no overlap in covariate distributions between the treatment groups.

2.2.2. Firth Bias Reduction

A penalized likelihood-based Firth logistic regression method may provide an effective solution to separation [16]. Firth's approach adds a bias term to the log-likelihood function as follows:

$$\ln L\left(\gamma \mid \boldsymbol{X}\right)^{*} = \ln L\left(\gamma \mid \boldsymbol{X}\right) + \frac{1}{2}\ln\left|I\left(\gamma\right)\right|.$$

This penalized likelihood eliminates bias even in the presence of complete separation.

2.2.3. Cauchy Distribution

In Bayesian linear regression, the prior distributions of the regression coefficients are important. Non-informative priors are often applied when there is no prior information. In extreme conditions such as separation, Lasso regression is equivalent to using the Laplace prior in Bayesian interpretation [17]. The Laplace distribution is centered around zero and has long tails. When applying such distributions to the regression coefficients, the posterior estimates shrink towards zero. Cauchy priors have features similar to those of the Laplace prior, whereas the normal prior is more diffuse around zero.

Gelman recommends the Cauchy distribution with a center of 0 and a scale of 2.5 as a default prior for regression coefficients in logistic regression because it allows inferences to be drawn even in the presence of separation [9]. Previous research has found that logistic regression coefficients are typically between −5 and 5. In logistic regression, a change of 5 corresponds to moving a probability from 0.01 to 0.5, or from 0.5 to 0.99 [9]. Here, we used a Cauchy distribution with a center of 0 and a scale of 2.5 (Cauchy (0, 2.5)).

2.2.4. Inclusion of Data Information

There are several approaches to the inclusion of data information in prior dis-

tributions for Bayesian analysis. Sullivan and Greenland calculated prior information, such as the median or variance from confidential intervals [18] [19]. In this study, we use a normal prior distribution that has information arising from true models; the mean is the true parameter value and the variance is 6.25, because the coefficients of logistic regression are almost always between −5 and 5, as described above.

3. Application

Simulations were conducted to evaluate the performance of the Bayesian simultaneous method. The simulation study comprised two parts. In "Part 1", we generated data from arbitrary distributions to confirm the basic performance of each model and prior distribution. In "Part 2", we generated data using a similar condition to that reported in the postoperative statin trial [15]. Each model and prior distribution were applied to the generated data. Simulations were performed using SAS, version 9.4 (SAS Institute, Cary, NC, USA).

3.1. Data Generation

In Part 1, we generated replicated datasets containing $N = 160$ observations for small trials, which led to quasi-complete separation, and $N = 1000$ observations for large trials, which is sufficient to eliminate quasi-complete separation. All the simulated datasets contain eight covariates C_i (C_1, …, C_8), four of which are continuous variables, and the remaining are discrete. Continuous variables were generated from a normal distribution with a mean of 0 and variance of 1, and discrete variables were assumed to follow a Bernoulli distribution with $p = 0.3$. X_i and Y_i are generated from Bernoulli distributions, and the relationship between them is described below.

$$\Pr\left(X_i = 1 \mid C_i, \gamma\right) = \frac{\exp\left(\gamma_0 + \gamma_1 C_{i1} + \cdots + \gamma_8 C_{i8}\right)}{1 + \exp\left(\gamma_0 + \gamma_1 C_{i1} + \cdots + \gamma_8 C_{i8}\right)} \tag{3}$$

and

$$\Pr\left(Y_i = 1 \mid X_i, C_i\right) = \frac{\exp\left(\alpha_0 + \alpha_1 X_i + \beta_1 C_{i1} + \cdots + \beta_8 C_{i8}\right)}{1 + \exp\left(\alpha_0 + \alpha_1 X_i + \beta_1 C_{i1} + \cdots + \beta_8 C_{i8}\right)}. \tag{4}$$

We used logistic regression in both steps and regarded Equations (3) and (4) as the true models used to generate data under the given simulation design. **Table 1** summarizes the characteristics of each simulation and the true values of the model coefficients. The simulation scenario and regression coefficients in Part 1 follow in **Table 1**. We assigned each value to the coefficient γ_n in Equation (3). Each scenario had different coefficient values (α_0, α_1, β_n) in Equation (4). We set $\gamma_8 = -4.4$ as the coefficient of confounder 8 to cause quasi-complete separation within $0 < g(\boldsymbol{C}, \gamma) < 1$ and $0 < \Pr(Y_i = 1 | X_i, C_i) < 1$. The allocation rate was set to 1:1. Under this condition, quasi-complete separation occurred in 16% of replicated datasets. This is not severe and can occur in clinical trials with small sample sizes. Using the conditions for small trials, we generated 8000 datasets

Table 1. Catalog of simulation designs including regression coefficient, outcome event rate per treatment group, number of subjects, and separation rate in replicated datasets. Each row corresponds to the true values of the regression coefficient used to generate the given simulation design. $X = 1$ and $X = 0$ are set for each treatment group.

Scenario	Regression Coefficient of Equation (3)									Regression Coefficient of Equation (4)			Outcome Event rate (%)	
	γ_0	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	α_0	α_1	β_n	$X = 1$	$X = 0$
1	(0.0, 0.5, 0.5, −0.5, −0.5, 0.65, 0.65, 0.65, −4.4)									0	−1	−0.3	23	41
2										0	0	−0.3	43	41
3										0	0	0.3	57	59
4										0	1	0.3	77	59

and sampled 1000 datasets that led to quasi-complete separation (small (100%)) and another 1000 that did not (small (0%)). Using the conditions for large trials, we generated 1000 datasets (large).

In Part 2, data were generated using conditions similar to those reported in a postoperative statin trial comparing AKI development between patients who used postoperative statins and those who did not [15]. In this trial, the sample size was insufficient, given the number of confounding variables, and quasi-complete separation occurred in 13% of the replicated datasets. From 10,000 generated datasets, we sampled 1000 datasets with quasi-complete separation (100%) and another 1000 without (0%). The replicated datasets contained 324 observations and 19 covariates (244 patients taking postoperative statins and 80 patients not taking postoperative statins). The outcome event rate was 21%. The association between outcome events and treatment variables, given confounders, and the distribution of confounders for patients taking postoperative statins and those who developed AKI are described in the original report. From 10,000 generated datasets, we sampled 1000 datasets that led to quasi-complete separation (100%) and another 1000 that did not (0%).

3.2. Prior Distributions and Models Applied to Replicated Datasets

In the Bayesian simultaneous approach, we assigned a normal distribution with a mean of 0 and variance of 10,000, a normal distribution with mean γ_n using information from the true parameter values of regression coefficients and variance of 6.25, and a Cauchy distribution with a center of 0 and a scale of 2.5. Table 2 lists the models applied to the data. We sampled from the posterior density using the Metropolis-Hastings algorithm and used posterior mean estimates obtained from MCMC chains of length 10,000 with the first 2000 discarded as burn-in. For comparison, we used the maximum likelihood estimation of the regression model (ML) and Firth penalized likelihood approach (Firth) as the sequential method.

Table 2. Catalog of simulation designs including the number of subjects, propensity score distributions between each group, separation rate of 1000 replicated datasets, regression coefficients, and the association between treatment and confounders (ML is a maximum likelihood estimation model).

Name of Model		ML	Firth	Normal	Cauchy 1	Cauchy 2	TrueCoef
Propensity score estimation (1st step)	Method	(Logistic regression) maximum likelihood	Firth's penalized likelihood	(Bayesian Logistic regression) Bayes MCMC			
	Prior	-		Normal (0, 10,000)	Cauchy (0, 2.5)	Cauchy (0, 2.5)	Normal (γ_n, 6.25)
Causal effect estimation (2nd step)	Method	(Logistic regression) maximum likelihood		(Bayesian Logistic regression) Bayes MCMC			
	Prior	-		Normal (0, 10,000)	Normal (0, 10,000)	Cauchy (0, 2.5)	Normal (γ_n, 6.25)

3.3. Analysis

We evaluated the numerical performance of estimating the causal effects of 1000 replicated datasets through bias, MSE, and 95% coverage. Causal effects were obtained from estimators of the regression coefficient a_1, which is the log odds ratio of the treatment effect. We referred to 95% confidence intervals or Bayesian credible intervals for calculating coverage.

We checked the balancing properties of the PS by using standardized differences to investigate whether the resulting covariate balance was optimized [11]. We conducted PS adjustment using regression with each confounder and PS as independent variables. We assessed the degree of imbalance in the distributions of each confounder between the treatment groups after PS adjustment using score *d*.

For continuous variables, balancing score *d* is defined as:

$$d = \frac{\left|\bar{C}_{X=1} - \bar{C}_{X=0}\right|}{\sqrt{\frac{SD_{X=1}^2 + SD_{X=0}^2}{2}}} \times 100(\%). \tag{5}$$

where $\bar{C}_{X=1}$ and $\bar{C}_{X=0}$ denote the sample mean of the confounder in each treatment group and $SD_{X=1}^2$ and $SD_{X=0}^2$ denote the sample variance of the confounder in each treatment group. For discrete variables, the score *d* is defined as:

$$d = \frac{\left|\hat{p}_{X=1} - \hat{p}_{X=0}\right|}{\sqrt{\frac{\hat{p}_{X=1}(1-\hat{p}_{X=1}) + \hat{p}_{X=0}(1-\hat{p}_{X=0})}{2}}} \times 100(\%). \tag{6}$$

where $\hat{p}_{X=1}$ and $\hat{p}_{X=0}$ denote the prevalence of the discrete variables in each treatment group.

4. Results

4.1. Impact of Quasi-Complete Separation and Model Comparison

Figure 1 shows the distributions of the causal effects. With a large sample size (N = 1000), the causal effect estimation variability is smaller than that with a small sample size (N = 160). The numerically summarized estimates of α_1 are presented in Table 3. Throughout the analysis, estimates of α_1 under Scenarios 1 and 4 present a large bias and MSE compared with those under Scenarios 2 and 3. Point estimates of α_1 using conventional frequentist estimation had the lowest MSE, but the coverage was slightly higher than 0.95. When using the Bayesian method with small sample sizes, Cauchy 2 shows superior performance relative to the other models, even though TrueCoef uses information from true coefficients. However, shrinking towards zero seems to lead to bias. With a large sample size, the distributions of α_1 estimates were less variable, regardless of the model and method of estimation used. In contrast, with small sample sizes, estimates of α_1 show large distributions and MSE regardless of the presence of quasi-complete separation. Bias depends on each scenario.

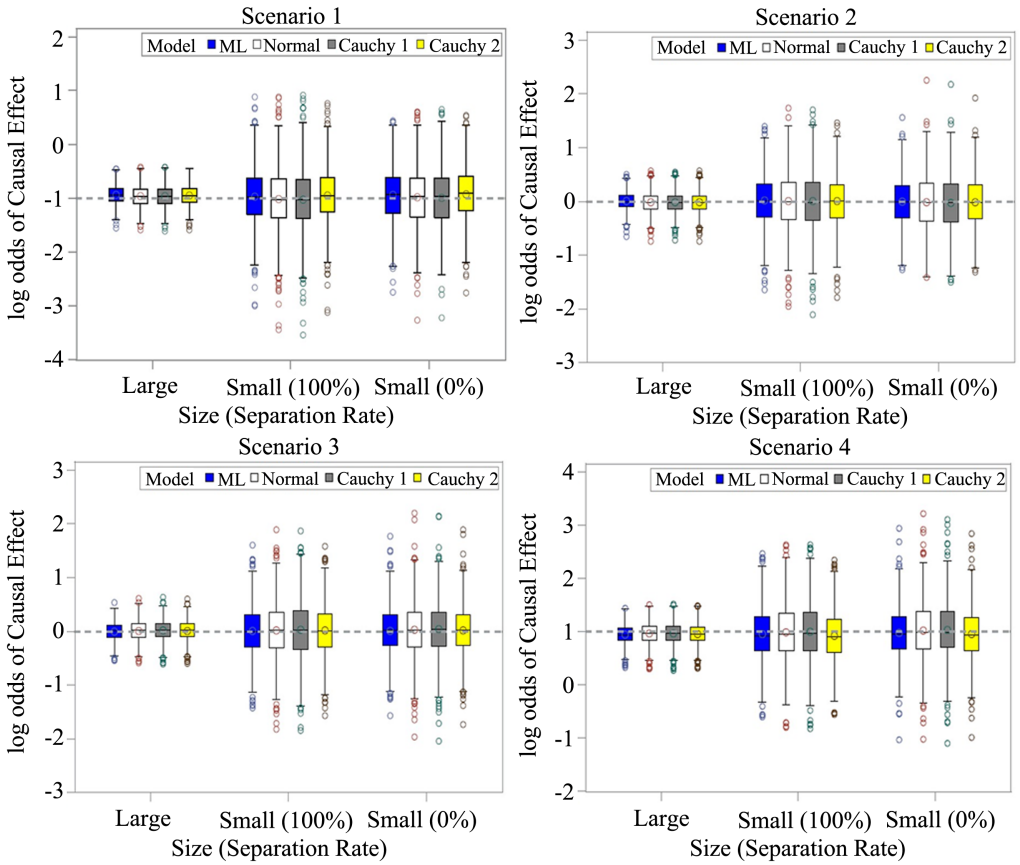

Figure 1. Distributions of estimate of causal effect (α_1) calculated from traditional sequential (ML, Firth) and Bayesian simultaneous method (Normal, Cauchy 2) under Scenario 2 and 3. Small-sized datasets contain 160 observations and Large-sized datasets contain 1000. Small (100%) shows all of replicated datasets lead to separation and Small (0%) shows none of them lead to separation. Horizontal dotted lines are the true parameter values.

Table 3. Numerical performance comparison of estimates of causal effect (α_1) to assess the impact of prior distribution assigned in the first step. Causal effect (α_1) is estimated from conventional sequential (maximum likelihood of regression, ML and Firth) and Bayesian simultaneous methods (Normal, Cauchy 1) with different specifications of prior distributions in the first step and the same specifications in the second step. Small datasets contain 160 observations and large datasets contain 1000 observations. "Small (100%)" shows that all of replicated datasets lead to quasi-complete separation and "Small (0%)" shows that none do.

Performance metrics	Separation	Scenario	N	Name of Model			
				ML	Firth	Normal	Cauchy 1
Bias	Small (100%)	1	160	0.027	0.025	−0.015	−0.026
		2	160	0.021	0.019	0.010	0.001
		3	160	0.003	0.005	0.015	0.023
		4	160	−0.052	−0.051	−0.016	−0.008
	Small (0%)	1	160	0.057	0.057	0.014	0.005
		2	160	−0.008	−0.008	−0.019	−0.027
		3	160	0.010	0.011	0.028	0.034
		4	160	−0.026	−0.026	0.017	0.026
	Large	1	1000	0.067	0.067	0.005	0.007
		2	1000	0.002	0.001	0.011	0.012
		3	1000	0.004	0.004	−0.015	−0.016
		4	1000	0.050	0.050	−0.011	−0.012
MSE	Small (100%)	1	160	0.263	0.262	0.334	0.335
		2	160	0.210	0.209	0.275	0.277
		3	160	0.206	0.206	0.269	0.273
		4	160	0.242	0.239	0.294	0.295
	Small (0%)	1	160	0.244	0.244	0.308	0.311
		2	160	0.196	0.196	0.258	0.262
		3	160	0.191	0.191	0.256	0.259
		4	160	0.233	0.231	0.290	0.289
	Large	1	1000	0.038	0.038	0.049	0.050
		2	1000	0.027	0.027	0.037	0.037
		3	1000	0.026	0.026	0.035	0.035
		4	1000	0.033	0.033	0.050	0.049
95% Coverage	Small (100%)	1	160	0.95	0.95	0.94	0.94
		2	160	0.96	0.96	0.94	0.94
		3	160	0.95	0.95	0.94	0.94
		4	160	0.97	0.97	0.95	0.95

Continued

Small (0%)	1	160	0.94	0.94	0.93	0.93
	2	160	0.95	0.95	0.94	0.93
	3	160	0.96	0.96	0.94	0.94
	4	160	0.96	0.96	0.94	0.95
Large	1	1000	0.93	0.93	0.94	0.93
	2	1000	0.96	0.96	0.92	0.93
	3	1000	0.97	0.96	0.94	0.94
	4	1000	0.96	0.96	0.93	0.93

Table 3 compares the impact of quasi-complete separation in the first step on each model. The performance of α_1 estimation is similar to that of the conventional PS estimation (ML) and Firth penalized method (Firth) despite quasi-complete separation in the first step. In the Bayesian simultaneous method, regardless of the scale parameter or shape of the prior distributions in the first step, there is little impact on the posterior estimates of α_1. The performance of models with the same specification of prior distributions in the second step (Cauchy 1, Normal) was similar, even if they had different specifications in the first step. Table 4 illustrates the performance of the models assigned different specifications for prior distributions in the second step (Cauchy 1, Cauchy 2, TrueCoef). A reduction in the scale parameter of the prior distributions in the second step led to an overall reduction in MSE and an improvement in 95% coverage. Estimates of α_1 are only slightly influenced by the specification of prior distributions in the first step but have a large impact from the second step.

4.2. Assessment of Covariate Balance of Models

Figure 2 shows the boxplot of the score d of confounder 8, which leads to quasi-complete separation, to assess the balancing properties of PS adjustment. Regardless of the prediction model used, PS adjustment significantly reduced the score d. However, the score d from the Bayesian simultaneous methods was larger than that from the conventional frequentist method. Additionally, the score d calculated from the Bayesian simultaneous methods varied by scenario, whereas those from the frequentist methods were constant.

The score d from Cauchy 1 when assigning Cauchy (0, 2.5) in the first step was superior to when using Normal (0, 10,000). Additionally, comparisons between Cauchy 1 and Cauchy 2 appear to show that the reduction in the scale parameter of prior distributions assigned in the second step is associated with an improvement in balancing properties.

4.3. Model Comparison in Postoperative Statin Use Trial

The results of the simulation using the conditions reported in the postoperative statin trial are described in Table 5. The true value of the odd ratio of causal

Table 4. Numerical performance comparison of estimates of causal effect (a_1) to assess the impact of prior distribution assigned in the second step. Causal effect (a_1) is estimated from conventional sequential (maximum likelihood, ML) and Bayesian simultaneous methods (Cauchy 1, Cauchy 2, TrueCoef) with different specifications of prior distributions in the second step and nearly identical specifications in the first step. Small datasets contain 160 observations and large datasets contain 1000 observations. "Small (100%)" shows that all of replicated datasets lead to quasi-complete separation and "Small (0%)" shows that none do.

Performance metrics	Separation	Scenario	N	Name of Model			
				ML	Cauchy 1	Cauchy 2	TrueCoef
Bias	Small (100%)	1	160	0.027	−0.026	0.051	−0.011
		2	160	0.021	0.001	0.008	−0.028
		3	160	0.003	0.023	0.014	−0.041
		4	160	−0.052	−0.008	−0.081	−0.120
	Small (0%)	1	160	0.057	0.005	0.078	0.014
		2	160	−0.008	−0.027	−0.016	−0.056
		3	160	0.010	0.034	0.020	−0.026
		4	160	−0.026	0.026	−0.050	−0.085
	Large	1	1000	0.067	0.007	0.004	0.003
		2	1000	0.002	0.012	0.010	0.008
		3	1000	0.004	−0.016	−0.013	−0.012
		4	1000	0.050	−0.012	−0.009	−0.008
MSE	Small (100%)	1	160	0.263	0.335	0.257	0.273
		2	160	0.210	0.277	0.212	0.229
		3	160	0.206	0.273	0.208	0.228
		4	160	0.242	0.295	0.235	0.254
	Small (0%)	1	160	0.244	0.311	0.244	0.252
		2	160	0.196	0.262	0.202	0.217
		3	160	0.191	0.259	0.196	0.210
		4	160	0.233	0.289	0.228	0.240
	Large	1	1000	0.038	0.050	0.046	0.048
		2	1000	0.027	0.037	0.036	0.036
		3	1000	0.026	0.035	0.034	0.034
		4	1000	0.033	0.049	0.046	0.048
95% Coverage	Small (100%)	1	160	0.95	0.94	0.95	0.95
		2	160	0.96	0.94	0.96	0.95
		3	160	0.95	0.94	0.95	0.95
		4	160	0.97	0.95	0.96	0.95

Continued

Small (0%)	1	160	0.94	0.93	0.94	0.94
	2	160	0.95	0.93	0.95	0.94
	3	160	0.96	0.94	0.95	0.94
	4	160	0.96	0.95	0.95	0.95
Large	1	1000	0.93	0.93	0.94	0.94
	2	1000	0.96	0.93	0.93	0.93
	3	1000	0.97	0.94	0.94	0.94
	4	1000	0.96	0.93	0.94	0.93

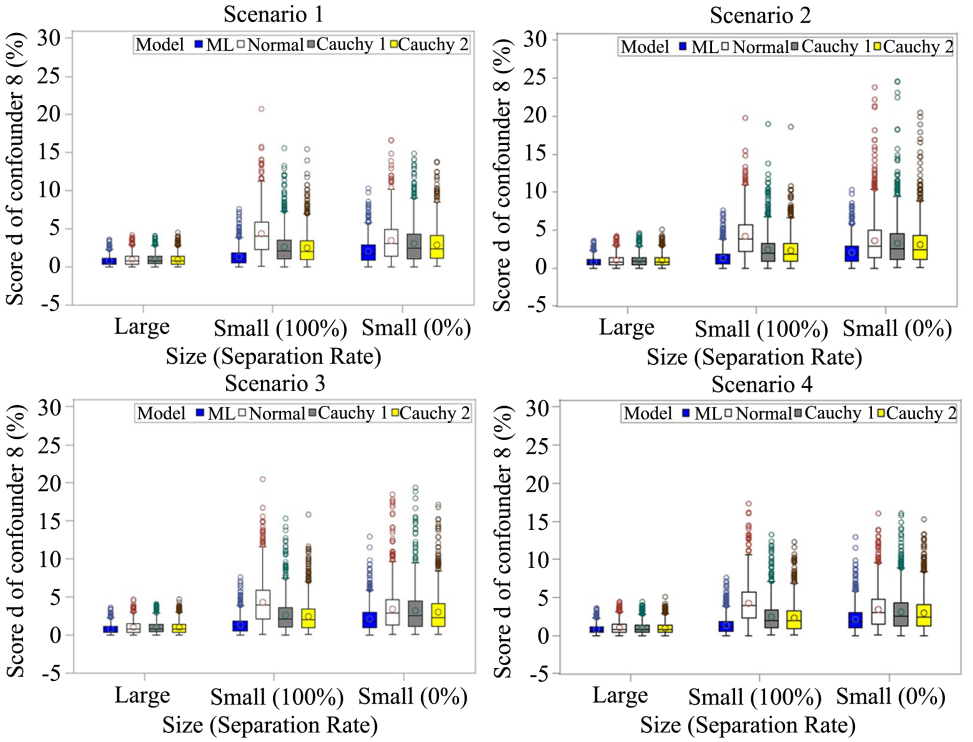

Figure 2. Distributions of balancing score d of confounder 8 which is cause of separation. Each regression coefficient was adjusted PS from traditional sequential (ML, Firth) and Bayesian simultaneous method (Normal, Cauchy 2, TrueCoef) and calculated balancing properties. Small-sized datasets contain 160 observations and Large-sized datasets contain 1000. Small (100%) shows all of replicated datasets lead to quasi-complete separation and Small (0%) shows none do.

Table 5. The consequence of estimates of causal effect (α_1) from conventional sequential (maximum likelihood, ML) and Bayesian simultaneous (Normal, Cauchy 2) methods in a postoperative statin trial setting. Replicated datasets contain 324 observations; "All (100%)" shows all of replicated datasets lead to quasi-complete separation and "None (0%)" shows none of them lead to quasi-complete separation.

Separation	Bias			MSE			95% Coverage		
	ML	Normal	Cauchy 2	ML	Normal	Cauchy 2	ML	Normal	Cauchy 2
All (100%)	0.024	−0.097	0.016	0.231	0.309	0.243	0.96	0.94	0.94
Non (0%)	0.062	−0.052	0.061	0.221	0.280	0.230	0.96	0.95	0.95

effects (0.3) was obtained from the trial report. Quasi-complete separation occurred in 12% of the replicated data. Assigning Cauchy 2 yielded a similar performance to that of conventional ML.

5. Discussion and Conclusions

In this study, we conducted a simulation assuming a small trial with quasi-complete separation caused by one specified confounder and compared the performance of each model.

With a small sample size, despite the presence of quasi-complete separation and using Firth's bias-eliminated estimation in the first step, the performance of the causal effect estimation is only slightly affected. Dimensionality reduction and sequential estimation in frequentist PS methods circumvent the adverse impact of a high-dimensional vector of confounders. Additionally, regression adjustment by PS exhibits superior performance if the outcome model is true [4]. In the Bayesian simultaneous method, the shrinkage estimation of regression coefficients in the first step has little impact on the performance of the causal effect estimation because calculations are fitted to the model of the second step [5] [6]. Hence, prior distributions applied to the coefficients in the second step are important, and the variability of α_1 is influenced by the specification of prior distributions in this step. In contrast, with large sample sizes, the impact of giving prior specifications is small. An insufficient sample size with respect to the number of confounders produced large variability in the PS estimates and led to a remarkable increase in the MSE of α_1. Therefore, an adequate sample size should be used.

Using Cauchy 2 drastically reduces the variance of causal effect estimates compared with other models, but the value of bias depends on the scenario. An adequate scale of prior distributions and shrinkage estimation in the second step seems to reduce the variability, and assigning Cauchy (0, 2.5) may be valuable regarding variance and coverage. However, further research that considers the trade-off between bias and variability is required. Using Cauchy 2 is superior to using TrueCoef because the normal distribution tends to be more diffuse around the mean [10] [17]. The conventional frequentist method (ML) exhibits a small bias and low MSE, but the coverage of α_1 is observed slightly higher than 0.95, without considering residuals.

The balancing property of confounder 8, which led to quasi-complete separation, significantly improved in all models. When using the Bayesian method, Cauchy 2 is superior because the shrinkage estimation of the coefficients caused by Cauchy (0, 2.5) improves the balancing properties. Score d calculated from posterior estimates varies depending on the scenario because simultaneous fitting for the model in the first and second steps provides flexibility in the estimation of γ_n. However, the first step is affected by the second step because it must be adjusted for outcome model fitting such that the posterior distribution of γ_n involves additional information from the second step. It has been argued that

variables included as measured confounders should not depend on the outcome model [12].

In small trials, the variability of causal effect estimation is large, without considering quasi-complete separation; therefore, it is necessary to select adequate prediction models and sample sizes related to the number of confounders. The Bayesian simultaneous PS method, assigning Cauchy (0, 2.5) to coefficients as the prior distribution, is superior in terms of predictive performance for estimating causal effects and improving the balancing properties of the confounder, when leading to quasi-complete separation.

Acknowledgements

I wish to thank Department of Biostatistics of Tohoku University Graduate School of Medicine for their support and encouragement.

Conflicts of Interest

No potential conflict of interest was reported by the authors.

References

[1] Rosenbaum, P.R. and Rubin, D.B. (1983) The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, **70**, 41-55. https://doi.org/10.1093/biomet/70.1.41

[2] Austin, P.C. and Mamdani, M.M. (2005) A Comparison of Propensity Score Methods: A Case-Study Estimating the Effectiveness of Post-AMI Statin Use. *Statistics in Medicine*, **25**, 2084-2106. https://doi.org/10.1002/sim.2328

[3] Gelman, A. and Hill, J. (2007) Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, New York. https://doi.org/10.1017/CBO9780511790942

[4] Lunceford, J.K. and Davidian, M. (2004) Stratification and Weighting via the Propensity Score in Estimation of Causal Treatment Effects: A Comparative Study. *Statistics in Medicine*, **23**, 2937-2960. https://doi.org/10.1002/sim.1903

[5] McCandless, L.C., Gustafson, P. and Austin, P.C. (2009) Bayesian Propensity Score Analysis for Observational Data. *Statistics in Medicine*, **28**, 94-112. https://doi.org/10.1002/sim.3460

[6] Zigler, C.M., *et al.* (2013) Model Feedback in Bayesian Propensity Score Estimation. *Biometrics*, **69**, 263-273. https://doi.org/10.1111/j.1541-0420.2012.01830.x

[7] Liao, S.X. and Zigler, C.M. (2020) Uncertainty in the Design Stage of Two-Stage Bayesian Propensity Score Analysis. *Statistics in Medicine*, **39**, 2265-2290. https://doi.org/10.1002/sim.8486

[8] Oganisian, A. and Roy, J.A. (2021) A Practical Introduction to Bayesian Estimation of Causal Effects: Parametric and Nonparametric Approaches. *Statistics in Medicine*, **40**, 518-551. https://doi.org/10.1002/sim.8761

[9] Gelman, A., Jakulin, A., Pittau, M.G. and Su, Y.S. (2008) A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models. *Annals of Applied Statistics*, **2**, 1360-1383. https://doi.org/10.1214/08-AOAS191

[10] Greenland, S. and Mansournia, M.A. (2015) Penalization, Bias Reduction, and De-

fault Priors in Logistic and Related Categorical and Survival Regressions. *Statistics in Medicine*, **34**, 3133-3143. https://doi.org/10.1002/sim.6537

[11] Austin, P.C. (2009) Balance Diagnostics for Comparing the Distribution of Baseline Covariates between Treatment Groups in Propensity-Score Matched Samples. *Statistics in Medicine*, **28**, 3083-3107. https://doi.org/10.1002/sim.3697

[12] McCandless, L.C., Gustafson, P., Austin, P.C. and Levy, A.R. (2009) Covariate Balance in a Bayesian Propensity Score Analysis of Beta Blocker Therapy in Heart Failure Patients. *Epidemiologic Perspectives & Innovations*, **6**, Article No. 5. https://doi.org/10.1186/1742-5573-6-5

[13] Rosenbaum, P.R. and Rubin, D.B. (1985) Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *Journal of the American Statistical Association*, **39**, 33-38. https://doi.org/10.1080/00031305.1985.10479383

[14] Rosenbaum, P.R. (1989) Optimal Matching for Observational Studies. *Journal of the American Statistical Association*, **84**, 1024-1032. https://doi.org/10.1080/01621459.1989.10478868

[15] Billings 4th., F.T., Pretorius, M., Siew, E.D., Yu, C. and Brown, N.J. (2010) Early Postoperative Statin Therapy Is Associated with a Lower Incidence of Acute Kidney Injury Following Cardiac Surgery. *Journal of Cardiothoracic Vascular Anesthesia*, **24**, 913-920. https://doi.org/10.1053/j.jvca.2010.03.024

[16] Firth, D. (1993) Bias Reduction of Maximum Likelihood Estimate. *Biometrika*, **80,** 27-38. https://doi.org/10.1093/biomet/80.1.27

[17] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society*: *Series B*, **58**, 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[18] Greenland, S. (2006) Bayesian Perspectives for Epidemiological Research: I. Foundations and Basic Methods. *International Journal of Epidemiology*, **35**, 765-775. https://doi.org/10.1093/ije/dyi312

[19] Sullivan, S.G. and Greenland, S. (2013) Bayesian Regression in SAS Software. *International Journal of Epidemiology*, **42**, 308-317. https://doi.org/10.1093/ije/dys213