

## Effect Estimation in Point-exposure Studies with Binary Outcomes and High-dimensional Covariate Data – a Comparison of Targeted Maximum Likelihood Estimation and Inverse Probability of Treatment Weighting

Menglan Pang, MSc<sup>1</sup>, Tibor Schuster, PhD<sup>1,2</sup>, Kristian B. Fillion, PhD<sup>1,2,3</sup>, Mireille E. Schnitzer, PhD<sup>4</sup>, Maria Eberg, MSc<sup>1</sup>, and Robert W. Platt, PhD<sup>2,5,6</sup>

<sup>1</sup>Centre For Clinical Epidemiology, Lady Davis Research Institute, Jewish General Hospital, Montreal, Quebec, Canada

<sup>2</sup>Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada

<sup>3</sup>Division of Clinical Epidemiology, Department of Medicine, McGill University, Montreal, Quebec, Canada

<sup>4</sup>Faculté de pharmacie, Université de Montréal, Quebec, Canada

<sup>5</sup>Department of Pediatrics, McGill University, Montreal, Quebec, Canada

<sup>6</sup>The Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada

### Abstract

Inverse probability of treatment weighting (IPW) and targeted maximum likelihood estimation (TMLE) are relatively new methods proposed for estimating marginal causal effects. TMLE is doubly robust, yielding consistent estimators even under misspecification of either the treatment or the outcome model. While IPW methods are known to be sensitive to near violations of the practical positivity assumption (e.g., in the case of data sparsity), the consequences of this violation in the TMLE framework for binary outcomes have been less widely investigated. As near practical positivity violations are particularly likely in high-dimensional covariate settings, a better understanding of the performance of TMLE is of particular interest for pharmcoepidemiological studies using large databases.

Using plasmode and Monte-Carlo simulation studies, we evaluated the performance of TMLE compared to that of IPW estimators based on a point-exposure cohort study of the marginal causal effect of post-myocardial infarction statin use on the 1-year risk of all-cause mortality from the Clinical Practice Research Datalink. A variety of treatment model specifications were considered, inducing different degrees of near practical non-positivity.

Our simulation study showed that the performance of the TMLE and IPW estimators were comparable when the dimension of the fitted treatment model was small to moderate; however,

they differed when a large number of covariates was considered. When a rich outcome model was included in the TMLE, estimators were unbiased. In some cases, we found **irregular bias and large standard errors with both methods even with a correctly specified high-dimensional treatment model**. The IPW estimator showed a slightly better root MSE with high-dimensional treatment model specifications in our simulation setting.

In conclusion, for estimation of the marginal expectation of the outcome under a fixed treatment, TMLE and IPW estimators employing the same treatment model specification may perform differently due to differential sensitivity to practical positivity violations; however, TMLE, being doubly robust, shows improved performance with richer specifications of the outcome model. Although TMLE is appealing for its double robustness property, such **violations in a high-dimensional covariate setting are problematic for both methods**.

## Keywords

Targeted Maximum Likelihood Estimation; Inverse probability weighting; Non-positivity; Doubly robust estimator; Plasmid simulation

## Introduction

Population-based treatment effects, i.e. the marginal contrast of outcomes from the same population under different counterfactual treatments, are of particular interest in epidemiological studies being conducted for health policy evaluation. Several methods have been developed for estimation of such marginal effects, in particular marginal structural models (MSM) using inverse probability weighting (IPW) and computationally intensive approaches such as G-computation.(1–3) Recently, a class of doubly robust estimators has been developed that beneficially employs efficient influence functions for the parameter of interest.(4). An interesting characteristic of both IPW and the doubly robust methods is their use of the inverse of the propensity score.(4) One such procedure for estimating marginal causal effects, named targeted maximum likelihood estimation (TMLE), was proposed by van der Laan and Rubin.(5–9) For estimation of the mean outcome under a fixed treatment regime, TMLE requires specification of both the treatment model and a conditional model for the outcome. TMLE is doubly robust meaning that only one of these two models has to be correctly specified, in the sense that it must correctly model the outcome or treatment as a function of a sufficient set of covariates, in order to obtain unbiased effect estimates.(8) Furthermore, TMLE has been shown to be locally efficient, meaning that the resulting estimator has minimum large-sample standard error when both models are correctly specified. For these reasons, TMLE is an appealing method to reduce bias due to model misspecification and to possibly increase precision of the effect estimate.

Simulation studies have been used to evaluate the performance of TMLE in certain study designs. For example, simulation has been applied in the context of time-to-event analysis considering different types of censoring.(10,11) Simulation-based investigations have also been conducted in longitudinal data settings where the performance of TMLE was compared to that of other methods in the presence of time-varying confounding.(12–15) Porter(16) conducted simulations in the common setting of a point-exposure study with binary

outcome; we were interested in extending these results to administrative data with a large (e.g., >200) number of covariates, which we will refer to as 'high-dimensional'. We were also interested in determining whether the inclusion of 'pure' treatment predictors that are independent of the outcome in the treatment model would be harmful in this context, and whether the inclusion of 'pure' predictors of the outcome would be beneficial.(17–19)

Pharmacoepidemiologic studies are often conducted using administrative claims and clinical data that are routinely collected for non-research purposes by insurers and governments. Since these data are not research data, control for confounding may require a large number of covariates. In this setting, TMLE has not yet been widely implemented (although software for TMLE has been developed (20,21)). This lack of application may be due to the method's novelty and theoretical complexity. However, before applying the TMLE method in such studies, its performance should be fully assessed. One key assumption that may be particularly problematic for this type of data is the positivity assumption.(22,23) The theoretical positivity assumption requires that the probability of receiving any level of treatment conditional on the covariates must be positive for each individual in the population. Near-, or practical, violations of the positivity assumption can occur when a specific combination of covariates and treatment is rare. Near violations of the positivity assumption and TMLE have also been investigated, in particular for study settings with a continuous outcome(24,25) and TMLE has been implemented in several settings with binary outcomes (9,26,27). However, for high-dimensional administrative data with binary outcomes, the consequences of a near positivity violation and TMLE have been less widely investigated and discussed. Collaborative TMLE (25,28,29) has been proposed to address positivity violations by variable selection for the treatment model; however, we limit this paper to the more widely used standard TMLE.

In this article, we review the conceptual idea and implementation of TMLE in a general point-exposure study with binary exposure and outcome. A comprehensive Monte-Carlo simulation study is conducted to evaluate the performance of TMLE compared to the IPW estimator and to verify its double robustness property. Moreover, a plasmode simulation study (30) incorporating real administrative data is used to assess the performance of TMLE in a high dimensional covariate setting emulating a typical pharmacoepidemiologic study.

## Notation and assumptions

Assume that  $X$  is a (perhaps high-dimensional) vector of covariates and  $A$  and  $Y$  are binary indicator variables for the treatment status and the observed outcome, respectively. Let  $Y_{A=a}$  indicate a patient's potential outcome under treatment  $A=a$ . In order to perform valid causal inference using IPW or TMLE, several necessary assumptions are needed. First, the so-called time ordering assumption is required: the covariates  $X$  precede treatment  $A$  and  $A$  precedes  $Y$  in time, and  $Y$  depends on  $A$  and  $X$  while  $A$  depends only on  $X$ . Taking only pre-treatment variables in  $X$  guarantees that one does not condition on collider variables (common effects of treatment and other covariates) which would potentially induce selection bias.(31) Second, the consistency assumption requires that an individual's potential outcome under the treatment he or she actually received is equal to his or her observed outcome;  $Y_{A=a} = (Y|A=a)$ . Third, the conditional exchangeability assumption (also known as no

unmeasured confounding) has to be made. Confounding occurs when  $Y$  and  $A$  share a common cause. The conditional exchangeability assumption states that the potential outcomes are independent of the observed treatment given measured covariates;  $Y_{A=a} \perp A / X$ , that is, that  $X$  includes sufficient variables to account for common causes (32). Last, as noted earlier, the positivity assumption is also necessary for all treatment levels  $a$  and observed covariate realizations  $x$ . Violations of this assumption can be either *theoretical*, when  $P(A=a | X=x) = 0$ , or *practical*, when  $\hat{P}(A=a | X=x) = 0$  for some observed combination of  $a$  and  $x$ .

## Inverse probability weighting

In point-exposure studies, the IPW estimator can be used to estimate marginal treatment effects in the presence of confounding. Each subject is assigned a weight equal to the inverse of the estimated probability of having received his or her own treatment, i.e. the inverse of the propensity score for a treated subject and the inverse of one minus the propensity score for an untreated subject. By weighting, a pseudo-population is created where the distribution of covariates is comparable across treatment groups.(1) Therefore, contrasts in marginal outcome between the treatment groups in this pseudo-population produce an unbiased estimate of the marginal treatment effect. A weighted unadjusted logistic regression for the outcome will produce an estimate of the marginal treatment effect. Including additional covariates (such as confounders or other risk factors) in the logistic model will produce an estimate of the conditional treatment effect. Due to the non-collapsibility of the odds ratio, these marginal and conditional effects are different quantities.(33)

## Targeted Maximum Likelihood Estimation

In brief, in a point-exposure study with binary exposure and outcome, implementing TMLE for the estimation of the marginal odds ratio consists of three steps.(9,20) To estimate this odds ratio, we simultaneously target  $P(Y_{A=a}=1)$  for  $a=0,1$ . Consistent estimation of the odds ratio requires that both of these parameters be consistently estimated.

Step 1: Initially, two probability models are specified and estimated based on the observed data: i) a model to predict the outcome given the treatment  $A$  and the covariates  $X$  (outcome model) and ii) a model to predict receipt of treatment  $A=a$  conditional on covariates  $X$  (treatment or propensity score model). Standard logistic regression is one possible and commonly used approach for fitting these two models; machine learning tools such as Super Learner may also be used.(34) Step 2: A parametric update model is fit using an intercept free logistic regression including “clever covariates” and an offset equal to the fitted linear prediction from the initial outcome model. The clever covariate is derived based on the efficient influence function for the parameter of interest, and is defined as  $\frac{I(A=a)}{P(A=a | X=x)}$  in the binary case under consideration. This covariate can be estimated from the treatment model in step 1. The coefficient of the clever covariate is referred to as a fluctuation parameter. Consequently, the probabilities of the counterfactual outcomes are estimated with the predictions from the update model by setting everyone’s exposure to  $A=a$ . This step aims to find a better model parameterizations to target an optimal bias-variance tradeoff for the parameters of interest  $P(Y_{A=a}=1)$ . Step 3: The empirical mean of the predicted probabilities

of the counterfactual outcome (computed in the previous step) is taken. The standard error of TMLE can be again estimated with the efficient influence function. The Wald-type 95% confidence interval and the statistics for hypothesis testing can be calculated correspondingly. Once estimates of  $P(Y_{A=i}=1)$  are calculated for both treatment levels, one

can combine them to obtain an estimate of the marginal odds ratio:  $\frac{P(Y_{A=1}=1)/P(Y_{A=1}=0)}{P(Y_{A=0}=1)/P(Y_{A=0}=0)}$ . More detailed description of the theory of TMLE can be found in the literature.(7,9,20)

## Simulation study A

### Simulated data generation

In order to evaluate the performance of TMLE compared to that of the IPW estimator as well as to demonstrate the double robustness of TMLE, different model specifications were employed for both approaches. In all cases, we were interested in estimating the marginal odds ratio. We considered generating four types of binary covariates: a confounding variable ( $C$ ) related to both the treatment and the outcome, an outcome predictor not related to the treatment (baseline risk predictor,  $BR$ ), a treatment predictor not related to the outcome (instrumental variable,  $IV$ ), and a noise variable ( $NV$ ) neither related to the outcome nor to the treatment. These covariates were generated from Bernoulli distributions with probabilities equal 0.5. The treatment ( $A$ ) was generated based on  $C$  and  $IV$ , with sampling

probability  $P(A=1|C, IV) = \frac{\exp(\beta_{0A} + \beta_{CA}C + \beta_{IVA}IV)}{1 + \exp(\beta_{0A} + \beta_{CA}C + \beta_{IVA}IV)}$ . The intercept  $\beta_{0A}$  was chosen to make the prevalence of treatment approximately 0.5. Analogously, we generated the outcome ( $Y$ ) as a function of  $A$ ,  $C$  and  $BR$  with probability

$P(Y=1|A, C, BR) = \frac{\exp(\beta_{0Y} + \beta_{AY}A + \beta_{CY}C + \beta_{BRY}BR)}{1 + \exp(\beta_{0Y} + \beta_{AY}A + \beta_{CY}C + \beta_{BRY}BR)}$ . The intercept  $\beta_{0Y}$  was chosen to obtain outcome prevalence of approximately 0.25. To consider two different scenarios, the coefficients ( $\beta_{CA}$ ,  $\beta_{CY}$ ,  $\beta_{IVA}$ ,  $\beta_{BRY}$ ) for the respective covariates and treatment were set to  $\ln(2)$  and subsequently (in a second scenario) to  $\ln(5)$ . The target of our analysis was the

marginal odds ratio (OR), i.e.,  $\frac{P(Y_{A=1}=1)/P(Y_{A=1}=0)}{P(Y_{A=0}=1)/P(Y_{A=0}=0)}$ . The true values of the marginal odds ratios were 1.94 and 3.80 respectively in our two simulated scenarios, derived from simulations where both counterfactual outcomes were generated for every subject. A total of 100,000 subjects were considered and corresponding parameter settings were used for the two scenarios. The average estimates of the marginal odds ratios were then determined by taking the mean contrasts of the counterfactual outcomes in the logit scale over 1000 simulations.

Sample size was set at  $n \in \{500; 1000; 10,000\}$ . To investigate the impact of omitting different types of variables in either the treatment or the outcome model, we assessed the performance of TMLE in all  $16 \times 16 = 256$  possible combinations of treatment and outcome model specifications, excluding interactions (Table 1). We compared the performance of IPW estimation with that of the TMLE with an unadjusted outcome model for each of the 16 possible treatment model specifications. Stabilized weights for IPW were estimated using weighted logistic regression. Simulation of each data scenario was repeated 1000 times and analyzed accordingly: we reported the empirical bias, standard error, mean squared error of the estimates achieved by the TMLE and IPW approach.

We will further refer to this simulation as “simulation study A”.

### Simulation results

Table 2 show the bias and the root mean squared error (RMSE) for the estimates (ln OR) of the treatment effect for selected TMLE approaches, for the setting with sample size  $n=10,000$  and model coefficients  $\beta_{CA}, \beta_{CY}, \beta_{IV_A}, \beta_{BR_Y}, \beta_{AY} = \ln(5)$ .

With the confounding variable  $C$  included in the outcome models (M2, M6, M7, M8, M12, M13, M14, and M16), TMLE always yielded unbiased treatment effect estimates regardless of the treatment model. When neither the confounder ( $C$ ) nor the instrumental variable ( $IV$ ) was included in the outcome models (M1, M3, M5, and M10), TMLE yielded unbiased estimates as long as the confounder ( $C$ ) was included in the treatment models (T2, T6, T7, T8, T12, M13, T14, and T16). The estimates were biased with the other treatment models. In particular, inclusion of the  $IV$  in models without  $C$  (T4, T9, T11, T15) produced the largest bias. When the outcome models included  $IV$  but not  $C$  (M4, M9, M11 and M15), even when controlling for  $C$  in the treatment model (T2, T6, T8, T13), TMLE produces a small amount of finite-sample bias. Only the joint inclusion of  $C$  and  $IV$  in the treatment model (T7, T12, T14 and T16), i.e. modeling the treatment generating process correctly, produces an unbiased estimate of the effect in this finite sample. In other words, TMLE appeared to take longer to converge when the treatment model was estimated conditional only on confounders rather than the true set of generating variables when the outcome model was overfit to the instrumental variables. The left panel of Figure 3 displays the bias for TMLE with outcome model M4.

The results for RMSE are presented for the TMLE with outcome model M2, M1 and M4 in the corresponding right panel of the figures. The RMSE is minimal for TMLE providing  $C$  is included in the outcome model. Among the TMLEs with the outcome models M1, M3, M5 and M10 (in which both  $C$  and  $IV$  were excluded), estimators with treatment models that include both  $C$  and  $IV$  (T7, T12, T14 and T16) tend to yield slightly larger RMSE compared to those that include  $C$  but not  $IV$ . The RMSE is maximal when the treatment models include  $IV$  but not  $C$  (treatment model T4, T9, T11, T15). When the outcome models include  $IV$  but not  $C$  (M4, M9, M11 and M15), the RMSE is minimal with the joint inclusion of  $C$  and  $IV$  in the treatment model (T7, T12, T14 and T16) within TMLE.

In this simulation study, the IPW estimator had the same bias and RMSE when compared to TMLE with the same propensity score model specification and a marginal outcome model with only treatment included as a covariate (using a marginal, or otherwise limited, outcome model is not recommended, as it eliminates the opportunity for correct specification of the outcome model; however, we include these results for completeness). Moreover, adding the baseline risk ( $BR$ ) variable in either the treatment or the outcome model did not result in obvious improvement for TMLE.

The results are similar for the other parameter setting ( $\beta_{CE}, \beta_{CY}, \beta_{IV_E}, \beta_{BR_Y}, \beta_{EY} = \ln(2)$ ,  $n=\{500; 1000\}$ ), although the contrasts between different models were less distinct for the smaller coefficients (results not shown).



## Plasmode simulation study

### Data sources

To further evaluate the performance of TMLE and IPW in high dimensional covariate settings with respect to estimation bias and precision, simulation studies were conducted to mimic a typical pharmacoepidemiologic study with known data generating process. Given the challenge of generating realistic high dimensional data, we used a plasmode simulation study. (30) A plasmode study starts with an existing cohort, so that associations between covariates reflect real-world patients. We then injected known signals (effects) into it. The true simulated effect was 0.72, a value based on a real word study of the effect of the statin use post-myocardial infarction (MI) on one-year risk of all-cause mortality (35). We will further refer to this simulation as “simulation study B”.

This is a retrospective population-based cohort study using the data from the Clinical Practice Research Datalink (CPRD) and Hospital Episode Statistics (HES). 32,792 patients aged 18 and older, and diagnosed as MI were drawn from the databases between April 1<sup>st</sup>, 1998 and March 31<sup>st</sup>, 2012. This cohort consists of 19,122 patients treated with statin and 13,671 patients not treated with statin within 30 days after the diagnosis of MI. All-cause mortality was evaluated as any death recorded in the databases during the one year follow-up period. A range of known potential confounders can be predefined from the study, including demographic characteristics (e.g. age, sex), time variables (e.g. year of cohort entry), clinical characteristics (e.g., smoking, alcohol use, obesity), comorbidities (e.g. diabetes mellitus, atrial fibrillation, coronary artery disease (recorded >30 days before the index MI), acute coronary syndrome, cerebrovascular disease, congestive heart failure, chronic obstructive pulmonary disease, hypertension, hypercholesterolemia, peripheral vascular disease, previous coronary revascularization, previous stroke, previous MI (recorded >30 days before the index MI), and previous medications prescribed (e.g., aspirin, angiotensin-converting enzyme (ACE) inhibitors, angiotensin receptor blockers (ARBs), beta-blockers, calcium-channel blockers, diuretics, fibrates, non-steroidal anti-inflammatory drugs (NSAIDs)). We also constructed variables for the number of prescriptions issued and the number of hospitalizations in the previous year, which are two proxies for overall health. Age, number of hospitalizations, and prescription count were categorized into groups and they were considered as dummy variables along with year of cohort entry. Additionally, 400 empirical covariates were identified as proxies for unmeasured confounding via the high-dimensional propensity score (HDPS) algorithm.(36–41) This study (protocol number: 14\_018) was approved by the Independent Scientific Advisory Committee of the CPRD and the Research Ethics Board of the Jewish General Hospital (Montreal, Quebec).

### Exposure and outcome generation

The covariates from the real data example were directly applied in the simulations and the coefficients estimated from the real data were used as our parameter setting. The treatment and outcome were stochastically generated based on the estimated coefficients as retrieved from the real study data analysis. We let all the measured confounding covariates and the 400 empirical HDPS-selected variables be the complete set of confounding covariates (denoted as *C*). The treatment (denoted as *A*) was generated, similarly to the first simulation

study, conditional on  $C$ . The parameter values for the intercept ( $\beta_{0A}$ ) and the covariates  $\beta_{CA}$  were set as the corresponding estimated coefficients from the real data fitting a binary logistic propensity score model with response variable  $A$  and explanatory variables  $C$ . Therefore, the treatment variable ( $A$ ) was generated from a Bernoulli distribution with

$$P(A=1|C) = \frac{\exp(\beta_{0A} + \beta'_{CA}C)}{1 + \exp(\beta_{0A} + \beta'_{CA}C)}.$$
 Likewise, in the outcome model, the intercept  $\beta_{0Y}$ , the coefficient of  $A$  (the conditional log odds ratio), and the coefficient of  $C$  ( $\beta_{CY}$ ) were fixed at the corresponding estimates of the coefficients from the real study. Finally, we generated the outcome variable based on the previously generated  $A$  and the covariates  $C$ , with the

probability 
$$P(Y=1|A, C) = \frac{\exp(\beta_{0Y} + \beta_{AY}A + \beta'_{CY}C)}{1 + \exp(\beta_{0Y} + \beta_{AY}A + \beta'_{CY}C)}.$$
 The true marginal effect (OR) was derived from a contrast of the two marginal potential outcome probabilities, which were computed by using the true values of the parameters  $\beta_{0Y}$ ,  $\beta_{AY}$  and  $\beta_{CY}$  and simulating exposed and unexposed counterfactual outcomes for all subjects in the population. A varying number of covariates were defined by four nested covariate sets (Table 3) inducing different levels of confounding adjustment as well as different degrees of non-positivity. The null covariate set included no variables. The simple covariate set ( $C_1$ ) included several predefined important confounders, including age, sex, obesity, smoking and history of diabetes. The intermediate covariate set ( $C_2$ ) included  $C_1$  and a variety of additional potential confounders defined in the last section. Finally, the full covariate set ( $C_3$ ) included all pre-specified potential confounders ( $C_2$ ) as well as 400 empirical variables selected by the HDPS algorithm. Sixteen TMLE and four IPW analyses were performed in this simulation, considering different model specifications (Table 3) for the treatment model in both approaches and also for the outcome model in the TMLE. In addition, to assess the methods under a more practical strategy with respect to extreme propensity score, all the analyses were repeated with truncation of the propensity score at 0.025 and 0.975. We set the sample size to the number of subjects in the real study ( $n=32,792$ ). Due to computational intensity, the simulation and the analyses were limited to 500 generated datasets.

## Simulation results

The true marginal effect (OR) of the treatment in this simulation study is equal to  $-0.333$ . Based on the 443 simulation runs where all the TMLE and IPW estimators converged (out of a total of 500), the bias, the empirical standard error and the root mean squared error (RMSE) for the estimates of the marginal effects without truncation of the propensity score are presented in Table 4a.

In the treatment generating process, 445 coefficients were directly applied from the real data example. The intercept was set to  $-4.55$  and most of the coefficients were very small. The maximum coefficient was 2.1, and the minimum and maximum of the probability of treatment were 0.0006 and 0.9998 respectively. Therefore, full positivity violations did not occur. However, in the analyses of the simulated data, extreme estimated propensity scores were present when all the confounders ( $C_3$ ) were included in the treatment model. Over the 500 simulation runs, 5.8% of propensity scores were smaller than 0.025, while 14.1% of propensity score were larger than 0.975. Near violations of positivity often occurred in the



IPW estimator and TMLE when treatment models included the full covariate set. The distribution of the estimated propensity score and weights from all the simulations is presented in tables 5 and 6.

The distributions of the estimates from the 20 models without truncation are presented by the boxplots in Figure 1a. In this simulation, the results revealed the property of double robustness of TMLE. There was minimal bias if all the confounders ( $C_3$ ) were included in either the treatment model or the outcome model. With unconditional outcome models (IPW1 to IPW4, TMLE1 to TMLE4), the IPW estimator and TMLE provided equal estimates and precision for non-adjustment and when adjusting for the simple covariate set ( $C_1$ ). These estimates were largely biased because many important confounders were omitted. After adjusting for the intermediate covariate set ( $C_2$ ), the effect estimated by IPW estimator had slightly smaller bias and standard error compared to the TMLE. These estimates were still biased since only the pre-specified confounders were included, although they were moved towards to the true effect. Adjusting for all present confounders (covariate set  $C_3$ ), the TMLE yielded more bias and much larger RMSE compared to the IPW estimator. This result indicated that including a larger number of covariates in the TMLE treatment model, inducing a near violation of positivity, may lead to biased estimates and poor precision. Moreover, numerical problems (infinite fitted values) were observed with TMLE for 14 out of 500 simulation runs, perhaps due to overfitting of the treatment model. The plasmode simulation results show that with an unadjusted outcome model TMLE may perform worse than IPW when adjusting for a sufficient but high-dimensional set of covariates in the treatment model. Adjusting for predefined important confounders (the simple covariate set  $C_1$ ) in the outcome model, approach TMLE5 with an unadjusted treatment model provided a similar estimate compared to TMLE2. Adjusting for all pre-specified confounders (the intermediate covariate set  $C_2$ ) in the outcome model, the TMLE9 with an intercept only treatment model and the TMLE10 (treatment model including  $C_1$ ) provided similar point estimates. Compared to TMLE3 (which only adjusted for  $C_2$  in the treatment model), the standard errors were 50% smaller. Adjusting for all confounders (the full covariate set  $C_3$ ) in the outcome model, TMLE13 to TMLE15 provided considerably reduced bias and standard error compared to TMLE4, where  $C_3$  was only adjusted for in the treatment model. This suggested that it might be more efficient to adjust for confounders in the outcome model rather than the treatment model when there is a near non-positivity issue. We found again that adjusting for a large number of covariates ( $C_3$ ) in the treatment model was associated with larger standard errors and may be the result of near non-positivity. Noticeably, extreme propensity scores caused numerical convergence problems for TMLE. Numerical problems were experienced 2.8%, 10.6%, 3.2% and 1.0% of the time respectively for TMLE4, TMLE8, TMLE12 and TMLE16 among the 500 simulation runs (Table 4a). Among all the TMLEs that included  $C_3$  in the outcome model, TMLE14 (including  $C_1$  in the treatment model) provided the estimates with the lowest bias and the highest precision.

Results with truncation of the propensity score are presented in Table 4b and the corresponding boxplots for the distributions of the estimates are presented in Figure 1b. Compared to the analyses without truncation, slightly more bias but better precision for IPW estimators that adjust for a large number of covariates (IPW3 and IPW4) was attained by

truncation. On the other hand, the standard error as well as the bias were both largely reduced for TMLE that includes the full covariate set  $C_3$  in the treatment model. Moreover, as was expected, all TMLE models with truncation of the propensity score avoided the numerical problems previously observed in the non-truncated case. In this simulation with truncation, TMLE with a marginal outcome model again yielded slightly larger bias and standard error compared to the IPW estimator when high dimensional covariates were considered.

## Discussion

TMLE produces semi-parametric efficient estimators of marginal causal effects in a population. In this paper we have focused on the application of TMLE and its comparison to the IPW estimator in both low and high dimensional covariate settings. We used parametric models for the initial estimation of the probabilities of the outcome and treatment respectively. This was done to simplify the fitting process in a high-dimensional setting, but alternative (non-parametric) methods are also possible. For example, cross-validation based machine learning approaches, such as the Deletion/Substitution/Addition (DSA) algorithm, can search over a large space of polynomial generalized linear models;(42,43) Alternatively, the Super Learner can be used to combine predictions over a set of candidate algorithms. (34,44,45)

Our comprehensive simulation study investigated several combinations of the treatment and outcome model specifications conditional on the baseline variables. We demonstrated the double robustness property of TMLE in basic finite sample settings. Double robustness enables consistent treatment effect estimation when either the propensity score or the outcome model is correctly specified on the full set of confounders. The IPW estimator and TMLE using an unadjusted outcome model performed equivalently in the absence of a near violation of positivity when using the same propensity score model. This result is perhaps not surprising, given that the two methods use the same information (the treatment model). When TMLE included a rich outcome model, bias was reduced.

Indeed, the positivity assumption was satisfied in simulation study A. The estimated propensity scores were always bounded between 0.07 and 0.93 in all the simulations. However, both methods experienced instability in a high dimensional setting when a large number of covariates were included in the propensity score model. Near practical violations of the positivity assumption caused problems when the propensity score model was over-adjusted. However, diagnostics for the propensity score should be performed in high dimensional settings. When near positivity violations are detected, TMLE's performance can be improved by truncation of the propensity score or by including the high dimensional covariates in the outcome model while keeping only a few covariates in the treatment model.

In the plasmode simulation studies (simulation study B), we showed that as the result of a near violation of the positivity assumption, extreme propensity scores, although correctly modeled, can lead to poor precision and biased estimates for both IPW and TMLE. They can even produce numerical instability for TMLE (specifically, prediction of infinite values). Positivity is an essential assumption and should be always verified in any analysis.

Examination of the distributions of the propensity score and the weights is therefore recommended for both IPW and TMLE. The bias induced by such violations can be further assessed by using parametric bootstrap and simulation.(46) Further TMLE and the IPW estimator perform differently in this setting even with the same modeling approach due to their different sensitivity to practical positivity violation. Although inverse weighting by the propensity score is used in both approaches, the involved estimating process is different. Within the IPW estimation approach, the propensity score is used to define the weights for creation of the pseudo-population. Large weights, however, can result if an extreme predicted probability (close to 0 or 1) does not correspond with the actual observed treatment status of an individual. Such inflated weights can have high impact on the analysis results thus leading to poor precision and/or biased effect estimates. In the TMLE procedure, the same inverse weights are embedded in the clever covariate, which enables estimation of parameters in the updating procedure, allowing for the subsequent computation of individual counterfactual outcome prediction. However, since the estimation equation is evaluated for each subject under both treatment and no treatment conditions, any extreme propensity score (close to 0 or 1) inflates the value of the clever covariates, and consequently causes unstable parameter estimation. Moreover, an extreme propensity score also has influence in estimating the parameter in the logistic parametric update model, consequently numerical non-convergence can occur due to a very unstable fluctuation parameter estimate.

Truncation of the propensity score may prevent poor precision with the price of minor bias, (47) and data-adaptive approaches have been proposed to select the level of truncation level minimizes the expected mean square error.(48,49) There is an option in the TMLE function in R to specify an amount of truncation to avoid extreme estimated probabilities of receiving treatment.(20,21) We provided results from additional analyses in which the propensity score was truncated at 0.025 and 0.975. The results indicate that the influence of the extreme propensity score can be controlled by truncation. Therefore, it should be carried out in practice especially for a high dimensional propensity score model. In our truncation analysis, IPW and TMLE (with a marginal outcome model) again performed differently. This again reflect the difference in the use of propensity score in these two estimating procedures. Collaborative-TMLE (28) is an extension of TMLE that allows for the data-adaptive selection of covariates into the propensity score, potentially avoiding the non-convergence and variance inflation that we encountered. This is another potential solution for high-dimensional data analysis with near-positivity violations (8).

Generalizability of the results and conclusions is a potential limitation in our plasmode simulation study. We did not assign a broad range of covariate distributions or parameter values. The covariates were instead derived directly from a real-world setting. The parameter values and the sample size are derived from the real-world study and then fixed for the simulation. Therefore, our simulation results are not necessarily representative of other settings and we cannot draw definitive conclusions about the comparison of the performance of TMLE and IPW estimators. However, this specific data setting is typical in a pharmacoepidemiologic study and has already provided useful insights and evaluation regarding the performance of TMLE in the high dimensional scenario.

In conclusion, TMLE and IPW estimators both use inverse propensity score weighting, but can perform differently. Both TMLE and IPW can be sensitive to violations of the positivity assumption; near-violations of this assumption are possible in high-dimensional covariate settings, and inference should be interpreted with caution. In such settings, TMLE with a high-dimensional outcome model and a reduced treatment model may be a better alternative. Collaborative TMLE may be a useful alternative in these settings; further exploration of these methods is warranted.

## Acknowledgments

This work was supported by the Canadian Network for Observational Drug Effect Studies (CNODES). CNODES, a collaborating centre of the Drug Safety and Effectiveness Network (DSEN), is funded by the Canadian Institutes of Health Research (CIHR). MP holds a studentship from the Fonds de Recherche du Québec – Santé (FQR-S). KBF and MES hold New Investigator awards from CIHR. RWP is a Chercheur-national (National Scholar) of the FQR-S and holds the Albert Boehringer I Chair in Pharmacoepidemiology

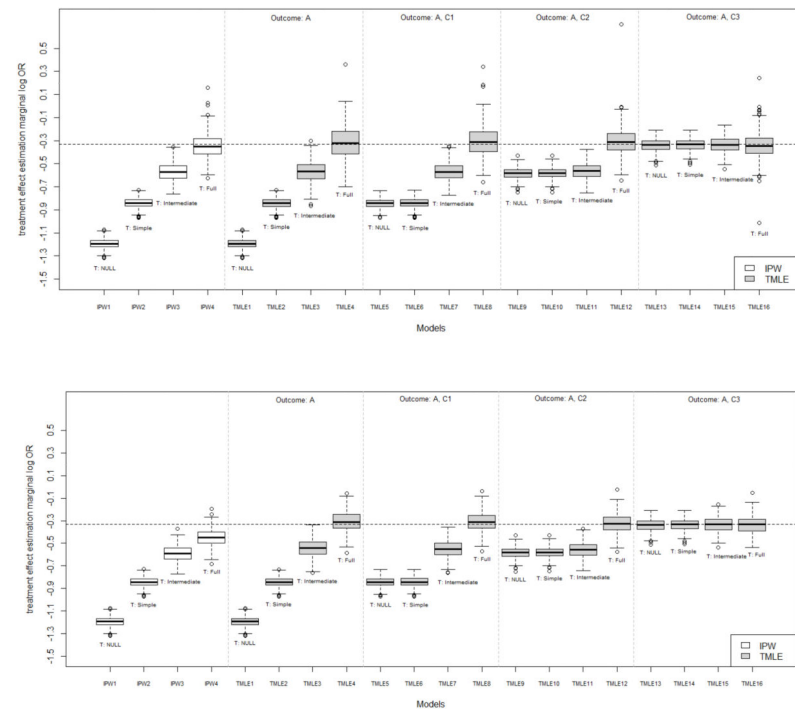
## References

1. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000; 11(5):550–60. [PubMed: 10955408]
2. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period —application to control of the healthy worker survivor effect. *Math Model*. 1986; 7(9):1393–512.
3. Snowden JM, Rose S, Mortimer KM. Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. *Am J Epidemiol*. 2011 Apr 1; 173(7):731–8. [PubMed: 21415029]
4. Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J Am Stat Assoc*. 1999; 94(448):1096–120.
5. Van der Laan MJ. Targeted maximum likelihood based causal inference: Part I. *Int J Biostat*. 2010; 6(2)doi: 10.2202/1557-4679.1211
6. Van der Laan MJ. Targeted maximum likelihood based causal inference: Part II. *Int J Biostat*. 2010; 6(2)doi: 10.2202/1557-4679.1241
7. Rosenblum M, van der Laan MJ. Targeted maximum likelihood estimation of the parameter of a marginal structural model. *Int J Biostat*. 2010; 6(2)doi: 10.2202/1557-4679.1238
8. Schnitzer ME, Lok JJ, Gruber S. Variable Selection for Confounder Control, Flexible Modeling and Collaborative Targeted Minimum Loss-Based Estimation in Causal Inference. *Int J Biostat*. 2016; 12(1):97–115. [PubMed: 26226129]
9. Moore KL, van der Laan MJ. Covariate adjustment in randomized trials with binary outcomes: Targeted maximum likelihood estimation. *Stat Med*. 2009; 28(1):39–64. [PubMed: 18985634]
10. Stitelman OM, De Gruttola V, van der Laan MJ. A general implementation of tmle for longitudinal data applied to causal inference in survival analysis. *Int J Biostat*. 2010; 8(1)doi: 10.1515/1557-4679.1334
11. Moore, KL., van der Laan, MJ. UC Berkeley Division of Biostatistics Working Paper Series. 2009. Application of time-to-event methods in the assessment of safety in clinical trials. Working Paper 248
12. Schnitzer ME, Moodie EE, Platt RW. Targeted maximum likelihood estimation for marginal time-dependent treatment effects under density misspecification. *Biostatistics*. 2013; 14(1):1–14. [PubMed: 22797173]
13. Van der Laan MJ, Gruber S. Targeted Minimum Loss Based Estimation of Causal Effects of Multiple Time Point Interventions. *Int J Biostat*. 2012; 8(1)doi: 10.1515/1557-4679.1370
14. Petersen M, Schwab J, Gruber S, Blaser N, Schomaker M, van der Laan M. Targeted Maximum Likelihood Estimation for Dynamic and Static Longitudinal Marginal Structural Working Models. *J Causal Inference*. 2014 Sep 1; 2(2):147–85. [PubMed: 25909047]

15. Schnitzer ME, van der Laan MJ, Moodie EEM, Platt RW. Effect of breastfeeding on gastrointestinal infection in infants: A targeted maximum likelihood approach for clustered longitudinal data. *Ann Appl Stat.* 2014 Jun; 8(2):703–25. [PubMed: 25505499]
16. Porter, KE. The Relative Performance of Targeted Maximum Likelihood Estimators Under Violations of the Positivity Assumption. 2011. <http://escholarship.org/uc/item/3hp4r33n.pdf>
17. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable Selection for Propensity Score Models. *Am J Epidemiol.* 2006 Jun 15; 163(12):1149–56. [PubMed: 16624967]
18. Austin PC, Mamdani MM. A comparison of propensity score methods: a case-study estimating the effectiveness of post-AMI statin use. *Stat Med.* 2006 Jun 30; 25(12):2084–106. [PubMed: 16220490]
19. Lefebvre G, Delaney JAC, Platt RW. Impact of mis-specification of the treatment model on estimates from a marginal structural model. *Stat Med.* 2008 Aug 15; 27(18):3629–42. [PubMed: 18254127]
20. Gruber S, van der Laan MJ. tmle: An R Package for Targeted Maximum Likelihood Estimation. *Journal of Statistical Software.* 2011; 51(13):1–35.
21. Schwab, JL., Petersen, M., van der, L. LTMLE: longitudinal targeted maximum likelihood estimation, 2013. 2014. <http://cran.r-project.org/web/packages/ltmle/index.html>
22. Neugebauer R, van der Laan M. Why prefer double robust estimators in causal inference? *J Stat Plan Inference.* 2005; 129(1):405–26.
23. Ertefaie A, Stephens DA. Comparing approaches to causal inference for longitudinal data: inverse probability weighting versus propensity scores. *Int J Biostat.* 2010; 6(2)doi: 10.2202/1557-4679.1198
24. Gruber S, van der Laan MJ. A Targeted Maximum Likelihood Estimator of a Causal Effect on a Bounded Continuous Outcome. *Int J Biostat.* 2010; 6(1)doi: 10.2202/1557-4679.1260
25. Porter KE, Gruber S, van der Laan MJ, Sekhon JS. The relative performance of targeted maximum likelihood estimators. *Int J Biostat.* 2011; 7(1):1–34.
26. Lendle SD, Fireman B, van der Laan MJ. Targeted maximum likelihood estimation in safety analysis. *J Clin Epidemiol.* 2013 Aug 1; 66(8):S91–8. [PubMed: 23849159]
27. Brown DM, Petersen M, Costello S. Occupational Exposure to PM2.5 and Incidence of Ischemic Heart Disease: Longitudinal Targeted Minimum Loss-based Estimation. *Epidemiology.* 2015; 26(6):806–14. [PubMed: 26079662]
28. Van der Laan MJ, Gruber S. Collaborative Double Robust Targeted Maximum Likelihood Estimation. *Int J Biostat.* 2010; 6(1)doi: 10.2202/1557-4679.1181
29. Gruber S, van der Laan MJ. An Application of Collaborative Targeted Maximum Likelihood Estimation in Causal Inference and Genomics. *Int J Biostat.* 2010; 6(1)doi: 10.2202/1557-4679.1182
30. Franklin JM, Schneeweiss S, Polinski JM, Rassen JA. Plasmode simulation for the evaluation of pharmacoepidemiologic methods in complex healthcare databases. *Comput Stat Data Anal.* 2014 Apr; 72:219–26. [PubMed: 24587587]
31. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology.* 1999; 10(1):37–48. [PubMed: 9888278]
32. Hernán, MA., Robins, JM. *Causal Inference.* Boca Raton: Chapman & Hall/CRC; 2016. forthcoming
33. Pang M, Kaufman JS, Platt RW. Studying noncollapsibility of the odds ratio with marginal structural and logistic regression models. *Stat Methods Med Res.* 2013 0962280213505804.
34. Van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol.* 2007; 6(1)doi: 10.2202/1544-6115.1309
35. Pang M, Schuster T, Filion KB, Eberg M, Platt RW. Targeted Maximum Likelihood Estimation for pharmacoepidemiological research: A Real-World Example Using the Clinical Practice Research Datalink. 2015 In Press.
36. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiol Camb Mass.* 2009; 20(4):512.

37. Stürmer T, Schneeweiss S, Brookhart MA, Rothman KJ, Avorn J, Glynn RJ. Analytic strategies to adjust confounding using exposure propensity scores and disease risk scores: nonsteroidal antiinflammatory drugs and short-term mortality in the elderly. *Am J Epidemiol.* 2005; 161(9): 891–8. [PubMed: 15840622]
38. Rassen JA, Schneeweiss S. Using high-dimensional propensity scores to automate confounding control in a distributed medical product safety surveillance system. *Pharmacoepidemiol Drug Saf.* 2012; 21(S1):41–9. [PubMed: 22262592]
39. Rassen JA, Glynn RJ, Brookhart MA, Schneeweiss S. Covariate selection in high-dimensional propensity score analyses of treatment effects in small samples. *Am J Epidemiol.* 2011; 173(12): 1404–13. [PubMed: 21602301]
40. Rassen JA, Avorn J, Schneeweiss S. Multivariate-adjusted pharmacoepidemiologic analyses of confidential information pooled from multiple health care utilization databases. *Pharmacoepidemiol Drug Saf.* 2010; 19(8):848–57. [PubMed: 20162632]
41. Rassen, JA., Doherty, M., Huang, W., Schneeweiss, S. *Pharmacoepidemiology toolbox.* Boston MA: <http://www.hdpharmacoepi.org>
42. Sinisi, SE., Van Der Laan, MJ. Loss-based cross-validated deletion/substitution/addition algorithms in estimation. 2004. Available from: <http://biostats.bepress.com/ucbbiostat/paper103/>
43. Sinisi SE, van der Laan MJ. Deletion/substitution/addition algorithm in learning with applications in genomics. *Stat Appl Genet Mol Biol.* 2004; 3(1):1069.
44. Sinisi SE, Polley EC, Petersen ML, Rhee S-Y, van der Laan MJ. Super learning: an application to the prediction of HIV-1 drug resistance. *Stat Appl Genet Mol Biol.* 2007; 6(1):7.
45. Zheng, W., van der Laan, M. Asymptotic Theory for Cross-validated Targeted Maximum Likelihood Estimation. 2010. Available at: <http://works.bepress.com/wenjing-zheng/22/>
46. Petersen ML, Porter KE, Gruber S, Wang Y, van der Laan MJ. Diagnosing and responding to violations in the positivity assumption. *Stat Methods Med Res.* 2010 0962280210386207.
47. Cole SR, Hernán MA. Constructing Inverse Probability Weights for Marginal Structural Models. *Am J Epidemiol.* 2008 Sep 15; 168(6):656–64. [PubMed: 18682488]
48. Xiao Y, Moodie EEM, Abrahamowicz M. Comparison of Approaches to Weight Truncation for Marginal Structural Cox Models. *Epidemiol Methods.* 2013 Jan 8; 2(1):1–20.
49. Bembom, O., van der Laan, M. Data-adaptive selection of the truncation level for Inverse-Probability-of-Treatment-Weighted estimators. UC Berkeley Division of Biostatistics Working Paper Series. 2008. Working Paper 230. Available at: <http://biostats.bepress.com/ucbbiostat/paper230>





**Figure 1a–b.**

Distributions of the estimates from IPW estimator and TMLE in the plasmode simulation study. Figure 1a is for models with non-truncated weights; figure 1b uses weights truncated at the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles.

**Table 1**

Model specifications in simulation study A

	<b>Treatment model</b>
T1	<i>I (intercept only model)</i>
T2	<i>C</i>
T3	<i>BR</i>
T4	<i>IV</i>
T5	<i>NV</i>
T6	<i>C+BR</i>
T7	<i>C+IV</i>
T8	<i>C+NV</i>
T9	<i>BR+IV</i>
T10	<i>BR+NV</i>
T11	<i>IV+NV</i>
T12	<i>C+BR+IV</i>
T13	<i>C+BR+NV</i>
T14	<i>C+IV+NV</i>
T15	<i>BR+IV+NV</i>
	<i>C+BR+IV+NV</i>

T16

\* Outcome models (M1–M16) in TMLE included the treatment and the same 16 different covariate combinations

**Table 2**

Bias and RMSE of treatment effect estimates using various outcome models and treatment models with sample size  $n=10,000$ ,  $\beta_{CE}$ ,  $\beta_{CY}$ ,  $\beta_{IV_E}$ ,  $\beta_{BR_Y}$ ,  $\beta_{EY} = \ln(5)$ .

Treatment Model	Outcome Model 1				Outcome Model 2		Outcome Model 4	
	IPW		TMLE		TMLE		TMLE	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
I	4.64	4.67	4.64	4.67	0.05	0.57	5.21	5.24
X	0.01	0.56	0.01	0.56	0.01	0.56	0.63	0.86
BR	4.64	4.67	4.64	4.67	0.05	0.55	5.21	5.24
IV	5.25	5.28	5.25	5.28	0.06	0.6	5.25	5.28
NV	4.64	4.67	4.64	4.67	0.05	0.57	5.21	5.24
X+BR	0.01	0.54	0.01	0.54	0.01	0.54	0.63	0.85
X+IV	0.01	0.64	0.01	0.64	0.01	0.64	0.01	0.64
X+NV	0.01	0.56	0.01	0.56	0.01	0.56	0.63	0.86
BR+IV	5.25	5.28	5.25	5.28	0.06	0.58	5.25	5.28
BR+NV	4.64	4.67	4.64	4.67	0.05	0.55	5.21	5.24
IV+NV	5.25	5.28	5.25	5.28	0.06	0.6	5.25	5.28
X+BR+IV	0.01	0.62	0.01	0.62	0.01	0.62	0.01	0.62
X+BR+NV	0.01	0.54	0.01	0.54	0.01	0.54	0.63	0.85
X+IV+NV	0.01	0.64	0.01	0.64	0.01	0.64	0.01	0.64
BR+IV+NV	5.25	5.28	5.25	5.28	0.06	0.58	5.25	5.28
X+BR+IV+NV	0.01	0.62	0.01	0.62	0.01	0.62	0.01	0.62

**Table 3**

Nested covariates for defining model specifications for TMLE and IPW estimator

Covariate set	Number of covariates	Covariates included in the model
NULL	0	No covariates
Simple set ( $C_1$ )	8	Predefined important confounders (age, sex, obesity, smoking, history of diabetes)
Intermediate set ( $C_2$ )	44	All pre-specified confounders
Full set ( $C_3$ )	444	All potential confounders ( $C_2$ + the 400 empirical variables selected by the HDPS)

**Table 4a**

Results for various IPW and TMLE approaches in the plasmode simulation study (without truncation in the propensity score)

Approach	Outcome Model	Treatment Model	Estimate	SE	Bias	RMSE	Finite fitted values <sup>*</sup>
IPW1	$A$	$NULL$	-1.19	0.04	0.86	0.86	100%
IPW2	$A$	$C_1$	-0.84	0.04	0.51	0.51	100%
IPW3	$A$	$C_2$	-0.57	0.08	0.24	0.25	100%
IPW4	$A$	$C_3$	-0.35	0.10	0.04	0.11	100%
TMLE1	$A$	$NULL$	-1.19	0.04	0.86	0.86	100%
TMLE2	$A$	$C_1$	-0.84	0.04	0.51	0.51	100%
TMLE3	$A$	$C_2$	-0.57	0.10	0.24	0.26	100%
TMLE4	$A$	$C_3$	-0.31	0.30	0.02	0.31	97.2%
TMLE5	$A, C_1$	$NULL$	-0.84	0.04	0.51	0.51	100%
TMLE6	$A, C_1$	$C_1$	-0.84	0.04	0.51	0.51	100%
TMLE7	$A, C_1$	$C_2$	-0.57	0.08	0.24	0.25	99.8%
TMLE8	$A, C_1$	$C_3$	-0.29	0.23	0.04	0.23	89.4%
TMLE9	$A, C_2$	$NULL$	-0.58	0.05	0.25	0.26	100%
TMLE10	$A, C_2$	$C_1$	-0.58	0.05	0.25	0.25	100%
TMLE11	$A, C_2$	$C_2$	-0.56	0.07	0.23	0.24	100%
TMLE12	$A, C_2$	$C_3$	-0.31	0.12	0.03	0.12	96.8%
TMLE13	$A, C_3$	$NULL$	-0.34	0.05	0.005	0.06	100%
TMLE14	$A, C_3$	$C_1$	-0.34	0.05	0.002	0.06	100%
TMLE15	$A, C_3$	$C_2$	-0.34	0.07	0.002	0.07	100%
TMLE16	$A, C_3$	$C_3$	-0.35	0.16	0.01	0.16	99%

\* Refers to settings in which the first step of the TMLE returned non-infinite predictions.

**Table 4b**

Results for various IPW and TMLE approaches in the plasmode simulation study (with truncation in the propensity score at 0.025 and 0.975)

Approach	Outcome Model	Treatment Model	Estimate	SE	Bias	RMSE	Finite fitted values <sup>*</sup>
IPW1	$A$	$NULL$	-1.19	0.04	0.86	0.86	100%
IPW2	$A$	$C_1$	-0.84	0.04	0.51	0.51	100%
IPW3	$A$	$C_2$	-0.59	0.07	0.26	0.27	100%
IPW4	$A$	$C_3$	-0.45	0.08	0.11	0.13	100%
TMLE1	$A$	$NULL$	-1.19	0.04	0.86	0.86	100%
TMLE2	$A$	$C_1$	-0.84	0.04	0.51	0.51	100%
TMLE3	$A$	$C_2$	-0.54	0.07	0.21	0.22	100%
TMLE4	$A$	$C_3$	-0.31	0.09	0.03	0.09	100%
TMLE5	$A, C_1$	$NULL$	-0.84	0.04	0.51	0.51	100%
TMLE6	$A, C_1$	$C_1$	-0.84	0.04	0.51	0.51	100%
TMLE7	$A, C_1$	$C_2$	-0.55	0.07	0.22	0.23	100%
TMLE8	$A, C_1$	$C_3$	-0.31	0.09	0.03	0.09	100%
TMLE9	$A, C_2$	$NULL$	-0.58	0.05	0.25	0.26	100%
TMLE10	$A, C_2$	$C_1$	-0.58	0.05	0.25	0.25	100%
TMLE11	$A, C_2$	$C_2$	-0.56	0.07	0.23	0.24	100%
TMLE12	$A, C_2$	$C_3$	-0.33	0.08	0.01	0.08	100%
TMLE13	$A, C_3$	$NULL$	-0.34	0.05	0.005	0.06	100%
TMLE14	$A, C_3$	$C_1$	-0.34	0.05	0.002	0.06	100%
TMLE15	$A, C_3$	$C_2$	-0.33	0.08	0.000	0.06	100%
TMLE16	$A, C_3$	$C_3$	-0.34	0.08	0.002	0.08	100%

\* Refers to settings in which the first step of the TMLE returned non-infinite predictions.



**Table 5**

Distribution of propensity score estimated from four different treatment models

Model	Maximum	Upper quartile	Median	Lower quartile	Minimum
NULL	0.58	0.58	0.58	0.58	0.57
Simple( $W_1^a$ )	0.78	0.69	0.54	0.46	0.32
Intermediate( $W_2^b$ )	0.998	0.91	0.66	0.23	0.001
Full( $W_3^c$ )	0.999	0.94	0.68	0.18	0.00

<sup>a</sup> $W_1$ : Predefined important confounders (age, sex, obesity, smoking, history of diabetes);<sup>b</sup> $W_2$ : All pre-specified confounders;<sup>c</sup> $W_3$ : All potential confounders ( $W_2$  + the 400 empirical selected variables).

**Table 6**

Distributions of the stabilized weights from four different propensity score models

Model	Mean	SD	Maximum	Median	Minimum
NULL	1	0	1	1	1
Simple( $W_1^a$ )	1	0.26	1.94	0.87	0.62
Intermediate( $W_2^b$ )	0.999	1.74	232.30	0.63	0.42
Full( $W_3^c$ )	1.00	2.78	1698	0.61	0.42

<sup>a</sup> $W_1$ : Predefined important confounders (age, sex, obesity, smoking, history of diabetes);<sup>b</sup> $W_2$ : All pre-specified confounders;<sup>c</sup> $W_3$ : All potential confounders ( $W_2$  + the 400 empirical selected variables).