ELSEVIER

Journal of Clinical Epidemiology 66 (2013) S91−S98

Journal of Clinical Epidemiology

Targeted maximum likelihood estimation in safety analysis

Samuel D. Lendle[a,b,*], Bruce Fireman[b], Mark J. van der Laan[a]

[a]*Division of Biostatistics, UC Berkeley, 101 Haviland Hall, Berkeley, CA 94720, USA*
[b]*Kaiser Permanente Division of Research, 101 Haviland Hall, Berkeley, CA 94720, USA*

Accepted 19 February 2013

Abstract

Objectives: To compare the performance of a targeted maximum likelihood estimator (TMLE) and a collaborative TMLE (CTMLE) to other estimators in a drug safety analysis, including a regression-based estimator, propensity score (PS)−based estimators, and an alternate doubly robust (DR) estimator in a real example and simulations.

Study Design and Setting: The real data set is a subset of observational data from Kaiser Permanente Northern California formatted for use in active drug safety surveillance. Both the real and simulated data sets include potential confounders, a treatment variable indicating use of one of two antidiabetic treatments and an outcome variable indicating occurrence of an acute myocardial infarction (AMI).

Results: In the real data example, there is no difference in AMI rates between treatments. In simulations, the double robustness property is demonstrated: DR estimators are consistent if either the initial outcome regression or PS estimator is consistent, whereas other estimators are inconsistent if the initial estimator is not consistent. In simulations with near-positivity violations, CTMLE performs well relative to other estimators by adaptively estimating the PS.

Conclusion: Each of the DR estimators was consistent, and TMLE and CTMLE had the smallest mean squared error in simulations.

Keywords: Safety analysis; Targeted maximum likelihood estimation; Doubly robust; Causal inference; Collaborative targeted maximum likelihood estimation; Super learning

1. Introduction

Evaluating the effectiveness and safety of health interventions through observational studies is made more challenging by issues, such as confounding, missing data, and complex longitudinal data structures [1−3]. In an ideal world, an investigator would perform a randomized controlled trial, but this is often impossible or impractical because of cost or ethical concerns. Additionally, it may be impossible to avoid missing data even in randomized trials; for example, a patient may drop out of the study before some outcome of interest is observed. In lieu of a randomized trial, investigators often attempt to answer the same questions with observational data. In observational studies, the health intervention a patient receives is generally not assigned randomly but is chosen by the patient or physician based on characteristics of the patient, such as age, sex, health conditions, or medications the patient is taking.

These issues raise two important questions: When is it possible to estimate the effect of some health intervention on a safety or an effectiveness outcome without bias? and if it is possible, how can we estimate the effect? To address the first question, we discuss the potential outcomes framework and use it to formally define a target causal parameter that we wish to estimate and briefly review conditions under which it is possible to estimate the parameter without bias in Section 2. To address the second question, in Section 3, we review common estimation methods including a method based on the G-computation formula, inverse probability of treatment weighting (IPTW), and propensity score (PS) matching, and compare them to doubly robust (DR) methods such as augmented IPTW (AIPTW), targeted maximum likelihood estimation (TMLE) and collaborative TMLE (CTMLE), an estimator that uses a data-adaptive estimate of the PS in collaboration with the outcome regression. We also discuss methods for estimating the outcome regression and PS, including the data-adaptive super learner algorithm. In Section 4, we compare methods in a real data example, a simplified version of a drug safety surveillance study to motivate our question of interest and demonstrate the estimation methods. The data set is a subset of data

Conflict of interest: The authors have no conflicts of interest related to the content of this article.

* Corresponding author. Tel./fax: 336-972-2725
E-mail address: lendle@stat.berkeley.edu (S.D. Lendle).

0895-4356/$ - see front matter
http://dx.doi.org/10.1016/j.jclinepi.2013.02.017

What is new?

- Performance of outcome regression based, PS based, and doubly robust estimators are compared in realistic simulated situations.
- Advantages of doubly robust estimators are demonstrated when at least one of PS or outcome regressions are consistent.
- Advantages of plug-in estimators and, in particular, CTMLE are demonstrated in near-positivity violation situations.
- Advantages of data adaptive techniques such as the super learner algorithm are demonstrated when parametric models are not sufficient.

from Kaiser Permanente Northern California formatted according to the specification of Food and Drug Administration's Mini-Sentinel drug safety surveillance program. In Section 5, we compare methods in simulation studies. We present some concluding remarks in Section 6.

2. Causal parameter and identifiability

To define a target causal parameter we are interested in, we use the potential outcomes framework, also known as the Neyman−Rubin causal model [4−6]. We begin by defining Y_1 and Y_0 as potential outcomes for a patient had the patient received treatment 1 or treatment 0 (sometimes no treatment or placebo), respectively. The average treatment effect (ATE) is defined as $E(Y_1 - Y_0)$, where E denotes expectation with respect to the distribution of potential outcomes for the population of interest. For a particular patient, one of Y_1 or Y_0 is unobservable and is called counterfactual. Other causal parameters can be defined, such as the causal odds ratio or risk ration when the outcome is binary, but we focus on the ATE in this article.

Define the observed data $O=\{W, A, Y\}$, where W represents baseline characteristics of a patient, A is 1 if the patient receives the target treatment of interest or 0 if she receives the comparator or control treatment, and Y is the patient's observed outcome. We observe n independent and identically distributed copies of O. We assume $Y = Y_A$, the potential outcome under the drug that patient actually received, which is known as the consistency assumption. To be able to estimate ATE, we need to write it as a function of the observed data distribution. If we can do this, we say the ATE is identifiable. Because the ATE depends on unobserved potential outcomes, identifiability requires some assumptions. The first, known as the randomization assumption, is that given a set of baseline covariates W, the treatment A is independent of the potential outcomes Y_1 and Y_0. This is also called the ''no unmeasured confounders'' assumption. The second is the positivity assumption, in which we assume that for any value of baseline characteristics W, it is possible to receive either treatment, or $0 < P(A=1|W) < 1$ for all W, where P denotes probability. Under these assumptions, then we can write

$$E(Y_1 - Y_0) = E[E(Y|A=1, W) - E(Y|A=0, W)] = \psi_0, \quad (1)$$

so the ATE is equal to a statistical parameter that is a function of only the observed data distribution. A proof of identifiability is provided in Section 2 of the Appendix (see at www.jclinepi.com) for pedagogical purposes, but the result is well known [6−8]. The selection of variables to be included in W requires careful consideration and is discussed in more detail by Greenland et al. [9], Pearl [6], and Howards et al. [10].

3. Estimation

To estimate the causal effect, in addition to the randomization and positivity assumptions, we need to specify a statistical model or a set of possible probability distributions for the observed data O. A probability distribution for O can be factorized into the distribution of Y given A and W, the distribution of A given W, and the distribution of W. Because in an observational study we generally do not have enough knowledge about the data to posit a parametric model, we will put no restrictions on the distribution of the data and use the nonparametric model. In other settings, we may have knowledge that lets us use a more restrictive or even parametric model; for example, in a randomized controlled trial, we know that treatment is independent of the covariates. Knowledge such as this can be incorporated into the statistical model.

Traditional methods for estimating ψ_0 are usually based on an estimate of $E(Y|A, W)$, which we call the outcome regression, or are based on an estimate of the probability of being treated given baseline covariates, $P(A=1||W)$, known as the PS [8]. Using an estimate of the outcome regression, ψ_0 can be estimated using the G-computation formula discussed in the Appendix (see at www.jclinepi.com). Estimators based on the G-computation formula are called plug-in estimators. In general, the coefficient on the treatment variable in an outcome regression cannot be interpreted as a marginal causal effect, but when the regression is correctly specified, it can be used to test the null hypothesis $\psi_0 = 0$. Common PS-based methods to estimate ψ_0 include inverse probability of treatment-weighted estimators (IPTW) [11] and PS matching estimators [8,12], discussed in the Appendix (see at www.jclinepi.com).

For outcome regression methods and PS-based methods to consistently estimate the parameter ψ_0, the initial estimator for the outcome regression or the PS must be consistent. By consistent, we mean that as the sample size increases, the estimator converges (in probability) to the true function,

either $E(Y|A,W)$ or $P(A=1|W)$. We discuss estimation of these functions in the later sections. Another type of estimator of ψ_0, called a DR estimator, combines initial estimators of the outcome regression and the PS such that it is consistent if either of the initial estimators is consistent. Two examples of DR estimators are AIPTW [13] and TMLE [14−16]. Under regularity conditions, both AIPTW and TMLE are regular asymptotically linear estimators, which means they can be expressed as ψ_0 plus the average of a mean zero function called an influence curve and a small remainder. Regular asymptotically linear estimators are asymptotically normal, and their variance is the variance of their influence curve [17]. In addition to being DR, AIPTW and TMLE are efficient when the initial estimates for the outcome regression and the PS are both consistent.

An efficient estimator achieves the minimum asymptotic variance of all regular estimators, with influence curve equal to the efficient influence curve (EIC). AIPTW uses the EIC as an estimating equation and estimates ψ_0 by solving the EIC equation directly and is discussed further in Section 3.3 of the Appendix (see at www.jclinepi.com). On the other hand, the TMLE updates the initial outcome regression and plugs this updated outcome regression into the G-computation formula, so it is a plug-in estimator as well. The initial outcome regression is updated in such a way that the EIC equation is implicitly solved. Plug-in estimators tend to be more stable than non−plug-in estimators because they guarantee that the estimate falls in the parameter space. This is an advantage over other estimators such as AIPTW in small samples, as shown in the simulations. Porter et al. [18] discuss and compare other DR estimators.

The first step in TMLE is to calculate initial estimates of the outcome regression and the PS, which we call $\widehat{E}(Y|A=a,W=w)$ and $\widehat{P}(A=1|W)$. We then update the initial outcome regression using the estimated PS. For a binary outcome as in our example, we perform this update on the logit scale, so

$$\operatorname{logit}\left(\widehat{E}^{*}(Y|A=a,W=w)\right)=\operatorname{logit}\left(\widehat{E}(Y|A=a,W=w)\right)+\widehat{\epsilon}h(a,w)$$

where $\operatorname{logit}(x)=\log(x/(1-x)), h(a,w)=a/\widehat{P}(A=1|W=w)-(1-a)/(1-\widehat{P}(A=1|W=w))$ and $\widehat{\epsilon}$ is the maximum likelihood estimate from the logistic regression model with

$$\operatorname{logit}(E(Y|A=a,W=w))=\operatorname{logit}\left(\widehat{E}\left(Y|A=a,W=w\right)\right)+\epsilon h(a,w) \quad (2)$$

Using standard logistic regression software with Y_i as an outcome and $\operatorname{logit}\widehat{E}(Y|A=A_i,W=W_i)$ as an offset with no intercept, $\widehat{\epsilon}$ is the estimated coefficient in front of the covariate $h(A_I,W_i)$. Once the updated estimate is obtained, the final estimate for ψ_0 is calculated by plugging $\widehat{E}^{*}(Y|A=a,W=w)$ into the G-computation formula. Additional details are provided in Section 3.4 of the Appendix (see at www.jclinepi.com).

CTMLE [16,19,20] is an extension of TMLE that uses an estimate of the PS that is estimated in collaboration with the outcome regression. Standard TMLE uses an estimate of the PS based on all W. When the initial estimate of the outcome regression is adjusting for some or all covariates in W very well, or some covariates are not related or only weakly related to the outcome, updating the initial outcome regression based on an estimate of the PS adjusting for all W can be harmful, increasing the variance of the estimate. CTMLE attempts to avoid this by constructing a sequence of updated outcome regressions based on PS estimates that incorporate an increasing number of covariates. Covariates are added to the PS estimate in a stepwise fashion and are chosen based on a penalized log-likelihood statistic from the logistic regression model in Eq. (2). The number of steps is chosen based on the cross-validated log-likelihood statistic. This can lead to gains in efficiency and more robustness in settings when the positivity assumption is nearly violated. We discuss the algorithm further in Section 3.5 of the Appendix (see at www.jclinepi.com), and Gruber and van der Laan [20] provide a detailed example.

Previously, we note that the consistency of an estimator of ψ_0 depends on the consistency of the initial estimator of the outcome regression or the PS. In the nonparametric model, the form of these functions is not known. A candidate estimator for $E(Y|A,W)$ or $P(A=1|W)$ could be a parametric estimator like logistic regression, a nonparametric or machine learning estimator like random forest [21] or support vector machines [22], or a semiparametric estimator such as a generalized additive model [23]. For more nonparametric estimators, fewer conditions are required for consistency, but the estimates tend to be more variable, particularly at smaller sample sizes. Parametric estimators require more conditions for consistency, like that the true outcome regression or PS follow a specified functional form, but usually have a smaller variance. In small samples, more parametric estimators can sometimes approximate the true function better than flexible nonparametric estimators in terms of variance at the cost of some bias. The outcome regression and PS can also be estimated with the super learner algorithm [16,24]. The algorithm chooses the best combination of a library of multiple parametric and machine learning estimators using crossvalidation to avoid overfitting in small samples and performs at least as well as the best candidate estimator in the library asymptotically. This allows the analyst to choose many candidate estimators ranging from machine learning algorithms to parametric estimators proposed by subject matter experts before looking at the data without having to choose only one estimator that may not perform well in a particular application. In the third simulation in Section 5, we see the importance of using such a data-adaptive algorithm in an

example where parametric estimators are not sufficient for consistency.

4. Data example

Using a real data set from Kaiser Permanente, we compare an unadjusted estimator, an outcome regression-based estimator, PS-based estimators, and DR methods. The population of interest is diabetic patients without prior cardiovascular disease who are new users of one of two antidiabetic pharmaceutical agents. The two drugs of interest are pioglitazone and sulfonylurea for our example, although these two drugs are not of primary interest in the Mini-Sentinel program. The outcome of interest in our example is the occurrence of an acute myocardial infarction (AMI) during the first 6 months of new antidiabetic drug use. Because of loss to follow-up, the outcomes for some patients are missing. When adjusting for missingness, we did not see a meaningful difference in the results; therefore, for simplicity, we only include observations with nonmissing outcomes in the analyses in this article. All methods in Section 3 can be extended to account for outcomes subject to missingness. The data set also includes about 50 covariates including demographic information, such as sex and age, comorbidities, and other drug use. Covariates were selected by cardiovascular epidemiologists during the study design phase of Mini-Sentinel program. Although for this example we use data only from Kaiser Permanente Northern California, the Mini-Sentinel program involves data from multiple collaborating organizations, so covariates were chosen based on availability at all data partners. In Figure 1 of Appendix (see at www.jclinepi.com), histograms of the distribution of the estimated PS by treatment are presented. Although for the sulfonylurea group ($A = 0$) the PS is lower as expected, the PS in both groups overlaps with most observations falling below 0.4. It is possible, because of limitations of the measured data, that the set of available covariates is not sufficient for the randomization assumption to hold, so we do not know if we can interpret an estimate of ψ_0 as an estimate of the ATE, but it is still useful to compare different methods of estimating ψ_0.

Table 1 summarizes (by drug) the number of patients at risk and number of AMIs observed in the first 6 months after starting a new antidiabetic drug. The unadjusted estimator estimates the ATE according to the difference between the proportion of AMIs in the pioglitazone and sulfonylurea groups. For this example, we use logistic regression for the outcome regression and the PS, with all baseline covariates, as main terms in the estimator. In addition to all baseline covariates, the logistic regression estimator for the outcome regression also includes an indicator of treatment of interest. Results are presented in Table 2. All methods estimate the ATE to be very close to zero; all estimates that adjust for confounders, other than PS matching, are closer to zero than the unadjusted estimate. Because the rate of AMIs is so low and we are not following patients for a long period, it is not surprising that we do not find a large difference in rates of AMI between the two drugs. Although the results from all methods are similar in this particular data set with a rare outcome, in general this will not be the case. Differences between the estimation methods are highlighted in the simulations in the following section.

Table 1. Summary of outcome by treatment

Variable	Treatment	Comparator	Total
Total patients	2,146	25,022	27,168
AMIs, *n* (%)	5 (0.23)	86 (0.34)	91 (0.33)

Abbreviation: AMIs, acute myocardial infarctions.

Table 2. Results from real data set

Method	Estimate	Standard error	*P*-value
Unadjusted	−0.0011	0.0013	0.39
G-computation	−0.0007	0.0014	0.61
PSM	−0.0013	0.0017	0.45
IPTW	−0.00005	0.0015	0.75
AIPTW	−0.0003	0.0015	0.86
TMLE	−0.0004	0.0015	0.80
CTMLE	−0.0010	0.0011	0.38

Abbreviations: PSM, propensity score matching; IPTW, inverse probability of treatment weighting; AIPTW, augmented inverse probability of treatment weighting; TMLE, targeted maximum likelihood estimator; CTMLE, collaborative targeted maximum likelihood estimator.

5. Simulations

In this section, we compare different estimators in simulation studies. To create realistic simulated scenarios, we use the empirical distribution of baseline covariates from a real data set to generate the simulated distribution of baseline covariates and specify the ''true'' outcome regression based on a regression estimated from the data. Instead of using pioglitazone and sulfonylureas, we create a false treatment that is not related to AMI (so the true ATE is known to be zero) and specify the PS as a function of baseline covariates. We conducted simulation studies based on three simulation schemes detailed in Section 4 of the Appendix (see at www.jclinepi.com). For each simulation, results are based on 1,000 data sets of each sample size. Results are presented in figures. For each simulation, tables of results and histograms of the distribution of the estimated PS by treatment are included in Section 5 of the Appendix (see at www.jclinepi.com).

In the first simulation, we chose 12 binary baseline covariates to be confounders that were strongly related to AMI in the real data set, and set the PS and outcome regression such that a logistic regression model with all 12 covariates as main terms is correctly specified. To investigate bias because of an inconsistent outcome regression or PS estimator, we left six of the 12 important confounders out

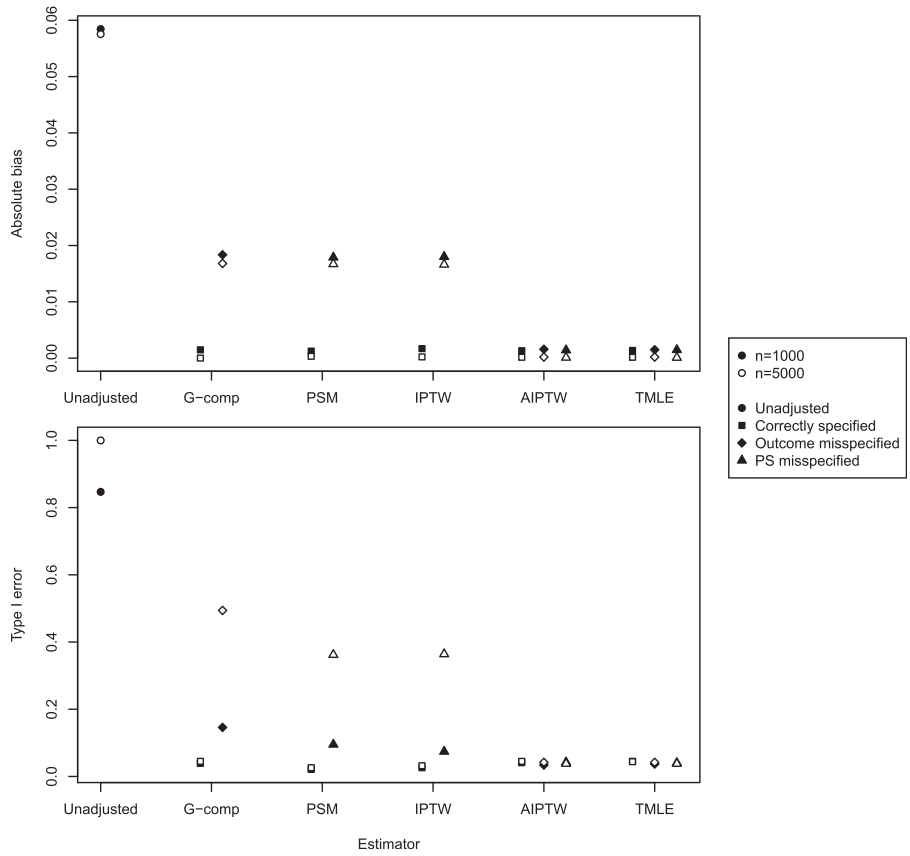

Fig. 1. Simulation results demonstrating double robustness. G-comp, G-computation; PSM, propensity score matching; IPTW, inverse probability of treatment weighting; AIPTW, augmented inverse probability of treatment weighting; TMLE, targeted maximum likelihood estimator.

of the logistic regression estimators. Although in practice, a researcher would generally not arbitrarily leave observed variables out of a regression estimator, this is analogous to failing to include important interaction terms or misspecifying a functional form with continuous covariates. Results are shown in Fig. 1. We see that all methods have low bias when the outcome regression and PS are correctly specified. We also see that when only either the outcome regression or the PS is correctly specified, both AIPTW and TMLE perform well, with bias going down quickly as sample size increases, demonstrating the double robustness property. The type I error rate is high and increases with sample size for the unadjusted estimator and the non-DR estimators when the initial estimator is not consistent.

In the second simulation, the true outcome regression is the same as in the first simulation, but the PS now depends on an additional covariate, an indicator of taking another diabetes medication at baseline that is not related to treatment. Given that a patient taking another medication at baseline, the probability of receiving treatment $A = 1$ is small, as low as 0.005, with the median being 0.0346. This is a near-positivity violation, making ψ_0 difficult to estimate. Because this additional covariate is related to treatment but not the outcome (other than through the other covariates), we can estimate ψ_0 without adjusting for it, although in reality we will often not know this ahead of time. Histograms representing the distribution of the estimated PS by treatment, with and without adjusting for the additional covariate, are provided in Figure 3 of the Appendix (see at www.jclinepi.com). When the covariate is excluded from the model, there are fewer observations with estimated PS very close to zero, particularly in the $A = 0$ group, the distributions of the PS are more similar in shape between the $A = 1$ and $A = 0$ groups. We note that even without adjusting for this additional covariate, there are still many observations with PS near 0, underscoring how difficult estimation of ψ_0 is in this setting.

Results for the second simulation are presented in Fig. 2. We see that the G-computation estimator performs the best because it relies only on the outcome regression; but we recall from the previous simulation that it is not DR. The mean squared error for both IPTW and AIPTW suffered because of inverse weighting by the PS, particularly at smaller sample sizes, and some estimates were outside the parameter space, from -1 to 1. Although TMLE also uses an estimate of the PS to update the outcome regression, it does so on the logit scale, so the effect of very small estimated PS values on the final estimate is smaller. We also observe that among the DR methods, CTMLE performs the best and almost as well as the G-computation estimator because it can exclude covariates from the PS estimator that do not help estimate the final

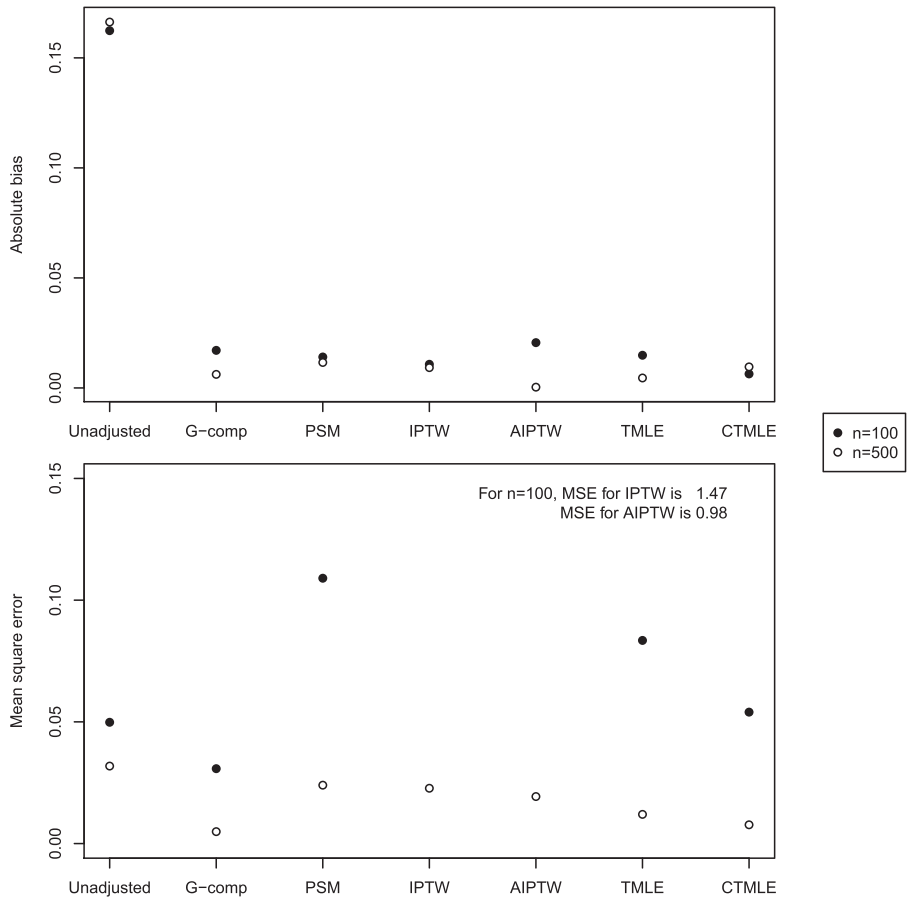

Fig. 2. Simulation results investigating near violations of the positivity assumption. G-comp, G-computation; PSM, propensity score matching; IPTW, inverse probability of treatment weighting; AIPTW, augmented inverse probability of treatment weighting; TMLE, targeted maximum likelihood estimator; CTMLE, collaborative targeted maximum likelihood estimator; MSE, mean squared error.

parameter via crossvalidation. These can include covariates that are already adjusted for in the outcome regression, or covariates that are related only to the treatment and not the outcome, in this case taking another diabetic medication at baseline.

In a third simulation, we chose four baseline covariates as confounders, and construct the outcome regression such as logistic regression estimator with the baseline covariates as main terms is correctly specified. For the PS, we included two pairwise interaction terms so that main terms only logistic regression would not sufficiently adjust for confounding. We compare estimators using main terms logistic regression for the PS, and the super learner algorithm that uses crossvalidation to choose the best combination of main terms logistic regression, stepwise logistic regression including interaction terms, and logistic regression with all interaction terms. For DR methods, we do not adjust the outcome regression so we can see how choice of PS estimator affects the performance. Results are shown in Fig. 3. For all methods, we observe that estimating the PS using the super learner algorithm results in a reduction in bias over main terms logistic regression, and for most methods, the bias is not decreasing quickly as sample size increases. We note that PS matching has low bias even when the PS is estimated with main terms logistic regression because of the small number of binary covariates that results in a very discrete estimated PS but will not be true in general. In this case, the matching estimator is able to match observations that are similar in the true PS although the estimated PS is not close to the true PS.

6. Discussion

In this article, we reviewed the identification of the ATE and discussed estimation methods, highlighting TMLE and CTMLE. We compared estimation methods in a real data example. In the example data set, the methods performed similarly; we do not expect this to be the case in general. We also examine the methods in simulation studies, exploring double robustness, sensitivity to near violations of the positivity assumption, bias because of inconsistent initial estimators, and the use of the data-adaptive super learner algorithm. In the Appendix (see at www.jclinepi.com), we provide detail on each of the estimators we discuss in the article.

In the result from the first simulation, we see the danger of an inconsistent initial estimator in terms of bias and type

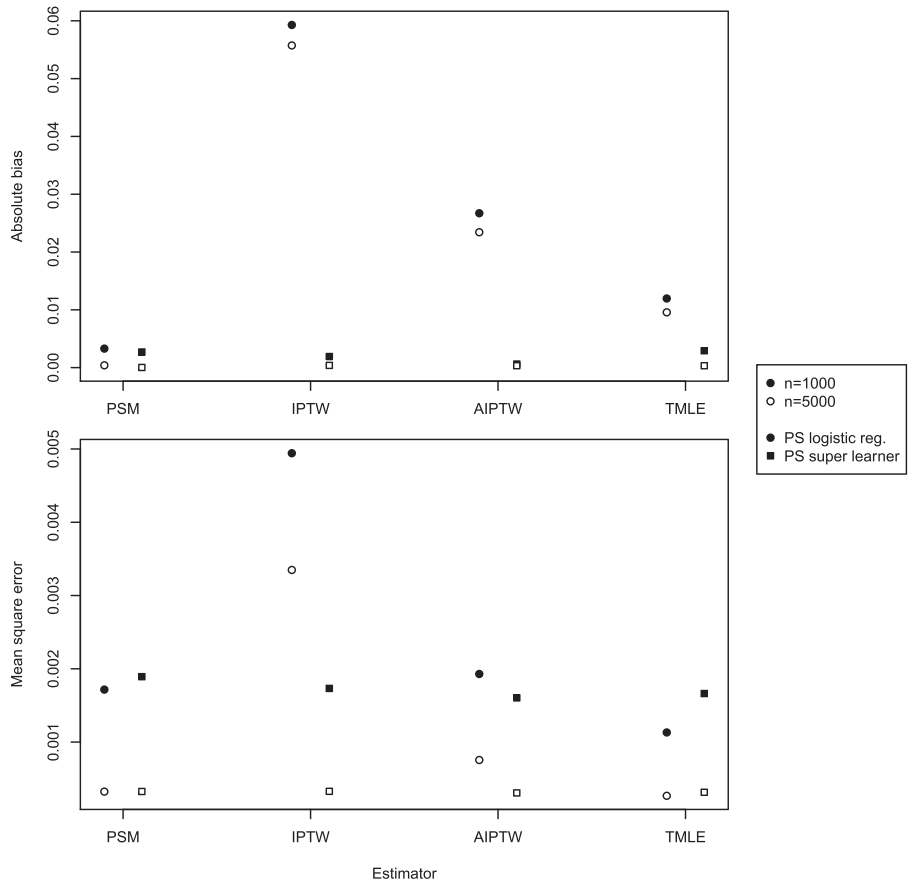

Fig. 3. Simulation results demonstrating the use of the super learner algorithm to estimate the propensity score. PSM, propensity score matching; IPTW, inverse probability of treatment weighting; AIPTW, augmented inverse probability of treatment weighting; TMLE, targeted maximum likelihood estimator.

I error rate. When an estimator is biased because of some sort of misspecification and with large sample sizes common to many epidemiological studies, this bias does not decrease and will always lead to rejection of the null hypothesis. This emphasizes the importance of DR estimation methods, which allow for two chances at reducing bias in estimates. In the second simulation, we saw how near violations of the positivity assumption can greatly harm estimation methods relying heavily on the PS, particularly through weighting. This example particularly highlights the importance of a data-adaptive estimation method like CTMLE, which chooses covariates for the PS with the goal of estimating ψ_0 not just predicting treatment.

All the methods in Section 3 require estimation of at least one of the outcome regression or the PS, and in practice, this is often done with parametric estimators like logistic regression, which are generally misspecified and therefore inconsistent. When the estimator of the outcome regression or PS is inconsistent, the final estimate of ψ_0 can have a bias that does not decrease with sample size, and data-adaptive or nonparametric methods are often preferable. In the third simulation, we see the benefit of using the super learner algorithm for estimating the PS to reduce bias in the final estimates of ψ_0. The super learner algorithm can also be used to estimate the outcome regression.

In addition to estimating the statistical parameter well, it is important to choose a causal parameter that answers the right question. Instead of the ATE, we may instead be interested in the ATE among a subset of the target population. For example, in the second simulation, we could have targeted the causal effect among patients not taking the other diabetes medication at baseline. The ATE among the treated is another causal parameter that requires slightly weaker positivity assumptions [25,26]. Crump et al. [27] and Stürmer et al. [28] propose estimation methods based on PS trimming that effectively change the statistical parameter from ψ_0 to a similar parameter for which there is more support in the data. The interpretation of these estimates causally is more difficult, however, because the populations in which they are defined are dependent on the data.

TMLE is a general method that provides a framework for developing semiparametric efficient estimators for parameters in a variety of settings and has been applied in randomized controlled trials, matched case−control studies, estimation of parameters of marginal structural models, longitudinal studies with multiple time point interventions, and estimation of direct effects [16,29,30].

Acknowledgments

We thank two anonymous reviewers for their insightful comments and suggestions. S.D.L. and B.F. were supported by Kaiser Permanente Northern California. M.v.d.L. was supported by National Institutes of Health grant R01 AI074345-04.

Appendix

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jclinepi.2013.02.017.

References

[1] Greenland S, Morgenstern H. Confounding in health research. Annu Rev Public Health 2001;22(1):189−212.
[2] Frangakis CE, Rubin DB. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. Biometrika 1999;86(2):365−79.
[3] Gill RD, Robins JM. Causal inference for complex longitudinal data: the continuous case. Ann Stat 2001;29:1785−811.
[4] Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol 1974;66(5):688.
[5] Holland PW. Statistics and causal inference. J Am Stat Assoc 1986; 81:945−60.
[6] Pearl J. Causality: models, reasoning, and inference. Cambridge, UK: Cambridge University Press; 2000.
[7] Rubin DB. Multivariate matching methods that are equal percent bias reducing, II: maximums on bias reduction for fixed sample sizes. Biometrics 1976;32:121−32.
[8] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70(1):41.
[9] Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology 1999;10:37−48.
[10] Howards PP, Schisterman EF, Poole C, Kaufman JS, Weinberg CR. Toward a clearer definition of confounding revisited with directed acyclic graphs. Am J Epidemiol 2012;176:506−11.
[11] Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000;11:550−60.
[12] Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. J Econ Surv 2008;22(1): 31−72.
[13] Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. J Am Stat Assoc 1994;89:846−66.
[14] van der Laan MJ, Rubin D. Targeted maximum likelihood learning. Int J Biostat 2006;2(1). http://dx.doi.org/10.2202/1557-4679.1043.
[15] Bembom O, Petersen ML, Rhee SY, Fessel WJ, Sinisi SE, Shafer RW, et al. Biomarker discovery using targeted maximum-likelihood estimation: application to the treatment of antiretroviral-resistant HIV infection. Stat Med 2009;28:152−72.
[16] van der Laan MJ, Rose S. Targeted learning: causal inference for observational and experimental data. New York, NY: Springer; 2011.
[17] Bickel PJ, Klaassen CAJ, Ritov Y, Wellner JA. Efficient and adaptive estimation for semiparametric models. Baltimore, MD: The Johns Hopkins University Press; 1993.
[18] Porter K, Gruber S, van der Laan M, Sekhon J. The relative performance of targeted maximum likelihood estimators. Int J Biostat 2011;7(1): 1−34.
[19] van der Laan MJ, Gruber S. Collaborative double robust targeted maximum likelihood estimation. Int J Biostat 2010;6(1). http://dx.doi.org/10.2202/1557-4679.1181.
[20] Gruber S, van der Laan MJ. An application of collaborative targeted maximum likelihood estimation in causal inference and genomics. Int J Biostat 2010;6. Article 18.
[21] Breiman L. Random forests. Mach Learn 2001;45(1):5−32.
[22] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995; 20(3):273−97.
[23] Hastie T, Tibshirani R. Generalized additive models. Stat Sci 1986;1: 297−310.
[24] van der Laan MJ, Polley EC, Hubbard AE. Super learner. Stat Appl Genet Mol Biol 2007;6(1).
[25] Hahn J. On the role of the propensity score in efficient semiparametric estimation of average treatment effects. Econometrica 1998;66(2): 315−31.
[26] van der Laan MJ. Estimation of causal effects of community based interventions. U.C. Berkeley Division of Biostatistics Working Paper Series; 2010. 268.
[27] Crump RK, Hotz VJ, Imbens GW, Mitnik OA. Dealing with limited overlap in estimation of average treatment effects. Biometrika 2009; 96(1):187−99.
[28] Stürmer T, Rothman KJ, Avorn J, Glynn RJ. Treatment effects in the presence of unmeasured confounding: dealing with observations in the tails of the propensity score distribution—a simulation study. Am J Epidemiol 2010;172:843−54.
[29] 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).
[30] Zheng W, van der Laan MJ. Targeted maximum likelihood estimation of natural direct effects. Int J Biostat 2012;8(1).