Doubly Robust Estimation of Causal Effect: Upping the Odds of Getting the Right Answers

Authors**:** Xiaochun Li[1] and Changyu Shen[2]

[1]Department of Biostatistics
Indiana University School of Medicine
Indianapolis, IN, USA

[2]Smith Center for Outcomes Research in Cardiology
Beth Israel Deaconess Medical Center
Harvard Medical School
375 Longwood Avenue|4th Floor
Boston, MA 02215

Corresponding author:
Xiaochun Li
Department of Biostatistics
Indiana University
School of Medicine & Richard M. Fairbanks School of Public Health
410 W 10th St., Suite 3000
Indianapolis, IN, 46202
Email: xiaochun@iu.edu
Phone: 317-274-2696

Abbreviations:
PS: propensity score
IPTW: inverse probability weighted
DR: doubly robust

Key words: confounding, propensity score, inverse probability weighting, outcome regression, doubly robust

Abstract word count: 121

Main text word count: 3341

_____

Abstract

Propensity score based-methods or multiple regressions of the outcome are often used for confounding adjustment in analysis of observational studies. In either approach, a model is needed: A model describing the relationship between the treatment assignment and covariates in the propensity score based-methods, or a model for the outcome and covariates in the multiple regressions. The two models are usually unknown to the investigators and must be estimated. The correct model specification, therefore, is essential for the validity of the final causal estimate. We describe in this paper a doubly robust estimator which combines both models propitiously to offer analysts two chances for obtaining a valid causal estimate, and demonstrate its use through a data set from the Lindner Center Study.

1. Background

Observational data is a rich source for investigation of exposure effects, for example, in the research of comparative effectiveness and safety of treatments. Exposure effects, or causal effects can be formulated in terms of potential outcomes in Rubin's causal model framework[1]. Given a choice of two exposures (mutually exclusive for a certain time interval), a patient has a pair of potential outcomes, one outcome for each exposure. In actuality, a patient is assigned to only one exposure and thus only one potential outcome is observed. Suppose we could rewind the time machine, and have the patient experience the other exposure and observe the outcome, the causal effect for this patient can be defined as some contrast (e.g. absolute difference) between the two potential outcomes under respective exposures, and the average causal effect (ACE) is the average of differences of the two potential outcomes across all patients in a relevant population. Because we can only observe one potential outcome in reality, patient-level causal effect is not directly estimable. Thus, it is usually the ACE that is the target of most studies. In fact, ACE can also be defined without the conceptual involvement of patient-level causal-effect. For example, risk differences, average causal effects for binary outcomes can be risk ratios or odds ratios.

Randomization in clinical trials is a tool that enables the estimation of ACE through randomly creating exposure groups, both of which are representative of the entire population with the only difference being the treatment assignment. However, although randomization is useful in ensuring exchangeability of treatment groups, the ACE on the efficacy of an intervention can only be estimated when the randomized trial is "perfectly"

executed, with perfect adherence to assigned treatment, no protocol deviations, perfect data capture and no loss of patients to follow up and so on. In contrast, the treatment assignments in observational studies are not randomized and often depend on patients' characteristics (confounding by indication), simple group comparisons tend to lead to biased causal estimates. For example, physicians make treatment decisions based on patients' demographics and pre-treatment health status that are associated with the outcomes, rendering differences in outcomes between treatment groups not exclusively attributable to the causal effect of the treatment on the outcomes. Therefore, factors affecting the treatment assignment and associated with the outcome need to be accounted for to remove the confounding bias in the analysis to produce valid causal estimate. In causal inference, the most essential assumption for unbiased estimation of a treatment effect is that we have at our disposal *all variables* (X) that are associated with both the treatment and the outcome (namely, confounders; this assumption is known as the assumption of no unmeasured confounders). Various statistical methods have been developed to adjust for confounding to enable valid comparisons, among which multivariable regression models of the outcome and propensity score-based approaches are often used in applications[2-6]. Notably, the topic of propensity score-based methods as well as their implementations is the focus of an earlier article[5] in the series of Primer on Statistical Methods of this journal.

In this report, we describe two common but contrasting approaches to causal effect estimation: multivariable regressions that focus on outcome modelling and inverse probability weighting using propensity score that focuses on exposure modelling. We discuss the common challenges to both, and then introduce their fortuitous combination — the doubly robust (DR) estimator. Using a public data set from the Lindner Center Study,

we demonstrate the steps of DR estimation to assess the effect of the treatment of abciximab on patients' 6-month mortality and the cardiac related costs incurred within 6 months of patients' initial percutaneous coronary intervention.

1.1 Multivariable Outcome Regressions

The most common approach to causal estimation is the prediction of potential outcomes for every subject via the use of the classic multivariable outcome regression models — for example, linear, logistic, Cox PH models, and so on — in which an outcome is modelled in terms of the exposure variable and a set of baseline covariates, which should include variables associated with both the exposure and the outcome (confounders), and preferably variables related to the outcome only (prognostic factors). The confounders are necessary for reducing bias in the causal estimate and improving its precision; the prognostic factors are primarily for further enhancement in the precision of the estimate.

We will focus on the two common outcomes, continuous and binary. Estimating the marginal ACE requires marginalization. Specifically, the estimated outcome regression model is used to produce a pair of predicted potential outcome values for each subject using this subject's baseline covariates, one for each exposure. The marginal mean of each potential outcome can be estimated by taking the average across all subjects. The ACE can then be constructed based on the specific metric of interest by using the two estimated marginal means, be it the treatment difference, the risk difference, the risk ratio or the odds ratio. The marginalization process is the same if two models are used for a continuous outcome with one model for each exposure group[7]. The outcome regression

estimator is consistent for ACE provided that there are no unmeasured confounders (all

is, all relevant confounders are available and included in the analysis) and the relationship

between the outcome and the treatment and a host of covariates is specified correctly in

the multivariable model for the outcome.

1.2 Inverse Probability Weighting using the Propensity Score

The propensity score (PS) of an exposure (or treatment) for a subject is defined as the

conditional probability of receiving the exposure given the values of the baseline

confounders of the subject. The PS is a one-dimensional representation of the multi-

dimensional (even high-dimensional in some cases) baseline confounders. The fact that

subjects with the same PS share a common distribution of the baseline confounders

(included in the PS model) across exposure groups enables researchers to control for

confounding through the use of the PS so as to derive causal estimate with minimal bias.

The PS adjustment is especially useful when the conventional approach of a multivariable

regression to simultaneously control many confounders is not feasible because of the rarity

of the outcome. The PS can be used in the analyses in a variety of ways: PS matching, PS

stratification, inverse probability weighting by PS (IPTW) or PS as a covariate in an

outcome regression[5]. These PS approaches are equivalent (see Appendix B) subject to the

validity of three assumptions: a) there are no unmeasured confounders, b) the PS is strictly

greater than zero and less than one, and c) models are correctly specified in respective PS

approach. In practice, PS matching is often used to estimate the average causal effect for

the treated rather than ACE. The approach of IPTW is to weight each subject with the

inverse of the subject's PS or 1-PS: each subject with confounder values X in a treatment

group is assigned a weight of 1/PS indicating that there are 1/PS subjects in the population with confounder values X; and each subject with confounders X in a control group is assigned a weight of 1/(1-PS) for similar reasons[8]. The consequence of such a weighting scheme is that the weighted groups are "similar" in confounders (that is, having the same distribution of the confounders) and thus comparable. Any differences in the distribution of the outcomes between the weighted groups can then be attributed to the treatment, after the differences in the confounders being accounted for by PS weighting. The ACE can then be estimated using the weighted outcomes from the two groups.

1.3 Model Estimation of Outcome Regression and PS

Regardless of which approach researchers take to estimate the ACE, a model is needed: an outcome regression model in the approach of modeling the outcome, or a PS model in the approach of modeling the exposure. Assuming no uncontrolled confounders, the validity of the final causal estimates is contingent on the validity of respective models. To build a model in general, we need to decide what independent variables should be included and how these variables are related to the dependent variable. Hence, either the PS model or the outcome regression model should include X. In addition, it is shown[9,10] that there may be further gains in efficiency to include in either models the covariates V that are not related to treatment exposure but are associated with potential response. Therefore, all confounders X and all covariates that are prognostic should be included provided the sample size allows for reliable estimation of the model.

Model checking and diagnostics for either models should follow the standard practice. Furthermore, for propensity-based methods in general, the PS distributions should be examined for overlap. In the case of substantial non-overlap, researchers may want to restrict their estimation task to the overlap, and as a consequence clarify the appropriate inferential population to which the causal estimate resulting from the subsample is applicable.

2 The Doubly Robust Estimator

The doubly robust (DR) estimator we discuss in this paper is a propitious combination of the IPTW and outcome regression, which is a consistent estimator (i.e., verging to be unbiased as sample size tends toward infinity) if either the PS model or the outcome regression model is correctly specified, and is the most efficient if both are correct[11]. The idea behind this estimator is proposed by Robins et al[11] to improve the IPTW by augmenting it with the prognostic information in the confounders. When the PS model is correct, the expectation of the IPTW term is the causal effect and the expectation of the second term is zero because of the expectation of the weighted residuals of the PS model is zero, even if the outcome regression model is wrong; when the outcome regression model is correct, the expectation of both terms combined yields the causal effect, even if the PS model is wrong. Of course, when both the PS and the outcome regression models are correct, the DR estimator estimates the causal effect of interest, whereas when both are wrong, the DR estimator does not necessarily estimate the causal effect. Nonetheless, the DR estimator does offer us two opportunities to get a valid estimate of the causal effects through the specifications of the PS and outcome regression models. Since the

DR estimator we present here is an extension of the IPTW, it is also called augmented IPTW method[5].

2.1 Implementation

The implementation of the DR estimator for ACE in terms of marginal differences in potential outcomes is straightforward as depicted in the Flowchart in Figure 1:

1. Estimate the PS model, and obtain the predicted probability $\hat{\pi}_i$ for every subject in the entire sample.

2. Estimate the outcome regression models, one for each treatment group. Once the models are estimated, obtain the predicted values for the *entire sample* using each estimated model. The predicted values $\hat{Y}_{i1}$ from the model estimated based on the exposed group are the counterfactual outcomes had all subjects been exposed; likewise, the predicted values $\hat{Y}_{i0}$ from the model estimated based on the non-exposed group are the counterfactual outcomes had all subjects not been exposed.

3. Calculate the DR causal estimate $\hat{\Delta}_{DR}$ (see the appendix for notations).

If one is interested in obtaining ACE in terms of the risk ratio or the odds ratio for a binary outcome, one can obtain the marginal means of the potential outcomes — the two terms as separated by the minus sign in Step 3 — and calculate the risk ratio or the odds ratio.

The DR estimator as described above is implemented in a SAS macro[12] developed at the University of North Carolina at Chapel Hill, where the macro and a sample dataset are available for download. The macro is suited to the scenario of a dichotomous treatment

variable, and either a dichotomous or a continuous outcome variable. The authors demonstrated in the manual through the sample data set that DR estimates are consistent as long as one of two models, PS or outcome regression, is specified correctly. The authors include in their manual examples with a binary and a continuous outcome.

In the next section, we will analyze one real-life data example using SAS, a popular platform for statistical analysis. SAS's new procedure (as of November 2016) `causaltrt` offers all three estimators, outcome regression, IPTW and DR.

3 Application to the Lindner data

3.1 The Lindner Center study

The Lindner Center Study is an observational study to evaluate the impact of adjunctive pharmacotherapy with abciximab platelet GP IIb/IIIa blockade during percutaneous coronary intervention (PCI) on costs and clinical outcomes in a high-volume interventional practice, Ohio Heart Health Center of The Christ Hospital, in 1997[13]. Confounding bias exists as the treatment was not assigned by randomization. In the original publication, PS stratification method was used to adjust for confounding bias. The average reduction in mortality rate at 6 months after abciximab therapy increases from the crude 3.4% to 4.9% after the PS stratification adjustment and no substantially higher cost is found associated with the treatment. However, due to the issues of data privacy and confidentiality, only a subset of the study data is included in the R package USPS[14] for public use. This dataset will be used in this paper to illustrate the DR approach to estimate the average treatment effect of abciximab on both the mortality and the cost.

The dataset in the R package `USPS` consists of observations on 996 patients who were followed for at least 6 months after receiving an initial PCI. Among them, 698 of them received the treatment with the abciximab platelet GP IIb/IIIa blockade during PCI, and the remaining 298 patients received usual-care-alone while undergoing their PCI. For each patient, we have data on 10 variables described in Table 1.

The variable `lifepres` assumes only two distinct values, 0 if died within 6 months, or 11.6 years otherwise. We define the 6-month mortality status `mort` as a binary variable, 1 if died within 6 months and 0 otherwise.

Mean baseline characteristics by whether they were treated with abciximab (1) or not (0) are summarized in the unweighted portion of Table 2. The unadjusted mortality rates were 5% for patients treated with the usual care and 2% for those treated with abciximab; the unadjusted average costs were $14612.22 and $16126.68 for those treated with the usual care and those treated with abciximab respectively.

From this simple summary, the two groups of patients do not appear to be readily comparable. The abciximab-treated patients seem to be in worse pre-treatment cardiac related conditions than the usual-care patients. For example, the number of patients who had acute myocardial infarction prior to PCI is two times more in the abciximab-treated cohort than in the usual-care cohort. If we look at the outcomes, abciximab-treated cohort has a 6-month mortality rate 3% lower than the usual-care cohort, but with about $1,500 more cost. An interesting question is whether the treatment is cost-effective, after

adjusting for the confounding factors. In the following sections, we will demonstrate our analysis using this data set.

3.2 Analysis

We begin by estimating the PS by using logistic regression with main effects of all the seven covariates in Table 2. The distributions of estimated PS for the two groups, abciximab treated and not treated, are displayed in Figure 2, where we see that the two distributions appear to overlap quite well. We will examine the balance of covariates in the two groups by summarizing the unweighted and the PS-weighted standardized differences for the seven covariates in Table 3, following the recommendations[15,16] of best practice. By Cochran's 0.25-rule, if the standardized difference of a variable exceeds 0.25, the variable is imbalanced between the two groups. The unweighted standardized differences of variables stent, acutemi and ves1proc as shown in Table 2 exceed 0.25 and thus those three variables are imbalanced between the treated and the untreated groups. In contrast, after inverse weighting by propensity score, the standardized differences of all seven covariates are well within 0.25 (Table 2). This indicates the appropriateness of the PS model, in addition to the Hosmer-Lemeshow test (p=0.37) of the goodness of the model fit.

Next we fit two logistic models for the mortality outcome, one for the abciximab treatment group, and one for the non-treated group. Then we use each model in turn to predict the counterfactual responses for *all* patients in the sample – now each patient has two predicted probabilities of mortality if he or she was treated by abciximab, and if he or she was not-treated. We can now employ the formulae (in the Appendix A) to compute

the ACE on mortality and its standard error. For the cost outcome, we follow the same steps, except that we fit two linear regression models instead because this outcome is continuous.

Thanks to a SAS's new procedure, `proc causaltrt` as released in SAS/STAT 14.2, all the above steps are packaged in a compact and user-friendly procedure. We used the procedure to obtain the doubly robust estimates for both outcomes of the 6-month mortality and the cost in Table 3, with standard errors by the robust sandwish formula and alternatively by bootstraping 1000 times. For comparison we also display the results from the IPTW and the outcome regression approach.

From Table 3 we see that all three estimators yield essentially the same results: treatment with abciximab afforded survival advantage by at least 5% while maintaining similar cost to the non-abciximab treated group. For both outcomes, the outcome regression estimator has the smallest SE and the IPW estimator has the largest SE.

The SAS code for this analysis is included in the Appendix C.

4 Discussion

In this paper we describe a doubly robust estimator that combines two causal estimators advantageously, the IPTW and the outcome regression, to offer us two chances to obtain a consistent causal estimate of a treatment effect. As long as we specify one of the two models correctly, the causal estimate is valid (in the sense that its bias diminishes as sample size increases). One can afford to get one of the two models wrong and still

obtain asymptotically unbiased causal estimate, which is certainly more advantageous in terms of robustness in estimation to a single model approach as in either IPTW or outcome regression. We outline the steps for the computation of the DR estimator and its standard error, with the relevant formulae in the appendix. We demonstrate the DR estimation process in contrast with IPTW and outcome regression through a real-life data set using a recently released SAS procedure `causaltrt`. From the results, we observe that although we arrive at the same conclusion by all three estimators, the estimators exhibit different variability. As a matter of fact, these estimators have the following behaviors in large samples. If both the PS model and the outcome regression model are specified correctly, the DR estimator has the smallest variance among all consistent estimators that involve 'inverse weighting'; if the PS is modeled correctly, the DR estimator will have smaller variance than the simple inverse weighted estimator; if the outcome regression is modeled correctly, the DR estimator will likely have larger variance than the regression estimator[9]. The implication of the above theoretical results is that when sample sizes are small, the inversed weighted estimators, IPTW and DR, may not have the precision needed for inference because the associated confidence intervals are wide. In contrast, outcome regression estimators have narrower confidence intervals, but they may be centered on the biased estimates if the outcome regression models are wrong. When sample sizes are large, bias becomes the upmost concern, DR is perhaps the best estimator among the three because it offers some protection against model-misspecification while IPTW and outcome regression do not. Finally, if both the PS and the outcome models are wrong, then DR estimator does not estimate the causal effect of interest and will give a wrong answer. On the other hand, if a non-doubly robust

estimator (IPTW or outcome regression) uses a wrong model, it also will not estimate the

cause effect of interest and the resulting estimate is wrong. Doubly robust estimators

afford protection against model misspecification by offering two chances at getting right

answers (consistent estimates).

References

1. Sekhon J. The Neyman-Rubin Model of Causal Inference and Estimation via Matching Methods. In: Box-Steffensmeier J, Brady H, Collier D, eds. *The Oxford Handbook of Political methodology.* Oxford: Oxford University Press; 2009.
2. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics.* 2005;61:962-972.
3. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology.* 2000;11(5):561-570.
4. Rubin DB. *Matched Sampling for Causal Effects.* Cambridge: Cambridge University Press; 2006.
5. Brookhart MA, Wyss R, Layton JB, Sturmer T. Propensity score methods for confounding control in nonexperimental research. *Circ Cardiovasc Qual Outcomes.* 2013;6(5):604-611.
6. Rosenbaum PR. *Observational studies.* 2nd ed. New York: Springer; 2002.
7. Williamson EJ, Forbes A, Wolfe R. Doubly robust estimators of causal exposure effects with missing data in the outcome, exposure or a confounder. *Stat Med.* 2012;31(30):4382-4400.
8. Li L, Shen C, Li X, Robins JM. On weighting approaches for missing data. *Stat Methods Med Res.* 2013;22(1):14-30.
9. Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Stat Med.* 2004;23(19):2937-2960.
10. Shen C, Li X, Li L. Inverse probability weighting for covariate adjustment in randomized studies. *Stat Med.* 2014;33(4):555-568.
11. Robins J, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association.* 1994;89:846-866.
12. Funk M, Westeich D, Davidian M, Weisen C. Introducing a SAS® macro for doubly robust estimation. 2007; https://www.med.unc.edu/pediatrics/research/harryguess/sas-macro-for-doubly-robust-estimation.
13. Kereiakes DJ, Obenchain RL, Barber BL, et al. Abciximab provides cost-effective survival advantage in high-volume interventional practice. *Am Heart J.* 2000;140(4):603-610.
14. Obenchain B. USPS: Unsupervised and Supervised methods of Propensity Score Adjustment for Bias. R package version 1.2-2. 2012. https://CRAN.R-project.org/package=USPS.
15. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med.* 2015;34(28):3661-3679.
16. Ho D, Imai K, King G, Stuart EA. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis.* 2007;15:199-236.

Variable Name	Definition
lifepres	Mean life years preserved due to survival for at least 6 month following PCI
cardbill	Cardiac related costs incurred within 6 months of patient's initial PCI
abcix	Treatment indicator: 1 for abciximab and 0 for usual-care-alone
stent	Coronary stent deployment: 1 for yes and 0 for no
height	Height in centimeters
female	Gender indicator: 1 for female and 0 for male
diabetic	Diagnosis of diabetes mellitus: 1 for yes and 0 for no
acutemi	Acute myocardial infarction within the seven days prior to PCI: 1 for yes and 0 for no
ejecfrac	The left ventricular ejection fraction
ves1proc	Number of vessels involved in the patient's initial PCI

Table 1: List of Variables in the Lindner Center Data

Variable	control group		abcixib group		Standardized diff (%)
	mean	sd	mean	Sd	
Unweighted					
stent	0.58	0.49	0.70	0.46	25
height	171	11	171	11	0.034
female	0.39	0.49	0.33	0.47	11
diabetic	0.27	0.44	0.20	0.40	15
acutemi	0.06	0.24	0.18	0.38	37
ejecfrac	52	10	50	10	18
ves1proc	1.2	0.48	1.5	0.71	43
Weighted					
stent	0.67	0.47	0.67	0.47	0.63
height	171	12	171	11	1.1
female	0.33	0.47	0.34	0.48	2.2
diabetic	0.24	0.43	0.22	0.42	5.2
acutemi	0.14	0.35	0.14	0.35	0.29
ejecfrac	51	10	51	10	0.049
ves1proc	1.4	0.7	1.4	0.66	6.5

Table 2. Mean, standard deviation and standardized difference between groups, unweighted and weighted.

Method	Outcome	μ_0	μ_1	ACE	Robust SE	Bootstrap SE	Wald 95% CI	Bootstrap Bias Corrected 95% CI
DR	Mortality	0.082	0.015	-0.066	0.025	0.026	(-0.115,-0.018)	(-0.119, -0.016)
	Cost	15744	15995	251.46	1128.8	1110.4	(-1961.0, 2464.0)	(-2205.7, 2146.7)
IPTW	Mortality	0.082	0.015	-0.067	0.029	0.031	(-0.123, -0.011)	(-0.144, -0.018)
	Cost	16009	15981	-27.85	1292.0	1396.1	(-2560.2, 2504.5)	(-3598.2, 2222.8)
REG	Mortality	0.068	0.015	-0.054	0.020	0.021	(-0.093, -0.014)	(-0.097, -0.014)
	Cost	15056	16037	981.6	906.2	915.1	(-794.5, 2756.7)	(-1040.5, 2570.3)

Table 3. μ_0, μ_1 are the averages of counterfactuals, ACE is the treatment effect. The Wald 95% CI is based on the Robust SE.
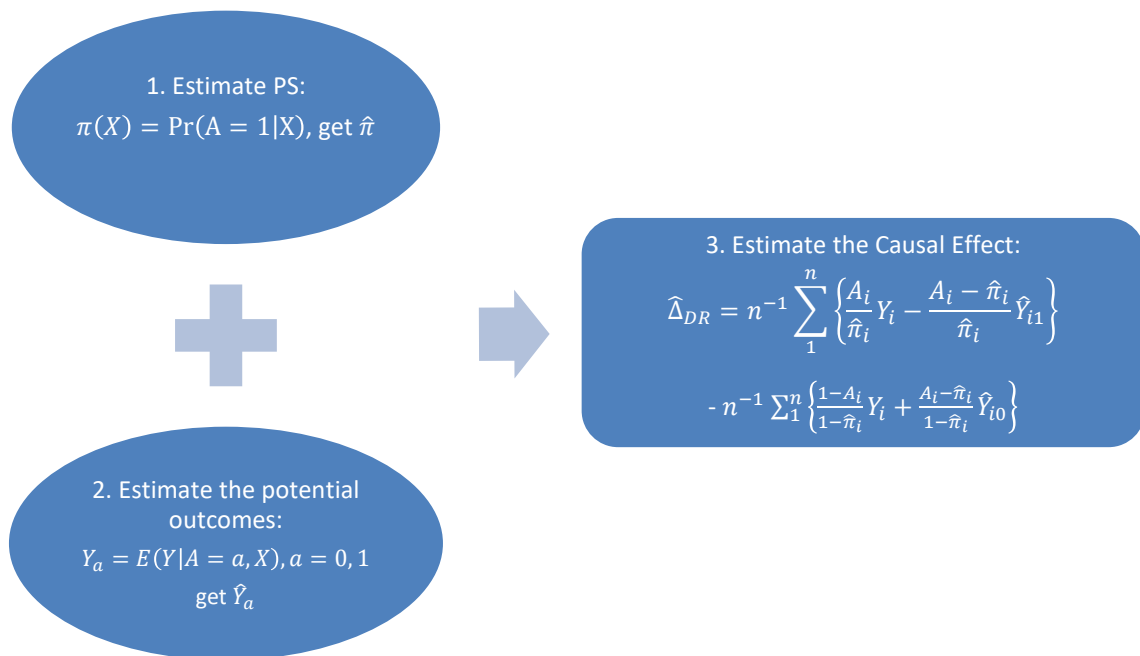
1. Estimate PS:
$\pi(X) = \Pr(A = 1|X)$, get $\hat{\pi}$

2. Estimate the potential outcomes:
$Y_a = E(Y|A = a, X), a = 0, 1$
get $\hat{Y}_a$

3. Estimate the Causal Effect:
$$\hat{\Delta}_{DR} = n^{-1} \sum_1^n \left\{ \frac{A_i}{\hat{\pi}_i} Y_i - \frac{A_i - \hat{\pi}_i}{\hat{\pi}_i} \hat{Y}_{i1} \right\}$$
$$- n^{-1} \sum_1^n \left\{ \frac{1-A_i}{1-\hat{\pi}_i} Y_i + \frac{A_i - \hat{\pi}_i}{1-\hat{\pi}_i} \hat{Y}_{i0} \right\}$$

Figure 1. Flowchart of steps of a DR analysis.
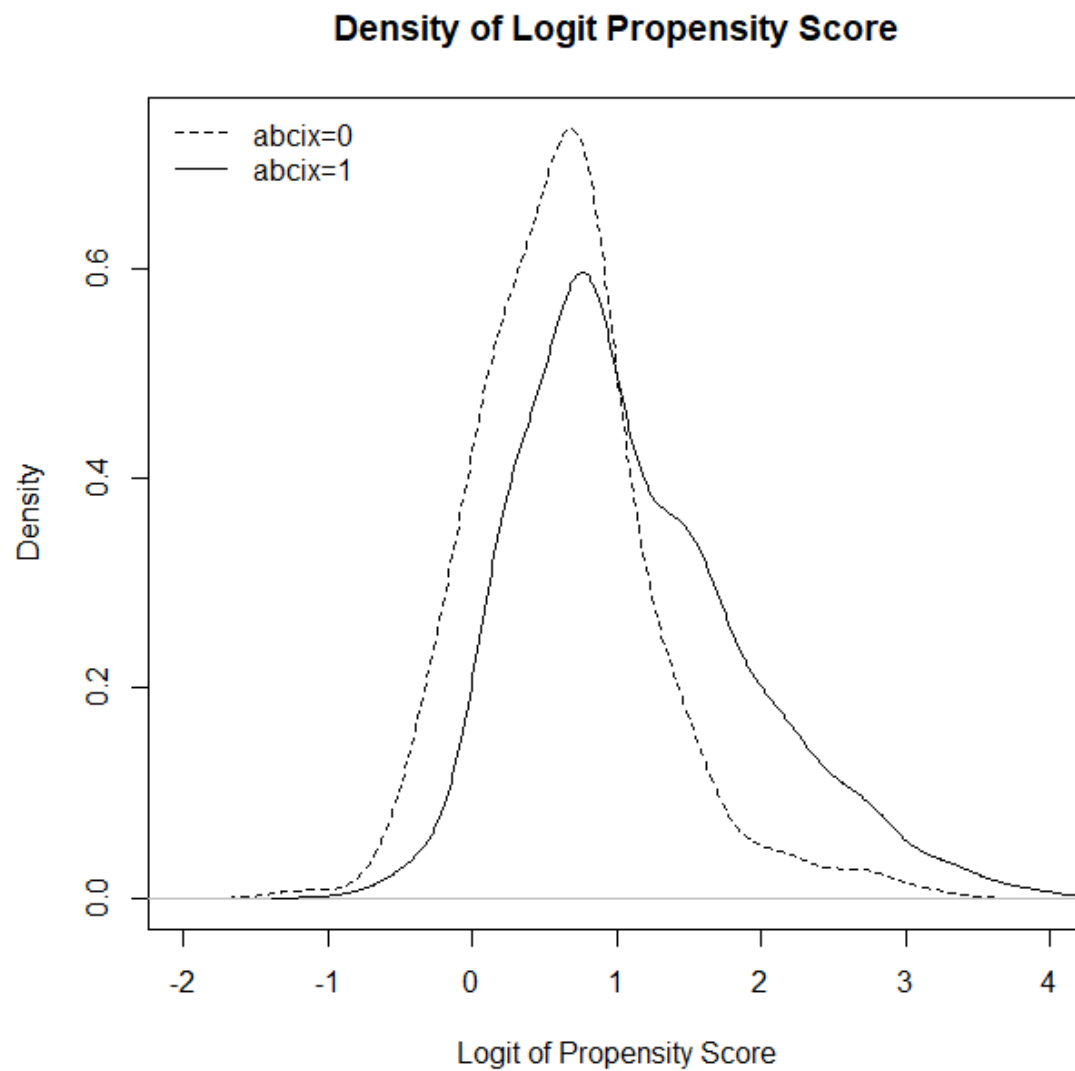
Density of Logit Propensity Score

Figure 2. Distribution of logit of PS for the two groups, one treated with abciximab (abcix=1) and the other not (abcix=0).

Appendix A: Doubly robust estimator for the average treatment effect

Suppose we have i.i.d. data $(Y_i, A_i, X_i), i = 1, \ldots, N$ on N subjects where Y_i indicates the observed outcome, $A_i = 0 \ or \ 1$ indicates the binary treatment, and X_i indicates a p-dimensional vector of pre-treatment covariates. We use $\pi(X) = Pr(A = 1|X)$ to denote the propensity score, the probability of being exposed conditional on the pre-exposure covariate vector X. Also we use Y_{i0} to denote a patient's potential outcome if the patient receives the control and Y_{i1} to denote a patient's potential outcome if the patient receives the treatment. Of course in reality, a patient is assigned to either control or treatment but not both and so only one potential outcome is actually observed. In this paper the parameter of interest is the average treatment effect, $\Delta_{ATE} = E(Y_{i1} - Y_{i0})$.

The DR estimator is defined as:

$$\widehat{\Delta}_{DR} = n^{-1} \sum_{i=1}^{n} \left\{ \frac{A_i}{\widehat{\pi}_i} Y_i - \frac{A_i - \widehat{\pi}_i}{\widehat{\pi}_i} \widehat{E}(Y|A = 1, X_i) \right\}$$

$$- n^{-1} \sum_{i=1}^{n} \left\{ \frac{1 - A_i}{1 - \widehat{\pi}_i} Y_i + \frac{A_i - \widehat{\pi}_i}{1 - \widehat{\pi}_i} \widehat{E}(Y|A = 0, X_i) \right\}$$

$$\equiv n^{-1} \sum_{1}^{n} \widehat{\Delta}_{i,DR}$$

where $\widehat{\pi}_i$ is the estimated propensity score at X_i and $\widehat{E}(Y|A = a, X_i), A = 0,1$, is the estimated potential outcome for a subject with covariate vector X_i under each of the two treatments using outcome regressions. Under the assumptions of the consistency (i.e., $Y_i = A_i Y_{i1} + (1 - A_i) Y_{i0}$), no unmeasured confounders (NUC, $Y_a \perp A|X, a = 0,1$) and positivity ($P(A = a|X) > 0$), $\widehat{\Delta}_{i,DR}$ is the doubly robust estimate of the treatment effect for subject i (that is, the difference in the two counterfactuals for this subject). Note that

taking only the first terms inside the curly brackets from both lines of the formula yields the IPTW estimator. The standard error for the DR estimator is[9],

$$SE_{\widehat{\Delta}_{DR}} = n^{-1} \sqrt{\sum_1^n \left(\widehat{\Delta}_{i,DR} - \widehat{\Delta}_{DR} \right)^2}.$$

Note that the standard error can also be estimated by simply employing the bootstrap method.

Appendix B: Equivalence of outcome regression and PS-based methods

Under the assumptions of the consistency, NUC and positivity (see Appendix A), seemingly different approaches of confounding adjustment are equivalent in the population sense based on the following equalities (subscript i is dropped for clarity in the equations below),

$$E(Y_a) = E[E(Y_a|X)]$$

$$= E[E(Y_a|A = a, X)]$$

$$= E[E(Y|A = a, X)] \quad\quad\quad (B.1)$$

$$= E[E(Y|A = a, \pi(X))] \quad\quad\quad (B.2)$$

$$= E\left[\frac{I(A = a)}{P(A = a|X)}Y\right] \quad\quad\quad (B.3)$$

Equations (B.1), (B.2) and (B.3) say that $E[Y_a]$ can be obtained by regressing observed Y on the treatment indicator A and covariates X (or the PS $\pi(X)$) followed by averaging over X (or $\pi(X)$), or through inverse probability weighting (IPTW). However, regression approaches motivated by Equations (B.1), (B.2) require additional model specification of the response of Y in terms of A and X or $\pi(X)$ compared to the IPTW approach of (B.3).

Furthermore, when X or $\pi(X)$ is discrete, (B.1) or (B.2) motivates a matching (also a stratification) approach through conditioning on the exact values of X or $\pi(X)$. Matching is an intuitive and popular approach to adjust for confounding in the estimation of causal effects. Matching approaches seek to balance variables in X within matched sets between groups as defined by A so as to estimate the effect of A on the outcome Y with minimal bias. Once the matched sets are formed, an average treatment effect is calculated by a weighted average of the differences of average outcomes between the treated and the

control patients within strata with weights proportional to the stratum sizes. When X is

discrete, we can stratify the entire sample (the treated and the control patients) by the

exact values of X. That is, patients with a given value of X form a matched set. In this

case, a saturated regression with strata defined by unique values of X is equivalent to the

matching approach on the exact values of X.

Exact matching on X is often difficult or near impossible, especially when X is high-

dimensional and contains continuous variables. Instead of matching exactly on X, strata

or matched sets can be formed based on some distance measure (a form of dimension

reduction), e.g., Mahalanobis distance, or distance derived from the probability of

receiving treatment, $\pi(X)$.

Appendix C: SAS code for the analysis of Lindner data

```
data lindner (drop = lifepres);
      set dat;
      mort = lifepres = 0;
run;

/* examine covariate means by treatment prior to weighting  */
proc means data=lindner maxdec=2;
      class abcix;
      var mort cardbill stent height female diabetic acutemi ejecfrac
ves1proc;
run;


/* DR for mortality. Plot of logit of PS for 2 treatment arms, and
produce standardized differences pre- and -post weighting.*/
ods graphics on;
proc causaltrt data=lindner method=aipw covdiffps poutcomemod
nthreads=2;
      class abcix mort;
      psmodel abcix (ref='0') = stent height female diabetic acutemi
ejecfrac ves1proc/plots=LPS;
      model mort (event='1') = stent height female diabetic acutemi
ejecfrac ves1proc/dist=bin;
      bootstrap seed=1234 plots=hist(effect);
      output out=ps_weights ipw=ps_weight;
run;

/* DR for cost (cardbill) */
proc causaltrt data=lindner method=aipw covdiffps poutcomemod
nthreads=2;
      class abcix mort;
      psmodel abcix (ref='0') = stent height female diabetic acutemi
ejecfrac ves1proc;
      model cardbill = stent height female diabetic acutemi ejecfrac
ves1proc;
      bootstrap seed=1234 plots=hist(effect);
run;

/* examine covariate mean by treatment post weighting by PS */
proc means data=ps_weights maxdec=2;
      class abcix;
      var mort cardbill stent height female diabetic acutemi ejecfrac
ves1proc;
      weight ps_weight;
run;

/* IPW for mortality */

proc causaltrt data=lindner method=ipw covdiffps;
      class abcix mort;
      psmodel abcix (ref='0') = stent height female diabetic acutemi
ejecfrac ves1proc;
      model mort (event='1') /dist=bin;
      bootstrap seed=1234 plots=hist(effect);
```

```
run;

/* IPW for cost (cardbill) */
proc causaltrt data=lindner method=ipw covdiffps;
      class abcix mort;
      psmodel abcix (ref='0') = stent height female diabetic acutemi
ejecfrac ves1proc;
      model cardbill;
      bootstrap seed=1234 plots=hist(effect);
run;

/* Regression for mortality */

proc causaltrt data=lindner method=regadj;
      class abcix mort;
      psmodel abcix (ref='0');
      model mort (event='1') = stent height female diabetic acutemi
ejecfrac ves1proc/dist=bin;
      bootstrap seed=1234 plots=hist(effect);
run;

/* Regression for cost (cardbill) */
proc causaltrt data=lindner method=regadj;
      class abcix mort;
      psmodel abcix (ref='0');
      model cardbill = stent height female diabetic acutemi ejecfrac
ves1proc;
      bootstrap seed=1234 plots=hist(effect);
run;
```