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,