Behaviour Research and Therapy 98 (2017) 76–90

Contents lists available at ScienceDirect

Behaviour Research and Therapy

journal homepage: www.elsevier.com/locate/brat

A practical guide to propensity score analysis for applied clinical research

Jaehoon Lee*, Todd D. Little

Department of Educational Psychology and Leadership, College of Education, Texas Tech University, United States

ARTICLE INFO

Article history:
Received 22 September 2016
Received in revised form
10 January 2017
Accepted 12 January 2017
Available online 19 January 2017

Keywords:
Propensity score
Matching
Subclassification
Weighting
R

ABSTRACT

Observational studies are often the only viable options in many clinical settings, especially when it is unethical or infeasible to randomly assign participants to different treatment régimes. In such case propensity score (PS) analysis can be applied to accounting for possible selection bias and thereby addressing questions of causal inference. Many PS methods exist, yet few guidelines are available to aid applied researchers in their conduct and evaluation of a PS analysis. In this article we give an overview of available techniques for PS estimation and application, balance diagnostic, treatment effect estimation, and sensitivity assessment, as well as recent advances. We also offer a tutorial that can be used to emulate the steps of PS analysis. Our goal is to provide information that will bring PS analysis within the reach of applied clinical researchers and practitioners.

1. Introduction

The centerpiece of applied clinical research is evidence-based practices, defined as interventions or treatments for which there is consistent scientific evidence of improvement in an outcome (Drake et al., 2001). Modern clinical research often relies on the use of experimental designs to strengthen causal arguments that the improvement among study participants is due to the treatment. In many real-world applications, however, observational studies are the only viable options—especially when it is unethical or infeasible to randomly assign participants to one of two (or more) treatment alternatives (Rosenbaum, 2002, 2010). A complexity arises when the participants who received the treatment and those who did not are considerably dissimilar. If some individual characteristics (e.g., demographics) are associated with treatment status and an outcome, these "covariates" may confound the relation of the treatment to the outcome, thereby obscuring any causal inference about the *true* treatment effect (Rosenbaum, 2005; Rosenbaum & Rubin, 1983).

Propensity score (PS) analysis, as originally proposed by Rosenbaum and Rubin (1983), has been available for more than three decades to aid in the appraisal of causal effects in observational research. Applied clinical researchers have found PS methodology useful in reducing confounding due to unbalanced covariates while examining, for instance, the effect of exposure to paracetamol during fetal life on neurodevelopmental problems (Vlenterie, Wood, Brandlistuen, Roeleveld, van Gelder, & Nordeng, in press); the effect of metformin on gastric cancer risk among patients with type 2 diabetes mellitus (Tseng, 2016); and the effect of a home-based palliative care program on healthcare use and costs (Brian Cassel et al., 2016). A meta-analytic study on coronary artery bypass grafting also showed that observational studies using PS methodology produced results similar to those from randomized trials (Olmos & Govindasamy, 2014). Furthermore, a few researchers have provided systematic reviews on the applications of PS methods in medical research (e.g., Austin, 2008a, 2008b; Weitzen, Lapane, Toledano, Hume, & Mor, 2004), and the advantages and disadvantages of PS analysis are also well documented (e.g., Brooks & Ohsfeldt, 2013; Stuart, 2010).

Although PS analysis has received growing attention in the methodology literature and applied accounts are available elsewhere, recent technical advances are not yet fully incorporated into the substantive literature. This may be, in part, due to the fact that there are few clear guidelines to aid applied researchers in their understanding, use, and evaluation of available PS techniques (e.g., Olmos & Govindasamy, 2015; Randolph, Falbe, Manuel, & Balloun, 2014). Therefore, the aim of this article is to help applied clinical researchers become more familiar with PS analysis by offering a

* Corresponding author.
E-mail addresses: jaehoon.lee@ttu.edu (J. Lee), todd.d.little@ttu.edu (T.D. Little).

http://dx.doi.org/10.1016/j.brat.2017.01.005
0005-7967/

step-by-step guidance on its application through readily accessible statistical software.

The PS methods described in this article include matching (Rosenbaum & Rubin, 1985; Dehejia & Wahba, 1999; Ho, Imai, King, & Stuart, 2007, 2010; Stuart & Rubin, 2008), subclassification (Lunceford & Davidian, 2004; Rosenbaum & Rubin, 1984), weighting (Hirano, Imbens, & Ridder, 2003; Robins, Hernan, & Brumback, 2000), as well as a few recent advances in the methodology literature. Empirical examples are given to demonstrate a comprehensive process of PS analysis: 1) estimating propensity scores; 2) checking balance on the propensity scores and covariates; 3) matching, subclassifying, or weighting the sample; 4) checking balance on the covariates after matching, subclassification, or weighting; 5) estimating the treatment effect; and 6) conducting a sensitivity analysis. Among the few statistical software available for PS analysis we showcase R (R Core Team, 2015), a free and open-source platform because many automated features of R packages facilitate the analysis process. Specifically, *MatchIt* (Ho, Imai, King, & Stuart, 2011) and *twang* (Ridgeway, McCaffrey, & Morral, 2016) will be used for parametric and non-parametric PS estimation, respectively; and *MatchIt* for matching and subclassification. Balance diagnostics will be performed graphically as well as numerically using *MatchIt* and *twang*. Finally, *rbounds* (Keele, 2015) will be utilized to examine sensitivity of a matching application. All the example R codes provided in Table 1 through Table 8 are also accessible online for a download (http://dx.doi.org/10.1016/j.brat.2017.01.005), which would help readers successfully implement a PS analysis by modifying the codes and reproducing various steps of the analysis.

1.1. Propensity scores

Observational studies are vulnerable to selection bias, a situation when individual characteristics (covariates) are related to the likelihood of receiving the treatment, and such relations lead to an inaccurate estimate of the treatment effect (Rosenbaum, 2002, 2010). The PS is the conditional probability quantifying the likely that a study participant is assigned to or selects the treatment given his or her values on the covariates at baseline. PS methodology utilizes this conditional probability to achieve balance on the covariates recreating a situation that would have been expected in a randomized experiment and thereby providing an unbiased effect estimate (Austin, 2011; Rosenbaum & Rubin, 1983).

The PS is calculated (estimated) using observed covariates; balance on the PS produces average balance on the measured covariates (Rosenbaum & Rubin, 1983). Here, the key assumption is that once treatment assignment is conditioned on the measured covariates, there are no unmeasured covariates that confound the relationship between the treatment and outcome, the so-called *strongly ignorable treatment assignment assumption* (Rosenbaum, 2005; Rosenbaum & Rubin, 1983). Unfortunately, this assumption cannot be empirically tested. One can only attempt to make a convincing case that all important covariates have been measured. The inability to balance unmeasured covariates is a major limitation for nearly all observational studies. Neither PS analysis nor traditional regression approach (i.e., covariate adjustment) can directly correct for bias from unmeasured covariates. Therefore, it is important to assess how results from a PS analysis are sensitive to unmeasured covariates (Rosenbaum, 1991a, 1991b). A more thorough overview on the PS and its assumptions can be found in Austin (2011), Harder, Stuart, and Anthony (2010), Shadish and Steiner (2010), and Stuart (2010). Readers more interested in the general issues relevant to design and analysis of observational research are also referred to Rosenbaum (2010).

1.2. Example data

To illustrate PS methods we will use a random subset of empirical data from Veeh, Severson, and Lee (in press). This study tracked adult men and women for four years since they were released from a state prison. While incarcerated, some of them voluntarily participated in a Serious and Violent Offender Reentry Initiative (SVORI) program that is intended to promote a successful reentry of incarcerated persons into communities. Treatment is the reentry program; and outcomes include clinical measures such as urine test results and recidivism measures such as level of recidivism risk, having a new conviction, etc. The program participants (n = 473; treatment group) and the no-program participants (n = 1000; comparison group) were dissimilar at the onset of the study, which could bias the observed treatment effect.

Although R can read data in many different formats, CSV (comma-separated value) format is most convenient to work with R. The R code in Table 1 (1^{st}–2^{nd} lines) shows how to import the example data in a.csv file ('example.csv') into R and save as an R dataset ('dat'). The *file.choose* function allows for interactively choosing a data file in a directory. The *head* function is useful to understand the structure of a dataset as it prints the column (variable) names and the first few rows (6 cases) in the dataset. In Fig. 1 the function output shows that the example dataset ('dat') contains a unique case number for each participant and his or her observations on seven variables.

2. Steps of propensity score analysis

2.1. Step 1. Estimating propensity scores

2.1.1. Covariate selection

The first step of a PS analysis is to decide which covariates should be included in estimating the PS for each participant. Because only a rich set of covariates can make the *strongly ignorable*

Table 1
Estimating propensity scores.

Task	R syntax
Import data into R	> dat < - read.csv(file.choose(), header = TRUE) > head(dat)
Run logistic regression, or	> param < - matchit(trt ~ age + male + white + prison_mo, data = dat) > dat$param_ps < - param$distance > head(dat)
boosted regression	> set.seed(123456) > nonparam < - ps(trt ~ age + male + white + prison_mo, data = dat, n.trees = 5000, interaction.depth = 4, shrinkage = 0.01, stop.method = "es.mean", estimand = "ATT") > dat$nonparm_ps < - nonparam$ps$es.mean > summary(nonparam$gbm.obj) > head(dat)