ELSEVIER

Journal of Clinical Epidemiology 66 (2013) S84−S90

Journal of Clinical Epidemiology

Prognostic score−based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research

Elizabeth A. Stuart[a,b,*], Brian K. Lee[c], Finbarr P. Leacy[d]

[a]*Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, 624 North Broadway, Hampton House, Room 804, Baltimore, MD 21205, USA*

[b]*Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA*

[c]*Department of Epidemiology and Biostatistics, Drexel University School of Public Health, 1505 Race St, Mail Stop 1033, Philadelphia, PA 19102, USA*

[d]*MRC Biostatistics Unit, Cambridge, Institute of Public Health, Robinson Way, Cambridge CB2 0SR, UK*

Accepted 17 January 2013

Abstract

Objective: Examining covariate balance is the prescribed method for determining the degree to which propensity score methods should be successful at reducing bias. This study assessed the performance of various balance measures, including a proposed balance measure based on the prognostic score (similar to a disease risk score), to determine which balance measures best correlate with bias in the treatment effect estimate.

Study Design and Setting: The correlations of multiple common balance measures with bias in the treatment effect estimate produced by weighting by the odds, subclassification on the propensity score, and full matching on the propensity score were calculated. Simulated data were used, based on realistic data settings. Settings included both continuous and binary covariates and continuous covariates only.

Results: The absolute standardized mean difference (ASMD) in prognostic scores, the mean ASMD (in covariates), and the mean *t*-statistic all had high correlations with bias in the effect estimate. Overall, prognostic scores displayed the highest correlations with bias of all the balance measures considered. Prognostic score measure performance was generally not affected by model misspecification, and the prognostic score measure performed well under a variety of scenarios.

Conclusion: Researchers should consider using prognostic score−based balance measures for assessing the performance of propensity score methods for reducing bias in nonexperimental studies.

Keywords: Causal inference; Confounding; Disease risk score; Matching methods; Nonexperimental study; Propensity score diagnostics

1. Introduction

The primary challenge facing nonexperimental studies is a lack of comparability of units in the various treatment conditions. Consequently, differences in outcomes may be due to either the treatments under consideration or to the preexisting differences. Propensity score methods, a key tool of comparative effectiveness research (CER), help address this problem by ensuring that the groups being compared are as similar as possible on the observed characteristics. Moreover, propensity score methods often yield more reliable estimates of treatment effects than traditional methods, such as regression adjustment [1].

Current guidelines advocate examining balance on baseline characteristics−how similar the treatment groups are to one another−to gauge the performance of a propensity score approach [2]. The theory underlying propensity scores implies that under certain distributional settings and outcome models, better balance will yield less bias in treatment effect estimates. However, in practice, there are two challenges to this theory. First, better covariate balance (at least by some measures) does not always yield less biased effect estimates [3]. Second, given the variety of balance measures available, assessing balance is not straightforward, either in terms of the measures for each covariate or in how to summarize across covariates. We also stress that of course a limitation of balance measures is that they cannot assess balance on unobserved confounders and thus cannot help diagnose potential bias because of unobserved confounders.

In this article, we propose a simple new balance measure based on prognostic scores, also known as ''disease risk scores'' when the outcome is binary [4,5]. Fundamentally, this new balance measure attempts to ensure that the groups

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

* Corresponding author. Tel.: 410-502-6222; fax: 410-955-9088.
E-mail address: estuart@jhsph.edu (E.A. Stuart).

0895-4356/$ - see front matter
http://dx.doi.org/10.1016/j.jclinepi.2013.01.013

What is new?

- Balance measures allow assessment of whether a propensity score approach is likely to lead to reduced bias in the treatment effect estimate. It is important to link the calculation of balance measures to how the propensity scores will be used in estimating treatment effects.
- Balance measures based on the expected prognosis under one condition (e.g., control) perform particularly well.
- Comparative effectiveness research studies can use the prognostic score−based balance measure to gauge the success of their propensity score approach.

being compared are similar in their baseline risk for the outcome (their ''prognosis''). Prognostic scores are estimated by modeling the outcome as a function of observed covariates in one of the treatment groups (usually the control or less active treatment condition). Predicted outcomes under that condition are then obtained for everyone in the sample and become the prognostic scores. In this study, we use the standardized difference in mean prognostic scores between treatment groups as a measure of balance.

The incorporation of prognostic scores or related concepts in propensity score diagnostics has been proposed previously [6]. Patrick et al. [7] suggest using information on the strength of the covariate−outcome relationship to inform variable selection in propensity score methods, and the high-dimensional propensity score method of Schneeweiss et al. [8] uses a similar strategy. Prognostic scores, or related ideas, have also been used in matching [5,9,10]. However, the use of prognostic scores as balance measures has not been investigated empirically.

This article presents two simulation studies investigating the correlation of various balance measures with bias in the treatment effect estimate. For simplicity, we adopt the common terms ''treatment'' and ''control'' to refer to the two groups being compared but recognize that in many CER studies these will be two active treatments under comparison.

2. Methods

2.1. Propensity scores

The propensity score is defined as the probability of receiving treatment, given the observed covariates [11]. The properties of the propensity score enable the formation of groups similar to one another on all observed covariates that went into the propensity score, by matching, subclassifying, or weighting using the propensity score [2]. Although propensity scores are often estimated using logistic regression, the diagnostics are not standard model diagnostics [12,13]. The key diagnostic for propensity score methods is the resulting similarity (balance) of the covariate distributions between the treatment and control groups.

In the spirit of separating the design and analysis stages of a study, traditional propensity score methods are conducted without use of the outcome variable. The rationale behind this separation is that a given study design or analysis will not be selected simply because it yields the desired result. This separation also means that one propensity score procedure can be used for multiple outcomes, just as one randomized clinical trial can be used to examine multiple outcomes. However, a consequence of ignoring outcome information during propensity score estimation and use is that propensity score methods prioritize variables by their importance in predicting the treatment received and not the outcome. As a result, variables that are strongly related to treatment assignment but unrelated to outcome, such as an instrumental variable, may have undue influence on the propensity score, which can lead to decreased precision and increased bias [13,14].

2.2. Existing balance measures

Various balance measures have been proposed, with their implementation tied to the propensity score approach by which the data are to be analyzed [15−17]. For example, when weighting is used, balance measures should be calculated using propensity score weights. The most common metric is the absolute standardized bias or absolute standardized mean difference (ASMD). Similar to the effect size, the ASMD is calculated as the absolute value in the difference in means of a covariate across the treatment groups, divided by the standard deviation in the treated group. Guidelines indicate that 0.1 or 0.25 represent reasonable cutoffs for acceptable ASMD's larger standardized biases indicate that groups are too different from one another for reliable comparison [15]. Other common balance measures include the Kolmogorov−Smirnov (K-S) test statistics and *t*-statistics, although caution is urged when using measures that conflate sample size and balance, such as hypothesis tests [16].

2.3. Prognostic scores

The prognostic score, formalized by Hansen [5], is defined as the predicted outcome under the control condition, reflecting baseline ''risk.'' It is estimated by fitting a model of the outcome in the control group and then using that model to obtain predictions of the outcome under the control condition for all individuals. The prognostic score generalizes and extends the unexposed-only disease risk score to continuous, categorical, and ordinal outcomes [4]. In the case of a binary outcome, the prognostic score and the unexposed-only disease risk score are equivalent.