Journal of Clinical Epidemiology 68 (2015) 122–131

Journal of Clinical Epidemiology

REVIEW ARTICLES

Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review

M. Sanni Ali[a,b], Rolf H.H. Groenwold[a,b], Svetlana V. Belitser[a], Wiebe R. Pestman[c], Arno W. Hoes[b], Kit C.B. Roes[b], Anthonius de Boer[a], Olaf H. Klungel[a,b,*]

[a]Division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Faculty of Science, Utrecht University, Utrecht, the Netherlands

[b]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands

[c]Catholic University of Leuven, Research unit for Quantitative Psychology and Individual Differences, Leuven, Belgium

Accepted 1 August 2014; Published online 26 November 2014

Abstract

Objectives: To assess the current practice of propensity score (PS) analysis in the medical literature, particularly the assessment and reporting of balance on confounders.

Study Design and Setting: A PubMed search identified studies using PS methods from December 2011 through May 2012. For each article included in the review, information was extracted on important aspects of the PS such as the type of PS method used, variable selection for PS model, and assessment of balance.

Results: Among 296 articles that were included in the review, variable selection for PS model was explicitly reported in 102 studies (34.4%). Covariate balance was checked and reported in 177 studies (59.8%). *P*-values were the most commonly used statistical tools to report balance (125 of 177, 70.6%). The standardized difference and graphical displays were reported in 45 (25.4%) and 11 (6.2%) articles, respectively. Matching on the PS was the most commonly used approach to control for confounding (68.9%), followed by PS adjustment (20.9%), PS stratification (13.9%), and inverse probability of treatment weighting (IPTW, 7.1%). Balance was more often checked in articles using PS matching and IPTW, 70.6% and 71.4%, respectively.

Conclusion: The execution and reporting of covariate selection and assessment of balance is far from optimal. Recommendations on reporting of PS analysis are provided to allow better appraisal of the validity of PS-based studies.

Keywords: Balance; Confounding; Pharmacoepidemiology; Propensity score; Reporting; Variable selection

1. Introduction

In observational studies, treated and control subjects often differ systematically on prognostic factors leading to treatment-selection bias or confounding in estimating the (adverse) effect of treatment on an outcome. Analytical tools such as the propensity score (PS) methods are applied to correct for such confounding bias. In their seminal article, Rosenbaum and Rubin described the PS as a balancing score: treated and untreated subjects with the same PS tend to have similar distributions of measured confounders given the PS [1]. In other words, assuming that there is no unmeasured confounding and having adequately measured confounders, conditioning on the PS allows one to obtain an unbiased estimate of the average treatment effect at that value of the PS.

PS analysis involves two key steps: deriving the PS from the data and estimating the treatment effect by using the PS to control for confounding. The first step involves an iterative process of fitting a PS model (eg, using logistic regression) on selected covariates until an optimal balance on those covariates is achieved [2]. Despite the growing popularity of PS methods in epidemiology, criteria for selecting variables for a PS model are not well developed compared with variable

Funding: The PROTECT project is supported by Innovative Medicines Initiative (IMI) Joint Undertaking (www.imi.europa.eu) under Grant Agreement no. 115004, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies in kind contribution. In the context of the IMI Joint Undertaking, the Department of Pharmacoepidemiology, Utrecht University, also received a direct financial contribution from Pfizer. The views expressed are those of the authors only and not of their respective institution or company.

Conflict of interest: O.H.K. received unrestricted funding for pharmacoepidemiological research from the Dutch private-public funded top institute Pharma.

* Corresponding author. Tel.: +31 6 288 31 313; fax: +31 30 253 91 66. *E-mail address*: O.H.Klungel@uu.nl (O.H. Klungel).

http://dx.doi.org/10.1016/j.jclinepi.2014.08.011
0895-4356/

What is new?

Key findings

- Balance of confounders between treatment groups is not properly checked and reported in propensity score (PS) analysis.
- *P*-values from significance tests are the most commonly used statistical tools for checking balance.

What this adds to what was known?

- Reporting of PS analysis including assessment of balance of confounders is far from optimal in the medical literature.
- Balance is more often checked in articles using PS matching and inverse probability of treatment weighing.

What is the implication and what should change now?

- The reporting of aspects of PS analysis such as covariate selection and balance assessment should be improved.

selection for conventional outcome models [3,4]. Once the PSs are derived, an intermediate step is using one of the four possible methods: matching, stratification or subclassification, covariate adjustment, and inverse probability of treatment weighting (IPTW) using the PS followed by checking the balance of covariate distribution between treatment groups using appropriate metric [2]. The choice of the PS methods depends on the specific research question, the target population, and inferential goals of the study [5–7], and it affects the way balance on covariates is assessed. Finally, the effect of treatment on the outcome is estimated using one of the PS methods chosen in the previous step.

Although the use of PS methods has shown a dramatic increase in the medical literature [8], previous literature reviews indicated that most authors do not adequately report information on the PS model development [9,10], the balance of covariates between the treatment groups in PS analysis [8,9,11,12], and those who report, often use inappropriate diagnostics [8,9,11]. In addition, researchers often ignore explicit discussion of the PS estimate (estimand) and its relationship with their research question [5]. However, the reviews were limited to PS matching [8,11], and detailed information on the current practice is very limited.

The PS methodology has evolved over the last few years, during which researchers have proposed recommendations on variable selection for PS model [4,13–16] and statistical tools for checking balance and/or selecting the optimal PS models [17–21] and advised against the use of some statistics such as significance testing or prematching C-statistic for evaluating balance and appropriateness of a PS model [17,22–24]. However, the current practice on selecting variables for PS model, choosing a specific PS method as well as measuring and reporting of covariate balance, is not well documented. Therefore, the objective of this study was twofold (1) to systematically review the practice of variable selection and PS model building with emphasis on assessment and reporting of balance when using PS analysis in the medical literature and (2) to provide practical recommendations on the reporting of PS analysis.

2. Methods

We performed a PubMed search to identify studies that used different PS methods. The search was conducted on June 2, 2012, using keywords: "propensity score(s)" or "propensity matching" in all fields (title, abstract, body, or references) identifying 2,317 unduplicated references. To assess the current practice, we limited our search to 6 months (December 2011–May 2012). Articles were excluded if they addressed only methodological or statistical aspects of the PS, if they are unrelated to medical research or published in languages other than English, or if they were reviews, editorials, or letters.

All authors discussed on identifying aspects of the PS analysis on which data had to be collected, but the extraction was conducted by one of the investigators (M.S.A.). From each article included for the review, we extracted information on the type of PS method used, the methods used to estimate the PS, how variables were selected for inclusion in the PS model, whether balance on confounder was checked, methods used for checking balance, and the "appropriateness" of the PS model. When PS matching was used, we recorded information on whether the articles mentioned the matching algorithm applied, the treated/untreated matching ratios used, size of the matched pairs and the starting population, and whether matching was taken into account in the analysis. When stratification on the PS was applied, we extracted information on the quantile of the PS used (deciles, quintiles, quartiles, or tertiles). In addition, information on the impact factor (IF) of the journals [25] and the SCImago Journal Rank (SJR) indicator from Scopus [25–27], a measure of quality of the journals, was extracted for articles included in the review to allow direct comparison of sources in different subject fields. Chi-squared tests were used to compare the frequency of reporting balance assessment and the use of different balance metrics among quintiles of the IF and the SJR of the journals in which the reviewed articles were published.

3. Results

The PubMed search identified 388 articles, of which 92 were excluded: methodological or statistical articles ($n = 20$), articles unrelated to medical research ($n = 63$),