Tips and tricks of the propensity score methods in the thoracic surgery research

Luca Bertolaccini[1], Alessandro Pardolesi[2], Piergiorgio Solli[1,2]

[1]Department of Thoracic Surgery, AUSL Romagna Teaching Hospital, Ravenna, Italy; [2]Department of Thoracic Surgery, AUSL Romagna Teaching Hospital, Forlì, Italy

Correspondence to: Luca Bertolaccini, MD, PhD, FCCP. Department of Thoracic Surgery, AUSL Romagna Teaching Hospital, Viale Vincenzo Randi 5, 48121 Ravenna, Italy. Email: luca.bertolaccini@gmail.com.

Abstract: Without randomization, the differences in the distribution of baseline covariates can confound the evaluation of outcomes between the groups. To estimate the causal effects of treatment on the results randomized clinical trials (RCT) are routinely used. When RCT are not feasible for cost, time, and ethical issues, the effect of treatment on an inevitable outcome could be investigated by using a non-experimental study design. The propensity score (PS) refers to the individual probability, for a subject involved in a study, of receiving a new treatment rather than the control treatment. PS matching is a valuable and increasingly popular tool for dealing with observational data and non-random treatment assignment. Nevertheless, surgeons need to proceed with caution and apply PS methods appropriately.

Keywords: Lung cancer; propensity score analysis; biostatistics; statistics; video assisted thoracic surgery

Submitted Feb 05, 2017. Accepted for publication Feb 27, 2017.
doi: 10.21037/jtd.2017.03.60
View this article at: http://dx.doi.org/10.21037/jtd.2017.03.60

Introduction

The evaluations of therapeutic interventions fall into two categories, observational studies and randomized controlled trials (RCT). The choice of treatment (in the observational studies) may be influenced by the patient characteristics (e.g., higher-risk patients may be more or less likely to receive the intervention). When considering the effect of intervention in RCT, confounding from both measured and unmeasured variables is avoided, and RCT is thus generally considered the highest form of scientific investigation. Nonetheless, accurate treatment effect estimates from observational databases can provide corresponding value to RCT (1). Also, when RCT are longitudinal and when subjects drop out, do not adhere to assigned treatment, or receive post randomization ancillary treatments or exposures, the substantial benefits of randomization dissipate. Without randomization, the differences in the distribution of baseline covariates (treated versus untreated subjects) can confound the evaluation of outcomes between the treatment groups. To estimate the causal effects of treatment on the results observational studies are routinely used. Hence, results cannot be compared directly between treatment groups. With large databases, proper analyses of observational data are becoming increasingly important. When RCT are not feasible for reasons such as cost, time, and ethical issues, the effect of treatment on an inevitable outcome could be investigated by using a non-experimental study design. However, in observational studies, treatment selection is influenced by patient baseline characteristics. In the absence of random treatment assignment, systematic differences in baseline characteristics between treatment groups may exist, leading to noncomparability between the groups, which is known as confounding bias (1). The propensity score (PS) theory was developed by Rosenbaum and Rubin and refers to the individual probability, for a subject involved in a study, of receiving a new treatment rather than the control treatment (2). Whatever the treatment received, items with close values of PS tend to have the same distribution of observed covariates. In the absence of unmeasured confounders, the treatment allocation and the potential outcomes are independent conditionally to the covariates. Moreover, the issue and the treatment assignment are also conditionally independent given the PS in the absence of unmeasured confounders. Consequently, the difference in outcome between the intervention and control proportions (for binary outcomes) is an unbiased estimate of the treatment effect conditionally on PS (3).

Overview of PS methods

PS methods are gradually being used in observational studies as an alternative to conventional covariate adjustment. Since the PS summarizes patient characteristics into a single covariate, it reduces (although does not eliminate) the potential for overfitting. Thus, the purpose of PS is to attenuate problems of features confusing and task to an intervention typically found in observational studies (1). Popular PS methods include the stratification, the matching, the inverse probability weighting, and the use of the PS as a covariate in a conventional regression model. PS stratification separates the dataset into several strata by the individual's PS alone, without reference to treatment group. A treatment effect is then assessed within each stratum, and an overall estimated treatment effect is calculated by taking an average across levels. An alternative method is to split the range of possible PS into similar parts, which results in fewer individuals in the more extreme levels. Stratification has the additional advantage that effect estimates are available for each layer, which may reveal potential heterogeneity of treatment effects across levels (1).

PS matching tries to find individuals with similar PS in the treatment and control groups. There are various methods to match people. Following matching, the treatment effect is calculated by applying either a conventional unmatched regression model or a matched pair analysis of the set of patients who are successfully matched. The matching process results in an analysis based on only those patients who are successfully matched. Therefore, if the treatment effect differs according to patients' characteristics and their likelihood of treatment, the treatment effect estimated from this subsection of patients may vary from the force in the original study population (1).

Inverse probability weighting uses the entire data set but reweights individuals to increase the weights of those who received accidental exposures. This procedure can be thought of as creating additional observations for those parts of the target population of which there were few data (1).

PS analysis step by step

A statistical analysis using PS has four main stages. First, the PS must be estimated. Second, the data need to be matched or grouped based on the estimated PS. Third, a balance must be assessed to ensure that the grouping produced similar pools of patients receiving both treatments. Finally, data can be analysed to estimate the treatment effect size and its clinical and statistical significance. The first three steps are used to frame a comparison around similar groups of patients, and they must be performed without looking at the outcomes data. However, two fundamental assumptions must be met for propensity matching to provide useful results. The most important prerequisite is that, given the covariates, the treatment assignment is independent of outcomes under the two treatment scenarios. In other words, the observed covariates contain all the information about the conditions relevant to potential issues. If the goal is to compare similar groups of patients receiving different treatments, all the factors that determine whether patients are comparable at the time of treatment allocation should be known.

The second condition is that, given the covariates, the patient needs to have a positive probability of receiving both treatments. Intuitively, this situation can be understood for example that there is no gain in asking what the potential benefits of surgery for a patient whose comorbidities preclude survival of an operation (4).

Report of the performance status analysis

PS methods are invaluable tools. However, like regression analysis, the quality of the results obtained depends on appropriate conduct using the consecutive steps. For a critical appraisal of a PS based study, the reader of papers should rely on the information provided. Nevertheless, despite substantial developments and standard applications of PS methods, reporting of aspects of the PS analysis is sometimes inconsistent, and this could be due in part to a lack of standards for conduct and report PS methods (5). Unlike p-values, the standardised difference is not confounded with sample size, and consequently, poise in the initial sample can be likened with that in the matched sample. It can also be used to compare the relative balance of variables measured. Some studies that use PS compare characteristics of matched with those of unmatched treated subjects. This comparison can provide useful information on differences between the treated patients for estimating the treatment effect and between the treated patients excluded from these analyses; in additions, it can provide useful clinical information and information on the generalizability of the results (6). PS methods are primarily aimed to balance treatment groups on covariate distributions, and it is relatively easy to detect and communicate by using simple statistics or plots (5).

Strengths and limitations of PS methods

PS methods, unlike regression techniques, can also warn investigators that a dataset cannot address the causal question without relying on untrustworthy the variables included in the PS model due to small overlap in covariate distributions between treatment groups. Furthermore, sensitivity analyses are invaluable tools to assess the plausibility of the assumptions underlying the PS methods and how violations of them might affect the conclusions. An additional limitation of PS methods is that they work better in large samples because the distributional balance achieved on measured covariates is an expected balance. Thus, in smaller studies, an imbalance of covariates is inevitable even if the PS model is correctly specified. Therefore, surgeons attempting to answer causal questions with the use of observational studies should explore large datasets with consistent qualities (5). Another limitation lies in the impossibility of capturing unobserved individual and contextual confounders. In fact, through control of a given set of related observed covariates, treatment status is supposed to be independent of potential outcomes (7).

And lastly, the selection of an equal number of exposed and unexposed subjects within PS enables the inclusion of exposed subjects without an exact unexposed match and may introduce bias from non-overlapping ranges for exposed and unexposed subjects at the extremes of the distribution of the PS. This bias can be circumvented by restricting analyses to the range of PS standard to both exposed and unexposed patients; plotting the PS distribution is an easy diagnostic for non-overlap (8).

Further warnings in PS methods

To correctly use PS methods, surgeons need to proceed with thoughtfulness. First, to reduce clear selection bias from observed covariates, all observed covariates related to both the action assignment and outcomes, should be considered in PS estimation models. It is indispensable to present comprehensive information on covariate selection to justify the inclusion of covariates in PS estimation models. It is also desirable to conduct sensitivity analysis to test the model sensitivity to hidden bias from potential unobserved covariates. To advance the efficiency of creating a matched control group, a significant sample size should be used for increasing shared support between the treatment and control groups. Finally, a set of covariates selected evaluating relationships with the treatment project and the outcomes will help to weak the influence of unwanted covariates on the estimation of treatment effects.

Conclusions

PS methods are a real statistical instrument for reducing selection bias in observational and non-RCT data. Because of the practical or ethical fences to conducting RCT, applying PS approaches to observational and non-RCT data, such as the electronic medical records, is an effective alternative to using RCT data to approximation treatment effects. PS methods are widely used by in a variety of disciplines. Not only is the PS approach humble and straightforward, but methods of addressing unobserved selection would require additional data (7). PS matching is a valuable and increasingly popular tool for dealing with observational data and non-random treatment assignment. Although here we have focused on the simplest case of a 2-level exposure variable, methods exist for continuous exposures and exposures with many levels. PS matching methods have been described in instructive detail and are now routinely incorporated into statistical software packages, increasing their ease of use (9-11). To obtain valid treatment effect estimates from observational and non-RCT data, surgeons need to proceed with caution and apply PS methods appropriately.

Acknowledgements

None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Elze MC, Gregson J, Baber U, et al. Comparison of

Propensity Score Methods and Covariate Adjustment: Evaluation in 4 Cardiovascular Studies. J Am Coll Cardiol 2017;69:345-57.

2. Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. Annu Rev Public Health 2000;21:121-45.
3. Leyrat C, Caille A, Donner A, et al. Propensity score methods for estimating relative risks in cluster randomized trials with low-incidence binary outcomes and selection bias. Stat Med 2014;33:3556-75.
4. McMurry TL, Hu Y, Blackstone EH, et al. Propensity scores: Methods, considerations, and applications in the Journal of Thoracic and Cardiovascular Surgery. J Thorac Cardiovasc Surg 2015;150:14-9.
5. Ali MS, Groenwold RH, Klungel OH. Best (but oft-forgotten) practices: propensity score methods in clinical nutrition research. Am J Clin Nutr 2016;104:247-58.
6. Austin PC. Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: a systematic review and suggestions for improvement. J Thorac Cardiovasc Surg 2007;134:1128-35.
7. Zhou X, Xie YU. Propensity Score-Based Methods versus MTE-Based Methods in Causal Inference: Identification, Estimation, and Application. Sociol Methods Res 2016;45:3-40.
8. Stürmer T, Joshi M, Glynn RJ, et al. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. J Clin Epidemiol 2006;59:437-47.
9. Winger DG, Nason KS. Propensity-score analysis in thoracic surgery: When, why, and an introduction to how. J Thorac Cardiovasc Surg 2016;151:1484-7.
10. Ali MS, Groenwold RH, Belitser SV, et al. Methodological comparison of marginal structural model, time-varying Cox regression, and propensity score methods: the example of antidepressant use and the risk of hip fracture. Pharmacoepidemiol Drug Saf 2016;25 Suppl 1:114-21.
11. Pan W, Bai H. Propensity Score Methods in Nursing Research: Take Advantage of Them but Proceed With Caution. Nurs Res 2016;65:421-4.

Cite this article as: Bertolaccini L, Pardolesi A, Solli P. Tips and tricks of the propensity score methods in the thoracic surgery research. J Thorac Dis 2017;9(4):920-923. doi: 10.21037/jtd.2017.03.60