Propensity-score analysis in thoracic surgery: When, why, and an introduction to how

Daniel G. Winger, MS, and Katie S. Nason, MD, MPH

ABSTRACT

Propensity score matching is a valuable tool for dealing with observational data and nonrandom treatment assignment, which often results in groups that differ systematically in numerous measured and unmeasured variables. When these systematically different variables are associated with both group assignment and the outcome(s) of interest, bias is introduced. Propensity score matching assigns a propensity for group assignment, which is then used to create 2 groups that are balanced across all possible variables that might influence exposure assignment. When used in the proper conditions, these analytics allow for more accurate and precise estimates of risk for a variety of outcomes. (J Thorac Cardiovasc Surg 2016;151:1484-7)

Central Message

Propensity score matching balances covariates associated with both exposure and outcome, enabling more accurate risk estimates.

Perspective

Analysis of observational datasets without a recognition of systematic differences in covariates influencing both group assignment and outcome is problematic. Propensity score matching balances these covariates, minimizing bias and improving estimates of risk. Thoracic surgeons should develop a thorough understanding of the strengths, weaknesses, and factors governing the use of propensity score matching.

With the advent of electronic records storage, investigators have increasingly greater access to a wide range of variables for many patients over extended periods. These databases provide excellent opportunities for cardiothoracic surgeons to evaluate risk factors and identify predictors for adverse outcomes, treatment response, recurrence, and survival for various treatments and disease processes.

Unfortunately, given the observational nature of these datasets, assignment to an “exposure” of interest, such as treatment or a specific postoperative outcome, is not random and results in groups that often differ systematically in numerous measured and unmeasured variables. These systematically different variables are often associated with both the exposure and the “outcome” of interest; for example, age and comorbidities often influence decisions regarding treatment and are also associated with survival.

To overcome these inherent limitations of observational datasets, propensity score analysis and propensity score matching have been developed. When used in the proper conditions, these analytics allow for more accurate and precise estimates of risk for a variety of outcomes compared with previous analytic approaches. Here we briefly review the factors governing the use of propensity score matching, as well as the strengths and weaknesses of this approach in observational datasets.
WHEN SHOULD PROPENSITY SCORE MATCHING BE USED?

The use of propensity score matching should be considered whenever assignment to an exposure of interest is nonrandom. To illustrate, Zahoor and colleagues were interested in comparing survival outcomes between patients treated with minimally invasive esophagectomy (MIE) and those treated with neoadjuvant therapy followed by MIE. Assignment to the exposure, primary MIE, was assigned nonrandomly, determined by the surgeons and/or oncologists on a case-by-case basis. As such, numerous preexposure variables differed between the groups at the outset. Among these significant differences, patients with primary MIE tended to be older and to have a greater comorbid disease burden. Given that older, sicker patients die sooner than younger, healthier patients, systematic group differences in these pre-exposure variables bias the comparison of survival outcomes against primary MIE. These baseline differences would need to be properly accounted for when estimating the effect of the exposure on the outcome, or else the findings would be biased.

At first glance, this sounds like a job for traditional multivariable analysis. One might think to include any variables that differ between groups in a multivariable Cox regression. However, in this case, 28 pre-exposure variables (representing 36 coefficients, counting categories) were considered; these candidate variables were chosen because they potentially could be related to exposure assignment as well as to the outcome of survival. Such a large number of variables would have overloaded the Cox regression model. Rather than relying on P value–based methods to whittle down the list of pre-exposure variables, propensity score matching can accommodate all of these variables in the propensity matching process and create well-balanced groups for a large number of pre-exposure variables, without having to worry about model misspecification in the propensity model. The propensity score model also can allow for the testing of many interaction terms that may not be feasible in traditional multivariable analysis, and some propensity model methods, such as boosted regression, do not require the user to manually specify all interaction terms.

Unlike multivariable adjustment, propensity score matching also has appealing similarities to the gold standard in clinical research, the randomized controlled trial (RCT). The gist of propensity matching is this: subjects are assigned a “propensity” for having received the exposure; in the simplest case, this is a probability produced from a logistic regression model. Ranging from 0 to 1, there is less overlap in the probability of receiving the exposure at the high and low ends of the probability range and larger overlap in the center values. This probability should be based solely on variables observed before the exposure and can include any variables that might influence exposure assignment. In the Zahoor and colleagues example, an older patient who is otherwise similar to a younger patient would have a higher probability to receive the exposure (primary MIE), because we know that the patients with primary MIE were significantly older. The matching process then matches an exposed subject and an unexposed subject with very similar propensity scores—say, a primary MIE patient who had a 32% chance (based on pretreatment characteristics) of receiving primary MIE, and a neoadjuvant patient who had a 31% chance of receiving primary MIE. In actuality, 1 patient received the exposure and the other did not, so matching them is as if they were randomly assigned to different exposures with equal probability. Matching can be done with a variety of methods and can be 1-to-1 or many-to-1.

Because propensity score matching aims to reduce bias across many potential differences between exposed and unexposed subjects, a postmatching balance assessment across all pre-exposure candidate variables is required to ensure that the matching process has been successful. Rules of thumb have suggested seeking standardized difference values <10% for most or all variables. The balancing of covariates can be graphically represented using covariance balance plots. Examination of the excluded patients is also imperative, to determine whether the process has systematically excluded groups that otherwise would be considered important. If this has occurred, then the analysis of outcomes would not apply to that group of patients. Following postmatching assessment, the desired
When Should Propensity Score Matching Be Avoided or Used with Caution?

Propensity score matching is not a magic bullet. Whereas RCTs involve actual random treatment assignment, which theoretically should balance all possible pretreatment characteristics, even the unmeasured ones, propensity score matching works only with the variables that are collected. Unfortunately, the collected variables are rarely the only variables that could affect treatment assignment, which leaves residual bias. In our example from Zahoor and colleagues, the failure to collect clinical staging data for the dataset would have rendered propensity matching highly suspect, because clinical stage is critical for determining treatment options. 

In summary, propensity score matching is a valuable and increasingly popular tool for dealing with observational data and nonrandom treatment assignment. Although here we have focused on the simplest case of a 2-level exposure variable, methods exist for continuous exposures (eg, different dosages) and exposures with many levels (eg, several treatment options). Basic propensity score matching methods have been described in instructive detail and are now routinely incorporated into statistical software packages, increasing their ease of use. Novel propensity score estimation and matching processes continue to be developed, and other uses of propensity scores, such as inverse probability treatment weighting, can be useful alternatives to matching. For a more detailed overview of the topic, the reader is referred to the article by McMurray and colleagues in the July 2015 issue of the Journal.

Conflict of Interest Statement

Authors have nothing to disclose with regard to commercial support.

References


**Key Words:** propensity score, epidemiologic methods, probability