



Editorial

On the Use of Propensity Score Matching in Biomedicine and Pulmonology



Looking for causality in medical research

A widespread mistake in scientific research is to confuse the concepts of association and causality. Although researchers have been aware of this conceptual difference in recent decades, improper techniques are often used for analysing biomedical data and claim causal relationships.

The fundamental assumption of this discussion is that there is no magic recipe for establishing a causal relationship, and more importantly, there is no causality without a theory of causality. The success of causal analyses depends on three essential ingredients: theory, data, and statistics. First, we need a theory of causality to hypothesize causal mechanisms, the direction of causality paths, and complex relationships among many variables. Then, data quality is critical because some assumptions underlying statistical methods may be untenable. Finally, the selection of proper statistical techniques depends on the assumptions we impose based on theoretical concerns and data availability.

The second critical consideration is that if we could always carry out Randomized Controlled Trials (RCT) with perfect randomization of participants in treatment and control groups, we would not need to consider matching techniques or other methods to control for possible confounders. All medical researchers are well aware that the effect of randomization (having a large sample) is that the average difference in the outcome (e.g. Peak Expiratory Flow Rate) between a group undergoing a treatment (e.g. Chest Physiotherapy) and that of the controls can be attributed to the treatment and is not significantly affected by patients' baseline characteristics. When we have large samples, we can be almost sure that the distribution of possible confounders is similar in both groups. In other words, through randomization, we can be sure that belonging to the treatment or control group is only due to chance and not to other known or unknown factors.

Now, suppose a doctor is treating a group of patients with acute respiratory failure. For example, the former aims to investigate the impact of using a specific portable oxygen tank on long-term damage to patients' lungs. Randomising people to be assigned to receive the oxygen tank or not is unethical. Effectively, those who do not use it may suffer or have long-term damage. Moreover, even if the doctor could try to do that, some individuals allocated to the

treatment group could decline to use the device. In contrast, people designated as controls may claim the right to receive the instrument. Thus, unfortunately, for several reasons, experiments are not always feasible.

The previous example is helpful because some might argue that the doctor could leave patients free to choose. The doctor could exploit those who do not want to utilize the portable oxygen tank as a control group in an experiment. In summary, it would seem that a doctor may have almost an interest in that some people against the device exist. The fundamental issue with this reflection is that “unluckily”, in such a case, the inclusion in the two groups is not the result of randomization but may depend on a myriad of other factors, e.g. general state of health, the severity of the disease, smoking, gender, age, education, and social background. All these variables make it more challenging to establish a causal link between the use of the device and long-term damage to patients' lungs; hence, we are in the field of observational studies.

In RCTs, statistical analyses are much more straightforward, whilst in observational studies, everything becomes more complex if we expect to find causal relationships with a high degree of reliability. Indeed, a RCT is like a safe castle, in which we are protected from enemies and can establish causal relationships with less risk of showing meaningless results. Instead, observational studies open the drawbridge to a series of hazardous enemies, i.e. confounders, spurious correlations, reverse causality, endogeneity, complex mediation and moderations effects, and thus the need for many control variables and more complicated statistical techniques. In observational studies, treatments cannot be randomly assigned. The treated and non-treated groups may have significant differences in their observed covariates, and these differences can lead to biased estimates of treatment effects.¹

For these reasons, the design of non-randomized studies for causal effect estimation and thus causal inference are vital topics in medical statistics and epidemiology today.² There are many formal frameworks for causal inference: structural equation modeling,³ direct acyclic graphs,⁴ and potential outcomes framework.⁵ The following section briefly introduces the basic idea behind the potential outcomes framework and propensity score matching technique in medical research.

Brief introduction on propensity score matching and potential outcome framework in medicine and pulmonology research

There is a “tiny” serious problem in real life and biological studies. After a statistical unit is treated, we can observe its outcome, but it is impossible to know what would have happened if it had been inserted in the controls. Conversely, if we do not treat the subject, we will never know what would have happened by treating her. In other words, in scientific research, we are constantly forced to decide whether an individual will be treated or not, which is a real pity! Effectively, the perfect situation would be to measure the outcome of a subject, both if treated and not treated, to grasp the difference immediately and, therefore, measure the effect due to the treatment. The basic idea of the potential outcome framework is precisely to try to approach such a situation. However, discovering a plausible counterfactual substitute is the essence of all sound causal inference.

Matching is a statistical technique that aims to discover in the group of controls those individuals who are similar to the treated in all relevant pre-treatment features (possible confounders) and can be used as counter-factual(s). There are several methods of matching individuals. The most common are exact covariate matching and propensity score matching (PSM). The former starts from the basic idea that all the combinations of the different modalities of the possible confounders can be used to form blocks. Units in the same block are pretty identical and thus can be compared. The main limitation of this technique is that in real life, the possible confounders are many and often are continuous variables. Therefore the number of possible combinations grows exponentially, giving rise to the so-called curse of dimensionality.

In observational studies, the major problem is that the attribution to the group of treated or controls is not random and could depend on other factors. For example, older people or people with severe breathing difficulties may be more likely to undergo surgery and be included in the treated group due to the severity of the disease and not due to chance. In other words, many possible confounding variables can influence both the outcome and likelihood that a subject is a control or a treatment. The genius insight behind PSM is that we need to check this likelihood. Indeed, the propensity score is the probability of participating in the program (being treated) as a function of the individual's observed characteristics, e.g. estimated via a logistic regression model. The basic idea is that if we compare the outcome of a treated subject with that of an untreated subject (or more than one) who has a very close propensity score (similar characteristics), it is almost as if we had compared the outcomes of the same individual undergoing both the treatment and the non-treatment (potential outcome).

Generally, the causal parameters of interest most frequently used in the literature are the average causal effect of the treatment on the whole population (ATE) and the average causal effect of the treatment on the sub-population of treated (ATT). ATE provides information on the expected impact of the treatment on a unit randomly selected from the population. Instead, ATT delivers knowledge about the expected effect of the treatment on a unit randomly selected from the sub-population of treated. Depending on how to select the estimate of the counterfactual outcome of generic unit i , there are different matching estimators,^{6,7} e.g. k -nearest neighbor (k -NN) with or without replacement, k -NN with caliper or radius, kernel.⁸

Randomized experiments are designed to balance treatment and control groups, often within blocks (i.e. within strata, subclasses or matched pairs) on all covariates. PSM tries to emulate this feature so that any differences between groups can only be attributed to the treatment. In statistical terms, the propensity score, expressed as the conditional probability of being treated

given the observed covariates, can be used to balance the covariates in the two groups and consequently decrease the bias in the estimates of treatment effects.⁹

Main advantages and disadvantages of propensity score matching

PSM makes the comparison of treated and control units more explicit with respect to multiple linear regression (MLR). In general, matching techniques are non-parametric (or semi-parametric if the propensity scores are calculated using a parametric model) and tend to focus attention on the common support condition. The latter aspect is crucial because after estimating the propensity score for all treated and untreated individuals, if there are no possible subjects to be matched as counterfactuals (because they have too different propensity scores), these are not considered in the analysis. Matching does not impose any constraint on the heterogeneity of treatment effects. Instead, MLR restricts the heterogeneity of effects to the assessed interactions incorporated into the model. However, if treatment effects are homogeneous (tough) or we know the proper functional form (also complicated), regression-based estimators are more efficient because they have a lower variance. In addition, differently from MLR, PSM offers tools to assess the quality of balance and overlap, and permits splitting the design and outcome analysis steps.

Nevertheless, PSM is not a miracle device for extracting causal information, but it decrypts the evaluation issue if the underlying conditions are met.

First, even though PSM can balance observed baseline covariates between exposure groups, it cannot deal with unmeasured characteristics and confounders. The limitation of all non-randomized studies relative to RCTs is that the latter reaches balance on all covariates, both observed and unobserved. Instead, in PSM, the so-called unconfoundedness assumption is strictly necessary. If the assumption is not plausible, an instrumental variables regression model is preferred.

Moreover, the sample size and the number of variables in the data play essential roles. In small samples, substantial imbalances of some covariates may be inevitable despite employing a sensibly estimated propensity score.⁹ In RCT, the expected covariate balance is achieved due to randomization. PSM aims to mirror, as closely as possible, the balancing properties of randomisation; however, if the number and quality of predictors used to estimate the propensity scores are low, it is challenging to obtain a satisfactory balance on the covariates. For this reason, simple diagnostics should be adopted to control the balance, and if the latter is not achieved, the propensity scores should be re-estimated, possibly enclosing transformations or interactions among the original covariates.⁹ Hence, the model's specification embraced in the first step to estimate the propensity scores (e.g. logistic regression) plays a vital role.

The possible applications of PSM in pulmonology are numerous. Providing examples would be reductive because every time we deal with observational studies with quality data to evaluate the effect of a treatment or risk factor on exposed and unexposed, PSM is a potential strategy to control the influence of confounders. Effectively, in the field of pneumology, in recent years, many exciting studies exploited PSM to answer causal questions [see e.g.^{10–12}].

References

1. Rutter M. Epidemiological methods to tackle causal questions. *Int J Epidemiol.* 2009;38:3–6.
2. Rubin DB. On principles for modeling propensity scores in medical research. *Pharmacoepidemiol Drug Saf.* 2004;13:855–7.
3. Wright S. The method of path coefficients. *Ann Math Stat.* 1934;5:161–215. <http://dx.doi.org/10.1214/aoms/1177732676>.

4. Pearl J. Causal diagrams for empirical research. *Biometrika*. 1995;82:669–88, <http://dx.doi.org/10.1093/biomet/82.4.669>.
5. Rubin DB. Assignment to treatment group on the basis of a covariate. *J Educ Stat*. 1977;2:1–26, <http://dx.doi.org/10.3102/10769986002001001>.
6. Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. *J Econ Surv*. 2008;22:31–72.
7. Becker SO, Ichino A. Estimation of average treatment effects based on propensity scores. *Stata J*. 2002;2:358–77.
8. Heckman JJ, Ichimura H, Todd PE. Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. *Rev Econ Stud*. 1997;64:605–54.
9. Rubin DB. Propensity score methods. *Am J Ophthalmol*. 2010;149:7–9.
10. Yousufuddin M, Sharma UM, Bhagra S, Murad MH. Hyperlipidaemia and mortality among patients hospitalised with pneumonia: retrospective cohort and propensity score matched study. *BMJ Open Res Res*. 2021;8:e000757.
11. Junjun L, Pei W, Ying Y, Kui S. Prognosis and risk factors in older patients with lung cancer and pulmonary embolism: a propensity score matching analysis. *Sci Rep*. 2020;10:1–8.
12. Crisafulli E, Guerrero M, Ielpo A, Ceccato A, Huerta A, Gabarru's A, et al. Impact of bronchiectasis on outcomes of hospitalized patients with acute exacerbation of chronic obstructive pulmonary disease: a propensity matched analysis. *Sci Rep*. 2018;8:1–12.

Fabrizio Maturo^a, Salvador Cruz Rambaud^{b,*}

^a *Department of Mathematics and Physics, University of Campania
Luigi Vanvitelli, Caserta, CE, Italy*

^b *Department of Economics and Business, University of Almería,
Almería, Spain*

Corresponding author.

E-mail address: scruz@ual.es (S.C. Rambaud).