Editorial Comments

Propensity scores: help or hype?

Wolfgang C. Winkelmayer1 and Tobias Kurth2

1Division of Pharmacoepidemiology and Pharmacoeconomics, and 2Division of Aging Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA

Keywords: bias; epidemiology; ESRD; late referral; nephrologist; propensity scores

In this issue of Nephrology Dialysis Transplantation, Kazmi et al. report an evaluation of the association between late nephrologist referral and mortality in a cohort of incident renal replacement therapy (RRT) patients [1]. After multivariable adjustment, they found that patients who reported having first been seen by a nephrologist < 4 months prior to RRT had a nearly 50% higher risk of 1 year mortality compared to those patients who had their first nephrologist referral earlier in relation to their first RRT [hazards ratio (HR) 1.44; 95% confidence interval (CI): 1.15–1.80]. In addition to standard multivariable regression adjustment, the authors used propensity score (PS) analysis to control for confounding and argued that this approach was a more robust method to balance covariates, and that it helped in their study to overcome confounding and selection bias compared with the traditional approach. However, after adjusting for quintiles of PS, their findings were virtually unchanged (HR = 1.42; 95% CI: 1.12–1.80).

In recent years, PS analyses have become a fashionable tool and its use is increasing particularly in pharmacoepidemiological studies [2]. It seems that lately, some journals and reviewers are in favour of this approach in observational outcomes research. However, it appears that there is much uncertainty among researchers regarding what PS can or cannot accomplish, or in which cases this technique is of no use. It is the purpose of this editorial to shed some light on these issues.

What is the propensity score?

In 1983, Rosenbaum and Rubin introduced PS analysis as an alternative tool to control for confounding [3]. The PS is the probability of receiving treatment, or more general any exposure of interest, for a patient conditional on the patient’s observed pre-treatment covariates. PS analysis is a two-step approach in which a model is first built to predict the exposure (treatment model), and secondly, a model incorporating the information on PS is constructed to evaluate the exposure–outcome association (outcome model). To estimate the PS, usually a logistic regression model is fitted that predicts the exposure and may include a large number of measured pre-treatment covariates. From this model, the summary of each study subject’s pre-treatment covariates yields the expected probability of receiving the treatment or exposure of interest for that individual. This expected probability is the person’s PS. In theory, it is expected that with increasing sample size the pre-treatment covariates are balanced between study subjects from the two exposure groups who have nearly identical PS.

The uses of propensity scores

There are several different options of how PS can then be used to control for confounding. These include regression adjustment in form of including the PS as a covariate or weight into the final outcome model, as well as stratification, or matching based on the PS. Each of these approaches has its advantages and disadvantages.

Kazmi et al. [1] decided to include the PS as covariates (for quintiles of PS) in the outcomes model. In such an approach, the only covariate other than the exposure of interest in the final outcome model is the PS, in this case categorized. One assumes that the association between the categories of the PS and the outcome is modelled appropriately and that no interaction exists between PS and the exposure of
interest. Often, predictive covariates are included in addition to PS in the outcomes model.

Another option is to stratify the overall population based on the PS and then run separate outcome models for each stratum. While technically appropriate, this approach has its limitations, because the sample size is reduced in each stratum, which results in diminished statistical power.

The remaining option is to match individuals from the two exposure groups on their respective PS. This is maybe the most intuitive way to use the PS. As it is important to match on PS as closely as possible, some individuals may be lost which would lead to reduced sample size and power. However, those subjects that could not be matched may constitute extreme observations, and may not reflect typical care situations. If such situations are also strongly associated with the outcome, confounding is avoided. However, if the association between exposure and outcome is different in individuals that cannot be matched (i.e. an effect modification exists) then a potential important exposure effect is ignored.

**The issue of confounding: multivariable model adjustment vs propensity scores**

At this point, the reader may still be wondering about the nature of the bias-reducing mechanism that makes the use of PS so appealing. Consider an illustration of a very simplified concept of confounding (Figure 1). We are interested to obtain an unbiased estimate of the exposure effect is ignored.

First and foremost, it is important to point out that, even though PS can balance observed baseline covariates between exposure groups, they do nothing to balance unmeasured characteristics and confounders. Hence, as with all observational studies and unlike randomized controlled trials, PS analyses have the limitation that remaining unmeasured confounding may still be present. In addition, approaches using the PS do not overcome initial selection bias.

Secondly, one cannot use covariates that may be affected by the exposure of interest in the model that estimates the PS. Kazmi et al. [1] elegantly circumvented this pitfall by excluding variables such as first treatment modality, serum albumin or haemoglobin at initiation of RRT from their estimation of PS. While these factors are associated with the outcome of interest, all-cause mortality, they are likely influenced by earlier nephrologist referral.

**The indications for use of propensity scores**

Why should the PS be used at all if we can accomplish the same goal with traditional multivariable regression modelling? We believe that in most situations, like in the study by Kazmi et al., the use of PS has no apparent advantage compared with traditional methods. In the current example, 255 deaths were observed during the first year of RRT, in which case, following a general rule that one covariate for each 10 outcomes observed can be included in a multivariable model [4], 25 covariates could have been included in that outcomes model. Hence, it is not surprising that the results were not different whether a PS analysis was used or not.

The overall utility of PS as a general analytical tool is rather uncertain for most analyses in which the amount of potential confounding covariates is moderate. Recently, Sturmer and colleagues [5] presented a review of all 25 original papers published in 2002 that used PS. Among the manuscripts that presented both results from traditional multivariable models and from
PS analyses, the authors found that the results between these two techniques were not materially different in most of these studies. Even if the results were different between the two approaches, one could not necessarily assume that the PS analysis yielded the ‘true’ answer, at least not without specifying underlying assumptions.

In general, the approach of using traditional multivariate regression adjustment is preferable if the sample size is sufficiently large and the outcome of interest is not rare. Only if the outcome is rare relative to the number of confounders and the number of study subjects in the smaller exposure group is sufficiently large to warrant multivariable PS estimation, then this statistical technique has a legitimate role to potentially reduce bias and expand the possibilities in observational outcomes research [6]. Only then, the use of PS can be regarded a substantial help, not just hype.

Conflict of interest statement. None declared.

[See related article by Kazmi et al. (this issue, pp. 1808–1814)]

References

6. Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. Am J Epidemiol 2003; 158: 280–287

Stem cell therapy of cardiac disease: an update

Jörg Honold, Birgit Assmus, Ralf Lehman, Andreas M. Zeiher and Stefanie Dimmeler

Molecular Cardiology, Department of Internal Medicine IV, University of Frankfurt, Frankfurt, Germany

Keywords: angiogenesis; cardiac repair; perfusion; stem cells

Despite the development of therapeutic strategies and electrophysiological and surgical treatment regimens, heart failure remains one of the major causes of mortality in the Western world. Heart failure can develop as a consequence of myocardial infarction despite revascularization procedures. Ischaemia-induced death of cardiac myocytes results in scar formation and reduced contractility of the ventricle. Various experimental studies provided evidence that the infusion or injection of stem or progenitor cells may reduce scar formation and fibrosis. Moreover, predominantly bone marrow-derived cells were shown to augment blood flow, thereby providing a novel therapeutic option for the prevention and/or treatment of heart failure.

This review article intends to summarize the current experimental and clinical evidence for a potential role for stem cells in cardiac regeneration.

Which cells?

In animal models, several stem and progenitor cells showed a potential to improve cardiac regeneration (Figure 1). Human embryonic stem (ES) cells have the capacity to generate all cell types of an organism and can give rise to cardiac myocytes [1]. However, at present, there are several challenges to be overcome before ES cells can be clinically applied. One of the most prominent problems is the tumorigenicity of these cells; another challenge is that ES cells may be rejected by the immune system. Although ES cells are by nature more versatile than their adult counterparts, increasing evidence suggests that stem and progenitor cells of the