COMMENTARY

What is a propensity score?

The paper by van Marwijk and colleagues[1] illustrates the application of propensity scores to the analysis of a cluster randomised trial. This commentary outlines the role of propensity scores in the analysis of non-randomised studies and randomised trials.

Propensity scores in non-randomised studies. Consider the example of a population-based register of angina patients. Suppose that a researcher wishes to compare the long-term survival of patients who received coronary artery bypass surgery (CABG) with those who did not receive surgery. Patients selected for CABG can be expected to differ from those that did not receive surgery in terms of important prognostic characteristics including the severity of coronary artery disease or the presence of concurrent conditions, such as diabetes. A simple comparison of the survival of patients who either did or did not receive CABG will be biased by these confounding variables. This 'confounding by indication' is almost invariably present in non-randomised studies of healthcare interventions and is difficult to overcome.

Rosenbaum and Rubin[2] proposed the use of propensity scores as a method for allowing for confounding by indication. Propensity may be defined as an individual's probability of being treated with the intervention of interest given the complete set of all information about that individual.[2] The propensity score provides a single metric that summarises all the information from explanatory variables such as disease severity and comorbity; it estimates the probability of a subject receiving the intervention of interest given his or her clinical status.[3] Individual subjects may have the same or similar propensity scores, yet some will have received the intervention of interest and others will not. For example, women of a certain age, with triple vessel disease and the same comorbidities, may have the same propensity for CABG but only some will receive surgery. An assumption of propensity score analysis is that a fair comparison of treatment outcomes can be made between subjects with similar propensity scores who either did or did not receive the treatment of interest. The propensity score may be estimated for each subject from a logistic regression model in which treatment assignment is the dependent variable. An attractive feature of this approach is that explanatory variables are selected on the basis of their ability to predict exposure to the intervention of interest, their possible associations with outcomes need not be considered.

Three methods are commonly employed to include propensity scores in analyses: matching, stratification, and regression adjustment. Matching requires that each treated individual is matched to an untreated individual with the same or similar propensity score. The process of stratification represents a more general extension of matching in which there is more than one treated or untreated individual per stratum. Once matched pairs or strata have been formed, the association of treatment with outcome is estimated by contrasting outcomes between treated and untreated sets of individuals with similar propensity for treatment. Propensity scores are, however, more commonly included in a regression model as an explanatory variable.

Propensity scores can only balance the observed patient characteristics between treatment groups.[4] Imbalances may remain even after propensity score adjustment if relevant subject characteristics were not measured or were only measured imprecisely. It is also advisable to check that propensity score groups are balanced with respect to patient characteristics, rather than assuming that such balance exists.[5]

Propensity scores in randomised trials. Randomisation is usually considered the optimal method for addressing problems of confounding by indication. However, imbalances in the distribution of subject characteristics between trial arms is especially likely in cluster randomised trials because clusters represent groups of individuals who may share characteristics that differ from subjects in other clusters. Imbalances are also more likely when the number of clusters in a trial is small. Van Marwijk *et al* have implemented a novel application of propensity scores to control for imbalance of individual subject characteristics between the arms of a cluster randomised trial.[1] In their study, subjects who were allocated to the intervention group were less likely to be married, to share a household, or to have higher levels of education or occupation (Table 1[1]). One approach would have been to use the variables in Table 1 to estimate, for each subject, the predicted probability of receiving the trial intervention given the pattern of observed subject characteristics. This propensity score could then be used to adjust analyses in which outcomes were compared between intervention groups. In the Discussion, the report suggests an analysis in which pairs of GPs are matched for propensity. This suggestion raises a question concerning how the levels of the individual subject and the cluster should be considered in the estimation and application of propensity scores.

Jennifer Nicholas,
Research Assistant, Department of Public Health Sciences, King's College London.

Martin C Gulliford,
Professor of Public Health, Department of Public Health Sciences, King's College London, Capital House, 42 Weston St, London SE1 3QD.
Email: martin.gulliford@kcl.ac.uk

REFERENCE

1. Van Marwijk HWJ, Ader H, de Haan M, Beekman A. Primary care management of major depression in patients aged ≥55 years: outcome of a randomised clinical trial. *Br J Gen Pract* 2008; **58:** 680–687.

2. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70:** 41–55.

3. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc* 1984; **79:** 516–524.

4. Braitman LE, Rosenbaum PR. Rare outcomes, common treatments: analytic strategies using propensity scores. *Ann Intern Med* 2002; **137:** 693–695.

5. Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Stat Med* 2007; **26:** 20–36.