ABSTRACT

The main advantage of the randomized controlled trial (RCT) is the random assignment of treatment that prevents selection by prognosis. Nevertheless, only few RCTs can be performed given their high cost and the difficulties in conducting such studies. Therefore, several analytical methods for removing the effects of selection bias in observational studies have been proposed. The first aim of this paper is to compare three of those methods: the multivariable risk adjustment method, the propensity score risk adjustment method, and the instrumental variable method. The second aim is to compare the results from observational studies using the instrumental variable method with those from RCTs aiming to answer the same study question.

INTRODUCTION

The randomized controlled trial (RCT) is considered as the best design for studying the effect of therapy because of its ability to take away ‘selection by prognosis’ [1, 2]. This is the phenomenon that therapy choice is based on variables related to prognosis. For example, physicians prescribe therapy to a patient because they feel this therapy best serves the interests and needs of that particular patient. This selection by prognosis induces selection bias (sometimes called confounding by indication), as patients receiving treatment A cannot be considered similar to those receiving treatment B. In the comparison of treatment outcomes, RCTs solve this problem by allocating treatment to patients in a random and concealed manner. In this way, any difference in outcomes is expected to be due to chance and not due to ‘selection by prognosis’. It is extremely difficult, maybe even impossible, to make causal claims for intended treatment effects from observational studies without randomization. Nevertheless, in the presence of a countless number of research questions related to therapy, only few RCTs can be performed. Apart from high costs, RCTs may be impossible (e.g. ethical objections), inappropriate (e.g. rare adverse events), inadequate (e.g. low generalizability), or unnecessary (e.g. dramatic effect shown in an observational study, like that of insulin in type I diabetes mellitus) [3]. Consequently, the results from observational studies are often used for medical decision-making regarding therapy [4, 5].

Several analytical methods for removing the effects of selection bias in observational studies have been proposed. The first aim of this paper is to compare three of those methods: the multivariable risk adjustment method, the propensity score risk adjustment method, and the instrumental variable method. The second aim is to compare the results from observational studies using the instrumental variable method with those from RCTs.
mortality in haemodialysis patients. Within this study, patients dialysing through native arteriovenous fistulas (AVF) were younger than those using a catheter. Therefore, one could expect that age would confound the association between vascular access and mortality among haemodialysis patients, because patients who receive AVF are younger and because of this, but also for other reasons, may have a better survival. This example illustrates the main problem in investigating the effect of therapy in observational studies: the comparison groups may not be similar because of selection by prognosis. In other words, the decision on the type of vascular access could be based on many variables related to prognosis, like patient preference, medical history, and clinical gut-feeling. As a consequence, the treatment assignment (vascular access) may be linked to the patient characteristics like age and other known and unknown factors.

The conventional modelling approach trying to eliminate selection bias in observational studies is the multivariable risk adjustment method. This method incorporates all known confounders and their potential interactions [7]. The identification of potential confounders has been described extensively elsewhere [8]. In the study by Pisoni et al. [6], the unadjusted Cox regression analysis [9] showed that patients using a catheter as vascular access were 79% more likely to die compared with patients having an AVF (hazard ratio (HR) = 1.79 (95% confidence interval (CI) 1.68–1.92)), whereas after adjustment for age, sex, black race, years with end-stage renal disease (ESRD), comorbidity, and weight, the HR was reduced (adjusted HR = 1.45 (95% CI 1.35–1.56)). Adjustment for confounders made the comparison groups (catheters versus AVF) more similar with respect to these known confounders and for this reason the (smaller) difference in outcome better represented the difference in outcome due to vascular access. In other words, after adjustment the HR came closer to the real effect of vascular access on mortality. Please note that multivariable adjustment only adjusts for confounding variables that have been measured in the study. This means that the results may still, to a significant extent, be affected by unknown confounders that were not measured or measured incorrectly or incompletely (e.g. severity of comorbidity) and by unknown confounders. Both may cause residual confounding.

Another method trying to eliminate selection bias is the use of propensity scores in the model. In a propensity score, information on several potential confounders is combined into one score [10]. Each confounder included in the propensity score should satisfy the criteria for confounding [8]. A propensity score has the advantage of gaining statistical power by using fewer variables in the model and therefore using fewer degrees of freedom. Therefore, the propensity score is most useful in small observational studies or in large studies with very few events [10]. Usually, the logistic regression method is used to obtain a propensity score, with the dependent variable being the probability of the treatment of interest and the independent variables being all potential confounders measured in the study. First, one performs such a model including all individuals in the study. Thereafter, one could calculate the propensity score for each individual, which is the probability of being treated based only on the values of that individual’s confounders. Although it is correct to use a propensity score in a large study, it is not needed [7, 10] as was also shown in the study by Pisoni et al. [6]. Table 1 shows that the adjusted HR derived from the model with adjustment for the propensity score analysis (adjusted HR = 1.46; 95% CI: 1.35–1.57) was virtually identical with that derived from the model with multivariable adjustment (adjusted HR = 1.45; 95% CI: 1.35–1.56). Like in the multivariable risk adjustment analysis, the propensity score analysis has made the comparison groups more similar with respect to confounders and, again, the investigators were only able to adjust for measured confounders.

Table 1. Three analytical methods within observational studies for removing the effects of selection bias in the study of Pisoni (2009)a: multivariable adjustment, the propensity score risk adjustment, and the instrumental variable method

<table>
<thead>
<tr>
<th>Model</th>
<th>Model adjustments</th>
<th>Catheter versus AVFb</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Univariable method (patient level)</td>
<td>1.79 (1.68–1.92)</td>
</tr>
<tr>
<td>B</td>
<td>Adjusted for age, sex, black race, years with ESRD, comorbidity, and weight</td>
<td>1.45 (1.35–1.56)</td>
</tr>
<tr>
<td>C</td>
<td>Adjusted for propensity score (based on the covariates in model B)</td>
<td>1.46 (1.35–1.57)</td>
</tr>
<tr>
<td>D</td>
<td>No adjustment for patient characteristics</td>
<td>1.26 (1.19–1.34)</td>
</tr>
<tr>
<td>E</td>
<td>Adjusted for age, sex, black race, years with ESRD, comorbidity</td>
<td>1.20 (1.12–1.28)</td>
</tr>
</tbody>
</table>

bAVF, arteriovenous fistula.
leaving room for potentially substantial residual confounding due to unmeasured confounders. It should be noted that propensity scores can also be applied by matching or stratification on the propensity score which is described in more detail elsewhere [11, 12]. However, adjustment for confounders or propensity risk scores in the statistical analysis may not be sufficient to take away selection bias, because it is not possible to adjust for confounders that are not measured (or even unknown) and for confounders that have been measured insufficiently. This means that even after adjustment for confounders or propensity scores, it is difficult to ascertain how much of the observed association between vascular access and mortality is really caused by the vascular access itself instead of by other factors. Therefore, in their capacity to remove confounding, propensity scores have the same limitations as multivariable adjustment and are therefore not more likely to remove bias due to unmeasured confounders when strong selection bias exists [7, 10].

**INSTRUMENTAL VARIABLE METHOD**

To a certain extent the instrumental variable method mimics an RCT. Using the instrumental variable method, one chooses a variable—the instrumental variable—which is related to the actual treatment, but at the same time can be considered to be ‘allocated’ to a patient at random, so independent of individual patient characteristics related to prognosis. The random allocation of this variable can be considered as a ‘natural experiment.’ Like an RCT, the instrumental variable method offers the particular advantage of preventing selection bias. As such, also unmeasured confounding is at least partially taken away. Crucial in the instrumental variable approach is that patients are not analysed according to the actual treatment received, but according to the instrumental variable (i.e. a kind of intention-to-treat analysis). This means that the instrumental variable method breaks the link between individual patient allocation and individual treatment prognosis. In the study by Pisoni et al. [6], the investigators chose the case-mix-adjusted percentage of patients with a catheter at a facility as the instrumental variable. As patients tend to visit a treatment facility in the direct neighbourhood of their homes, in principle the facility’s treatment strategy (here expressed as percentage catheter use) can be considered to be allocated (at least partially) at random to a patient and may serve as an instrumental variable. However, being treated by a particular facility may still not be fully at random. Therefore, the facility percentage of patients with a catheter was adjusted for centre case-mix. The instrumental variable method in the study by Pisoni et al. did not examine the vascular access for that individual patient per se, but instead it used the percentage of patients with a catheter at a facility as a patient characteristic for all haemodialysis patients in that facility. The results showed that the HR per 20% increase of patients with a catheter at a facility was 1.26 (95% CI 1.19–1.34) (Table 1). As being treated by a particular facility was still not fully at random (centres differ in characteristics of patients treated), the investigators additionally adjusted for the differences in patient characteristics. After this additional adjustment, the adjusted HR per 20% increase of patients with a catheter at a facility was 1.20 (95% CI 1.12–1.28) (Table 1). There are some problems in the one-to-one comparison of effect estimates for characteristics directly assessed from patients (catheter versus AVF) with those derived from centres (% catheter use in centre). This will be discussed later in this paper. Please note that the instrumental variable method as used in the study by Pisoni et al. [6] is only possible in large multicentre studies like DOPPS or in registries where data on vascular access are available. Moreover, compared with standard modelling, instrumental variable analysis may be more suited to answering policy questions than specific clinical questions [7]. Table 2 provides an overview of the main differences between the instrumental variable method in comparison with the multivariable risk adjustment method and the propensity score adjustment method.

**Assumptions**

It has been argued that the instrumental variable method only mimics an RCT if three assumptions related to the use of the instrumental variable are satisfied [13, 14] (Figure 1):

(i) The instrumental variable must be related to the treatment individually assigned. This assumption may be fulfilled in the study by Pisoni et al. because the centre percentage of catheter use was likely higher in patients using a catheter than in patients dialysing through an AVF. Please note that if the instrumental variable is not strongly related to the actual treatment, one generally gets a dilution of the effect and the instrumental variable may even erroneously show no effect on the outcome.

(ii) The instrumental variable must be unrelated to observed and unobserved prognostic factors. This assumption is at least partially fulfilled: the differences in observed prognostic factors were substantially reduced if patients were compared on the basis of case mix-adjusted centre percentage of catheter instead of on the basis of individual vascular access.

(iii) The instrumental variable must be unrelated to outcome, except through pathways that operate via the treatment individually assigned. We cannot guarantee that the third assumption of the instrumental variable method is met, because the possibility remains that the facility percentage of catheter use may be associated with other facility practice patterns that could contribute to better outcomes (e.g. better quality of care in general).

(iv) There may be difficulties in finding an appropriate instrumental variable satisfying these criteria. Notably, it may not always be possible to verify to what extent these assumptions hold [15].

**Use of instrumental variables in nephrology**

Instrumental variable methods have been used in nephrology when investigating the effect of therapy on the outcome [6, 16–22]. In addition, the method can be used for other
research questions as well. Mendelian randomization enables the estimation of causal relationships in observational studies using genetic variants as instrumental variables. A recently published paper by Chmielewski et al. used the apoE genotype as an instrumental variable to assess the association between cholesterol and mortality in ESRD patients [23]. This is possible because the gene for apoE, which is polymorphic with three common alleles (ɛ2, ɛ3, and ɛ4), affects cholesterol concentration, with ɛ2 being associated with the lowest and ɛ4 associated with the highest cholesterol levels. The results yielded that ɛ2 carriers, associated with low cholesterol, had a slightly decreased risk of cardiovascular mortality and a significantly increased risk of non-cardiovascular mortality. An extensive explanation of Mendelian randomization is provided by Verduijn et al. [24].
estimate depends on the unit of the determinant: the adjusted HR for the instrumental variable will be much lower when choosing an HR per 1% increase of patients with a catheter at a facility instead of 1.26 reported for 20% increase.

Nevertheless, the true relationship of catheter and graft use with the risk of mortality may be more closely reflected by results based on the instrumental variable than those based on the patient-level analysis. The results of the study by Pisoni et al. [6] showed that the patient-level catheter use was strongly associated with many patient characteristics, while, in contrast, facility catheter use was not. In other words, these analyses showed that across the range of the instrumental variables (a facility characteristic), there were no trends in most measured confounders. Such balance suggests that practice is provided independently of the great majority of measured patient factors, and that practice may also be provided independently of most unmeasured confounders [6]. However, using the instrumental variable, some measured and unmeasured patient characteristics may still differ between the facilities. When measured confounders differ between the comparison groups, one could adjust the instrumental variable analysis only for these confounders. This means that using an instrumental variable analysis may better approach the true effect of catheter use than the patient-level analysis, but the results of the instrumental variable analysis may still be subject to selection bias.

**COMPARISON OF STUDY OUTCOMES: INSTRUMENTAL VARIABLE METHOD VERSUS RCTS**

Within the nephrology literature, observational studies using the instrumental variable method and RCTs have investigated the same association. Examples are the relationships between statin use and mortality in haemodialysis patients [17, 25–27], between the glomerular filtration rate at the start of dialysis and mortality [20, 28], and between intravenous vitamin D and mortality in patients on haemodialysis [21, 29, 30].

In the latter example, observational studies using multivariable adjustment found a survival benefit for those patients receiving vitamin D [31–33]. In contrast, Tentori et al. [21] investigated the same association using a case-mix adjusted percentage of patients with vitamin D prescription as instrumental variable and found no significant difference in mortality in patients in facility practices where the vitamin D prescription was more frequent (RR = 0.99; 95% CI 0.94–1.04). The results of the RCTs and meta-analysis confirmed the results of the instrumental variable method, in that the use of vitamin D was indeed not associated with mortality risk [21, 29, 30]. In the other examples on statin use [17] and glomerular filtration rate at the start of dialysis [20], both the patient-level analysis and instrumental variable analysis provided statistically significant associations between the exposure and outcome. In contrast, the RCTs on these topics did not find significant results [25–27, 28]. In these studies [15, 18], the instrumental variable analysis may better approach the true effect than the patient-level analysis, but the results of the instrumental variable analysis may still be subject to selection bias.

However, a comparison of the results of the instrumental variable method and RCTs is not straightforward. The relative risks of the instrumental variable method presented for example by the study of Tentori et al. have a different interpretation than that of the relative risks obtained from RCTs. Finally, it is important to note that as a study design the RCT is still the gold standard to assess the intended effects of treatment. However, regarding the generalizability of RCT results, there may be limitations regarding the time period or the population (inclusion and exclusion criteria) the study was performed in. Also RCTs need to satisfy specific criteria like being large enough so that statistical power is sufficient. On the other hand, well-conducted observational studies may add valuable information.

**CONCLUSION**

The main advantage of the RCT is the random assignment of the intervention that prevents selection by prognosis. However, because of difficulties conducting RCTs, well-designed observational studies are still needed to examine the effect of therapy. Within the observational studies methods exist to reduce confounding by indication. Propensity score risk adjustment has the same limitations as multivariable analysis and is not more likely to remove bias due to unmeasured confounders when selection bias exists. Like an RCT, the instrumental variable method offers the particular advantage of accounting for unmeasured confounders. Therefore, of the three techniques discussed, only instrumental variable analysis may approach the way in which RCTs remove the selection bias.

**CONFLICT OF INTEREST STATEMENT**

None declared.

**REFERENCES**

2. Vandenbroucke JP. When are observational studies as credible as randomised trials? Lancet 2004; 363: 1728–1731
6. Pisoni RL, Arrington CJ, Albert JM et al. Facility hemodialysis vascular access use and mortality in countries participating in


Received for publication: 20.1.2012; Accepted in revised form: 3.5.2012