B. Ollivere
Editor-in-Chief
editor360@boneandjoint.org.uk

The school of hard knocks: trials, registries, and propensity matching

I recently went through a painful experience that will be familiar to many readers of *360*: receiving an email informing me that, "Your paper didn't quite make it on this occasion." This happens to us all and, even as time goes on, such rejections don't become any easier.

This particular paper was one of my students' PhD theses, a randomized trial evaluating rehabilitation in wrist fractures, and the only trial in trauma to evaluate physiotherapy against just a generic advice sheet. The trial fits within a James Lind alliance priority area, a Cochrane identified research need, and a National Institute for Health Research (NIHR) priority area. The paper in question had been submitted to a high-profile British medical journal, and rejected without review. Given the nature of the paper, and due to it fitting within a direct area of research priority – although, admittedly, this was by luck rather than judgement! – I queried the decision with the editor of the journal. After a few emails backwards and forwards, eventually I was told that trials of clinical effectiveness in rehabilitation were not a priority for this journal, and nor were orthopaedic publications that were original research. This was because the journal had recently set a strategy to help maintain their impact factor and, as a result, would be concentrating on big data studies unless the trials were in their area of clinical interest.

Aside from my own personal unhappiness about the individual decision, I have great concerns about the ramifications for trauma and orthopaedics. Our own journals do their utmost to improve the quality of their published work, and to look not only to the United Kingdom but also internationally. However, as a speciality, we are unfortunate enough to be considered 'not exciting' by more general medical journals unless we are reporting tens or hundreds of thousands of patients.

While the arthroplasty and trauma registries have their strengths, they also very much have their weaknesses. Outcomes are limited to those defined as part of the registries data set, usually little in the way of demographic or surgical data is collected, and the data sets themselves are designed for surveillance and audit rather than research. It is true that a big National Joint Registry (NJR) paper is likely to get more citations than a less interesting randomized trial or cohort study – but is it more valid?

The difficulty with all observational studies is they are just that – a description of what happened, constrained by the variables in the data set. The earliest orthopaedic registry papers published just reported unadulterated audit data, perhaps with a little survival data thrown in. However, as the statistics have become more advanced, we are now seeing papers describing data based on multivariable models. Here, the aim is to adjust for confounders. In the most common statistical methods, this is done linearly (such as with Cox proportional hazards models) where the effect of any covariate is assumed to linearly and proportionally contribute to outcome, for example, death in trauma with age. With this kind of model, there is no facility to model a non-linear effect of age, just the average increased risk of death per year lived by the patient. The downsides of this approach are obvious.

What has started to appear (and there are a number of papers reported on in this edition of *360*) is the concept of propensity matching. I thought it was worth touching on what it is, and both what it offers and what it doesn't. As there is a button click function in all the major statistical packages (including R, Stata, and SPSS) offering a version of propensity matching, you can expect more and more papers based on this technique described originally by Rosenbaum and Rubin.[1] The aim of the technique is to allow for the accurate estimation of effect sizes by accounting for the covariate effects in observational data. Unlike other matching techniques, it does not suffer from the dimensionality problem associated with matching. The dimensionality problem refers to the issues with matching intervention patients to control patients when multiple covariates are used. The numbers of patients needed in the matching set rises exponentially with each additional dimension to be matched for. In practice, this limits the clinical applicability of this technique. However, although propensity score matching allows for this problem, it doesn't overcome the major problem of observational data – the unknown confounder.

If there is a confounder that is not recorded in the data set, is as yet unknown to science, or is thought incorrectly to be unimportant to the investigators, then any systems such as propensity score matching cannot account for this confounder. The only way to do so is with a prospective randomized trial, where patients are allocated randomly to groups and known and unknown confounders will be evened out.

An example in this month's issue of *360* would be the sixth study summarized in the knee roundup section.[2] The authors aim to establish whether a 'one-midnight' stay following total knee arthroplasty is as safe as a 'two-midnight' stay based on registry data. While the authors have used propensity score matching to allow for known confounders between the two groups, they do not actually know the reason why the patients did not go home at 'one midnight' in the intervention group. Without knowing this information, there must be unknown confounders, as the groups cannot be matched to account for this. If the reasons were described in the known covariates (such as age or comorbidity), then the analysis is sound. If not, then we have drawn inferences from inherently biased data.

So how does this fit in with the rejection letter I received? In the continued push for high-impact publications, which is faced particularly within the UK academic system, orthopaedic surgeons are being continuously pushed towards publication in general medical journals with higher impact factors. This, in turn, when combined with the stated aim of these journals to publish 'big data' papers, will risk over-analysis of, and over-publication from, registry data. This in itself will limit the quality of the research outputted, with more elaborate statistics applied to observational data sets leaving many questions unanswerable (such as whether in-person physiotherapy benefits patients following wrist fracture).

As a community, the problems that we treat are universal, with wide-ranging estimates; somewhere around 10% of presentations to a family doctor are musculoskeletal in origin. We must continue to research relevant questions, and to get ahead of the modern statistical and data analysis techniques. However, I fear that we face an uphill struggle, with trauma and orthopaedics perceived as a niche speciality. We will increasingly be unable to publish anything other than the most highly citable papers in the major medical journals. We have a responsibility to ensure that we utilize big data sets wisely, as it would be very easy to pull the rug from underneath ourselves in pursuit of that 'big publication' by misrepresenting causal and associative relationships in big data sets.

REFERENCES

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

2. Charpentier PM, Srivastava AK, Zheng H, Ostrander JD, Hughes RE. Readmission rates for one versus two-midnight length of stay for primary total knee arthroplasty: analysis of the Michigan Arthroplasty Registry Collaborative Quality Initiative (MARCQI) database. *J Bone Joint Surg [Am]* 2018;100-A:1757-1764.

 DOI: 10.1302/2048-0105.81.360672