COMMENTARY

In search of the perfect model

Blayne Welk, MD

Department of Surgery, Western University, London, ON, Canada

Administrative data has obvious advantages: it is inexpensive, time-efficient, representative of real-world results and patients, and there are massive amounts of it available in various countries around the world; however, it can have potentially significant limitations when it is used for clinical research. Nayan et al investigated a common inadequacy — lack of adjustment for smoking status and obesity.[1] Numerous studies have identified the importance of these two variables as risk factors for poor outcomes across various disease states. Often, administrative data does not have access to these kind of lifestyle variables; however, surrogates (such as a physician billing code for smoking cessation counselling or operative interventions on morbidly obese patients) or linked household survey results can be used to attempt to address some of this known residual confounding. The thought of prospective data collection to supplement an administrative data study is quite daunting, and thus often avoided if possible. In Nayan's study, the addition of these two covariates made little difference to their multivariable models predicting mortality after kidney cancer.

Residual confounding can exist in an observational study due to known confounders, which were not measured, or unknown confounders, which are not apparent to even the most astute of reviewers. More complex methods of assessing residual confounding from unknown variables are complicated, and few can account for multiple potential unobserved confounders (which is the reality of most administrative data studies).[2]

In this study, the known confounders of smoking and obesity were assessed. One of the reasons why there was only a small difference with the inclusion of these two variables may be because the existing variables, which are commonly available and used in clinical studies, (such as age, gender, and comorbidity score) drive most of the variability in the outcome. Among the various comorbidity scores derived from administrative data, most have a similar statistical performance, and none are perfect due to comorbidities that are inadequately captured with the available data sources.[3] Another explanation for the study's findings is that smoking and obesity are relatively equally distributed between statin users and non-statin users in this patient cohort, contrary to prior studies. In different patient populations, with higher levels of smoking and obesity among statin users relative to non-statin users, the additional adjustment for these factors might be more impactful.

Overall, this study serves to remind us that observational studies are rarely perfect or conclusive, no matter how statistically complicated they may sound. Although investigators may acknowledge weaknesses in the study design (such as residual confounding from unmeasured covariates), these factors do not necessarily invalidate the study conclusions. In this case, the study's null results prevailed, and the point estimates and confidence intervals were essentially unchanged, meaning that smoking and obesity do not appear to confound the relationship between statin use and kidney cancer mortality in this population.

References

1. Nayan M, Hamilton RJ, Finelli A, et al. The value of complementing administrative data with abstracted information on smoking and obesity: A study in kidney cancer. *Can Urol Assoc J* 2017;11:167-71. http://dx.doi.org/10.5489/cuaj.4569
2. Stürmer T, Schneeweiss S, Rothman KJ, et al. Performance of propensity score calibration — a simulation study. *Am J Epidemiol* 2007;165:1110-8. https://doi.org/10.1093/aje/kwm074
3. Schneeweiss S, Maclure M. Use of comorbidity scores for control of confounding in studies using administrative databases. *Int J Epidemiol* 2000;29:891-8. https://doi.org/10.1093/ije/29.5.891

Correspondence: Dr. Blayne Welk, Department of Surgery, Western University, London, ON, Canada; bkwelk@gmail.com