available at www.sciencedirect.com
journal homepage: www.europeanurology.com

European Association of Urology

Platinum Priority – Editorial

Referring to the article published on pp. 877–882 of this issue

Trouble in Paradise: Unmeasured Confounding in Registry-based Studies of Etiologic Factors

Jennifer R. Rider [a,b,*]

[a] *Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts 02118 USA;* [b] *Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115 USA*

In the new era of big data, population scientists using Nordic registries can be regarded as pioneers. Using a unique personal identification number to readily link demographics, health care utilization, and disease outcomes—not only for individuals but also across generations—has offered unparalleled opportunities to pursue public health research. In fact, some have coined the Nordic countries "a paradise for epidemiologists."

As well as describing patterns in disease incidence, mortality, and health care costs, registry linkage studies have been effectively used to address disease etiology. In the cancer realm, Nordic registry studies have helped to clarify the enigmatic origins of stomach cancer by investigating risks associated with duodenal and gastric ulcers [1]; have played an instrumental role in identifying the inverse association between nonsteroidal anti-inflammatory drugs and colorectal cancer [2]; and have helped to uncover the heritability of cancer risk through twin studies [3]. Ultimately, the appeal of these large studies comes down to their ability to estimate more modest causal effects with a greater degree of certainty. But the threat of unmeasured confounding looms large even in big data.

In this issue of *European Urology*, Pottegård et al [4] capitalize on the availability of Danish registry data to test the hypothesis of an inverse association between statin use and renal cell cancer. Smaller previous studies identified positive, null, and inverse associations. With 4606 histologically confirmed cases and 46 060 cancer-free controls, the authors have sample size on their side. When adjusting for covariates available through the registry, primarily other prescription drugs and health conditions, the authors encounter evidence of substantial confounding of the crude association: the unadjusted odds ratio for the association between long-term statin use and renal cancer of 1.48 (95% confidence interval [CI] 1.29–1.69) is attenuated to a statistically nonsignificant 1.06 (95% CI 0.97–1.16) in the multivariable model. The authors conclude there is no evidence of an overall association between long-term statin use and renal cell carcinoma.

But what about the potential impact of unmeasured confounders not readily available through the Danish registries? Pottegård et al are transparent about the fact that two established risk factors for renal cell carcinoma—smoking and obesity—are not accessible through their linkage. If smoking status or obesity influences prescribing patterns for statins, residual confounding could persist. The lack of apparent association in the adjusted estimates should not put us at ease regarding residual confounding, as confounding is a wily foe capable of yielding observed estimates in the direction opposite of the true effect.

Epidemiologists have at their disposal various tools that permit investigation into the impact of unmeasured confounding on observed associations. In general, these approaches require information on the following parameters: (1) the strength of association between unmeasured confounder(s) and the event of interest in the unexposed group; (2) the strength of the association between unmeasured confounder(s) and the exposure in the underlying study population; and (3) the prevalence of unmeasured confounder(s) in the underlying study population. The problem, of course, is that these parameters are usually unknown from the existing data and must be estimated from the literature or external data sources.

DOI of original article: http://dx.doi.org/10.1016/j.eururo.2015.10.020.
* Boston University School of Public Health, 715 Albany Street, T317E, Boston, MA 02118, USA. Tel. +1 617 6385035; Fax: +1 617 6384458.
E-mail address: rider@bu.edu.

http://dx.doi.org/10.1016/j.eururo.2015.10.044

Let us consider one potential confounder, smoking, using an oversimplified example with the following assumptions: (1) overweight and obese individuals are twice as likely to develop renal cell carcinoma as healthy-weight persons (the literature suggests a ~30% increase in risk per 5 kg/m^2 increase in body mass index [5]); and (2) the proportion of overweight/obese individuals is 15% and 35% among those not exposed and exposed, respectively, to long-term statin use. In this scenario, the true relative risk observed could then be 0.90, representing a modest 10% risk reduction associated with long-term statin use [6]. As long as overweight/obesity is more common in long-term statin users and positively associated with renal cancer development, the true relative risk will be lower than 1.06; the magnitude of the bias depends on our specific, untestable assumptions. An analogous argument can be made for obesity. We may then conclude that we cannot rule out a modest inverse association between statin use and renal cell carcinoma, consistent with the previously conducted meta-analysis [7].

Importantly, the real-world consequences of unmeasured confounding are nuanced in ways that render simulations and remedies using external data sources challenging. For instance, confounders do not operate in isolation, and the joint distribution of these factors can exacerbate residual confounding. Moreover, in the real world, exposures are not dichotomized. Accurate characterization of obesity and smoking as potential confounders would require consideration of timing of onset, the intensity and duration of those variables themselves, and with respect to their association with statin use. To further complicate the issue, imperfect measurement of measured confounders can exacerbate the bias from unmeasured confounders in certain scenarios [8]. Handling unmeasured confounding at the analytic stage, particularly when the goal is to detect modest associations, is messy business.

Given that lifestyle factors are rarely captured by population registries but are nonetheless potentially important sources of bias due to unmeasured confounding, careful appraisal of potential confounders and collection of these data at the study design stage represent an alternative solution. When it is not feasible to gather data for the entire cohort, case-cohort or nested case-control study designs can accommodate sampling of more detailed confounder information in a subsample. More sophisticated two-phase case-cohort designs [9] can use the valuable data available for the entire cohort, such as prescription drug usage and comorbid conditions. While these studies may at the outset appear smaller, they may be better suited to address etiologic questions.

In the end, this large study does not bring us any closer to answering the principal question of whether statins influence the development of renal cell cancer. The advantages gained by more case numbers are overcome by the shortcomings of the readily available data for addressing this particular research question. Consequently, the appropriate public health message regarding a potentially modifiable risk factor for kidney cancer remains unclear. As long as an epidemiologist's paradise is uncovering causes of disease, the deliberate and creative use of study design will always be indispensable.

Conflicts of interest: The author has nothing to disclose.

References

[1] Hansson LE, Nyren O, Hsing AW, et al. The risk of stomach cancer in patients with gastric or duodenal ulcer disease. N Engl J Med 1996;335:242–9.
[2] Gridley G, McLaughlin JK, Ekbom A, et al. Incidence of cancer among patients with rheumatoid arthritis. J Natl Cancer Inst 1993;85:307–11.
[3] Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 2000;343:78–85.
[4] Pottegård A, Clark P, Friis S, Hallas J, Lund L. Long-term use of statins and risk of renal cell carcinoma: a population-based case–control study. Eur Urol 2016;69:877–82.
[5] Chow WH, Dong LM, Devesa SS. Epidemiology and risk factors for kidney cancer. Nat Rev Urol 2010;7:245–57.
[6] Schlesselman JJ. Assessing effects of confounding variables. Am J Epidemiol 1978;108:3–8.
[7] Zhang XL, Liu M, Qian J, et al. Statin use and risk of kidney cancer: a meta-analysis of observational studies and randomized trials. Br J Clin Pharmacol 2014;77:458–65.
[8] Fewell Z, Davey Smith G, Sterne JA. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. Am J Epidemiol 2007;166:646–55.
[9] Breslow NE, Lumley T, Ballantyne CM, Chambless LE, Kulich M. Using the whole cohort in the analysis of case-cohort data. Am J Epidemiol 2009;169:1398–405.