Oh Data, Where Are Ye? Using Administrative Databases in Epilepsy Research

We live in the age of electronic health records, databases, and data linkage. Data captured systematically through well conducted health surveys (either omnibus or focused), and those obtained during the health care delivery process (administrative health data), have become a rich source of population based information for clinicians, health services researchers and policy makers. In the past, Canadian epidemiological studies using database linkage were deemed suspect of poor validity, but they are now being published in high impact journals such as JAMA and Lancet[1,2], and in numerous specialty journals.

The report by Schiariti et al in this issue of the CJNS illustrates how large administrative population-based datasets can be harnessed to answer important clinical questions[3]. The authors explored the two-year period prevalence of epilepsy and seizures in children in the province of British Columbia (BC), and assessed the impact of socio-demographic factors on prevalence. The overall period prevalence of childhood epilepsy in BC was 5.5 per 1,000, an all too familiar figure, analogous to those of prior population-based Canadian studies[4-6]. They also confirmed several findings from earlier Canadian epidemiological studies that used diverse methods. For example, prevalence was lower in visible minorities or immigrants (presumably due to the "healthy immigrant" effect), it did not differ by gender or urban or rural place of dwelling, and it was higher in those of lower socioeconomic status and of Aboriginal descent. Some region-specific differences in prevalence were unearthed which could not be disentangled from the effect of socioeconomic status and ethnicity. On the other hand, they found unexpected differences that remain unexplained, such as higher prevalence in infants and preschoolers, and association with socioeconomic status only in the youngest and oldest age groups.

As Schiariti et al point out, the only other Canadian study using administrative data to estimate epilepsy prevalence was Kozyrzky et al's Manitoba pediatric study[6], in which case ascertainment was not only based on ICD coding but also on a drug prescription database. In that study, the prevalence of treated epilepsy in 1998/99 was 4.7 per 1000 children and it increased with age, in keeping with prior international population-based studies that used administrative data[7,8]. Interestingly, many are unaware that the classic epidemiological studies by Hauser et al on the incidence and prevalence of epilepsy in Rochester, Minnesota, also identified epilepsy cases using coded administrative data, followed by chart review to validate the diagnosis[9]. That study also found increasing prevalence rates of epilepsy with age.

Schiariti et al's ability to link large, comprehensive datasets accounts for two of the study's strengths[3]. That is, their large sample size (8,125 children with seizures and 5,621 with epilepsy) and their ability to capture presumably all children receiving health care in the province. This is meaningful because in the Canadian health care system virtually all children with epilepsy are presumed to receive health care, therefore allowing for a near complete population sample. The question of how many children receive "traditional" health care or are not captured by physician reimbursement records is important, but the proportion is probably small.

The strength of Schiariti et al's methods also entails some inherent weaknesses[3]. Specifically, one wonders about the impact on their findings of the unknown accuracy of ICD coding for epilepsy case ascertainment; of the decision to use only codes in the primary diagnostic position; and of the decision not to require more than one epilepsy-coded encounter to establish diagnosis. Although not stated clearly, one presumes that they used only the specific subset of codes pertaining to seizures or epilepsy (780.3 and 333.2). Use of more than these specific subsets would result in overestimates of prevalence. In addition, the use of code V17.2, which does not identify children with seizures, but rather reflects a risk factor for epilepsy, would also over estimate prevalence. Finally, one also presumes that the children were living in BC for the duration of the study period. All of these issues come as no surprise because little attention has been paid in epilepsy to the methodological challenges raised by using administrative databases for research purposes.

In addition to the new and confirmatory epidemiological data uncovered by this study, what have we learned from Schiariti et al's and from similar Canadian studies?[3] We propose four areas of heuristic value.

First, we are beginning to see a picture emerge in the epidemiological canvass of epilepsy in Canada, with contributions from diverse groups and using several methods. Schiariti et al's work adds another stroke to the picture[3]. This is of enormous consequence. The distribution of health care resources, the planning and implementation of specific health services, and the development of research agendas are all informed pre-eminently by these types of data.

Second, it is clear that Canadian epilepsy research is moving well into the era of electronic linkage of health-related databases. We can learn from those that have preceded us. Canada has an excellent track record in research using linkage of health databases in such areas as cardiovascular disease, public health, and health care delivery, among others. We need not re-invent the methodological wheel for epilepsy, or for other emerging areas. Instead, as Schiariti et al did, we need to have stronger and wider collaborations with experts in related areas[3].

Third, we need to do better. We need to do a better job at assessing the validity and accuracy of coded data in administrative databases. As we unravel the epidemiological epilepsy picture, we must pay the highest attention to methodological quality[10]. Studies using ICD-coded data without attention to their validity or accuracy are becoming endemic. It is known that ICD codes are used differently by different individuals, institutions and countries, and their accuracy varies accordingly. Instances exist of excellent but also of useless validity in various conditions. Epilepsy in its many forms is fraught with potential for error. Are we getting a clear picture? Formal exploration should inform better analyses and reports of results.

Fourth, we need to move forward. Without neglecting high quality descriptive studies, we need to move from epidemiological diagnosis to epidemiological intervention. Descriptive Canadian studies are sending coherent messages about frequency, distribution and determinants of epilepsy, as well as areas of greatest and special need. We are learning much about co-morbidity, disparities, health behaviours, and variations in health care practice in epilepsy. The question is what are we doing about it? Here, again, we must learn from precedent efforts and collaborate with experienced individuals and groups from diverse disciplines, including decision-makers. What interventions are most pressing and most likely to succeed, and how are they implemented and measured? A systematic approach to these questions will be one of the frontiers for epilepsy researchers seeking to improve the health of people with epilepsy.

Samuel Wiebe, Nathalie Jette
Calgary, Alberta, Canada

REFERENCES

1. Hackam DG, Thiruchelvam D, Redelmeier DA. Angiotensin-converting enzyme inhibitors and aortic rupture: a population-based case-control study. Lancet. 2006;368:659-65.
2. Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. JAMA. 2007;297:278-85.
3. Schiariti V, Farrell K, Hoube S, Lisonkova S. Period prevalence of Epilepsy in children in BC: a population-based study. Can J Neurol Sci. 2009;36:36-41.
4. Wiebe S, Bellhouse DR, Fallahay C, Eliasziw M. Burden of epilepsy: the Ontario Health Survey. Can J Neurol Sci. 1999;26:263-70.
5. Tellez-Zenteno JF, Pondal-Sordo M, Matijevic S, Wiebe S. National and regional prevalence of self-reported epilepsy in Canada. Epilepsia. 2004;45:1623-9.
6. Kozyrskyj AL, Prasad AN. The burden of seizures in Manitoba children: a population-based study. Can J Neurol Sci. 2004;31:48-52.
7. Holden EW, Thanh NH, Grossman E, Robinson S, Nelson LS, Gunter MJ, et al. Estimating prevalence, incidence, and disease-related mortality for patients with epilepsy in managed care organizations. Epilepsia. 2005;46:311-9.
8. Wallace H, Shorvon S, Tallis R. Age-specific incidence and prevalence rates of treated epilepsy in an unselected population of 2,052,922 and age-specific fertility rates of women with epilepsy. Lancet. 1998;352:1970-3.
9. Hauser WA, Annegers JF, Kurland LT. Prevalence of epilepsy in Rochester, Minnesota: 1940-1980. Epilepsia. 1991;32:429-45.
10. Reid A, Jette NJ, Quan H, Hill M, Wiebe S. Using administrative databases for research in epilepsy: Validation of ICD-9 and ICD-10 epilepsy codes. Can J Neurol Sci. 2008;35(Suppl 1):S49