Editorials

Stroke Tracked by Administrative Coding Data
Is it Fair?

Simona Sacco, MD; Francesca Pistoia, MD, PhD; Antonio Carolei, MD

Community-based registries are considered the best source of information on stroke incidence, risk factors, characteristics, and outcome. They can be used for developing adequate strategies for stroke care and prevention. Criteria to perform an ideal stroke incidence study had been defined and refined in the past 2 decades and include standards for definitions, methods, and data presentation.[1–4] Complying with those criteria makes the obtained results reliable and the studies comparable to one another. However, community-based registries are challenging and are extremely time, labor, and resource demanding.[5,6] Cross-sectional, cohort, and case-control studies, not at the community level, are routinely performed to obtain reliable data on several aspects concerning stroke characteristics and care. They can be as challenging as community-based registries, particularly when high numbers of subjects need to be included to ensure statistical power. Furthermore, those same studies do not reflect disparities across a nation because they are usually performed in geographic regions that more likely involve centers with active teams promoting better stroke care.

Administrative Coding Data: Advantages and Pitfalls

Alternatively, administrative coding data (also called claims data or billing) have been introduced as a tool for stroke research because they are associated with less resource wasting. Nearly all Western countries currently collect hospital discharge data in some form. They are anonymous, plentiful, inexpensive, and widely available in electronic format.[7] In addition, they are quickly retrievable, thereby offering a remarkable advantage in terms of cost and affordability. Furthermore, because stroke is prominently a hospital-managed condition, administrative data might be suitable for monitoring its epidemiology. The advantage is greater when multiple years of data are needed for the analysis, such as when monitoring time trends or assessing the impact of different treatment strategies by using observations before and after the change. Administrative files typically contain information about patients (eg, age, sex, ethnicity, residence), their care (eg, diagnostic examinations and procedures), and hospitalization (eg, admission and discharge dates, length of stay, source of admission, discharge disposition, and charges). The application of administrative coding data to surveillance purposes can, therefore, often be accomplished with relatively modest additional efforts and costs.

Despite the advantages of administrative coding data, they are not flawless, and the degree of their weakness raises the question of the validity of some of the studies that have used them. Users of administrative coding data should recognize that data, particularly diagnosis-related group codes, were not designed for surveillance purposes but rather to justify reimbursement. Each admission is assigned to a diagnosis-related group, and depending on that diagnosis-related group the hospital is paid a predetermined lump sum, regardless of the costs associated with care. In addition, possible intrinsic and unaffordable limitations affect administrative coding data. They do not provide any information on the care for stroke that is borne by health services outside the hospital and by families of affected patients. Data are retrieved after the patient has been discharged by the hospital, so data acquisition is only retrospective. Furthermore, problems related to border crossing may complicate the use of administrative coding data (this is also a well-recognized problem that affects community-based registries). However, despite the already reported possible intrinsic limitations, a question that still requires an answer to legitimate the use of those data concerns the reliability of the recorded diagnoses. For this reason, the concordance between administrative coding data and medical records has been scrutinized in different settings.[8,9]

In the current issue of *Stroke*, Haesebaert et al[10] reported the results of their study designed to help us understand whether concern about the reliability of administrative coding data in stroke research is ill placed or whether such concern is justified or whether the answer lies somewhere in between the 2 poles. The study assessed the accuracy of the French hospital discharge database (HDD) to identify subjects with acute ischemic stroke. The French HDD (Program de Médicalisation des Systèmes d'Information) is a national administrative coding data system, derived from the American diagnosis-related group, established to investigate resource allocation. It gathers data for each patient stay in public or private French hospitals. In detail, the authors compared cases of acute ischemic stroke retrieved by their hospital-based multicenter prospective cohort study conducted in the Rhône area with those retrieved through electronic search in the HDD. In the hospital cohort, 465 cases with a confirmed diagnosis of ischemic stroke were included, whereas from the electronic search in the HDD, the authors identified 329 records with the same primary diagnosis code. Considering the hospital-based cohort as the less biased source for diagnosing ischemic stroke, the use of the HDD led to a false-negative rate (stroke cases not identified by the HDD) of 33% and to a false-positive rate (nonstroke

cases misdiagnosed by the HDD) of 5%. The sensitivity of the HDD (ability to identify ischemic stroke cases among the whole population of ischemic strokes) was 67%, whereas the positive predictive value (proportion of patients with a confirmed ischemic stroke among those identified as ischemic stroke in the HDD) was 95%. The main factors associated with inappropriateness of HDD coding were, referring to the false-negative cases, lack of precision regarding the cause of stroke (I64 code: stroke without indication of ischemic or hemorrhagic etiology) and misdiagnosis of stroke as transient ischemic attack. Regarding the false-positive cases, incorrectly coded as ischemic stroke in the HDD, the actual diagnosis was mostly transient ischemic attack or hemorrhagic stroke. The main limitation of the French HDD, as emerged by this study, was the high proportion of false-negative cases (ie, cases that would have been missed if the diagnosis had been based only on the search in the database). However, positive predictive value was good, reaching a value of 95%.

The accuracy of the French HDD for the diagnosis of stroke had already been evaluated in a study, by Aboa-Eboulé et al,[11] that tested its accuracy versus cases collected in the Dijon Stroke Registry.[11] In that study, the overall sensitivity of the French HDD was higher (77%), whereas the positive predictive value was lower (69%) than the one reported by Haesebaert et al.[10] Furthermore, in the same study,[10] sensitivity and positive predictive values were heterogeneous across stroke types and were higher for intracerebral hemorrhage as suggested by several previous studies.[9,12–15] Aboa-Eboulé et al[11] also found that the overall positive predictive value of the French HDD improved from 2004 up to 2008 by 30%. False-positive cases in that study[11] were mainly because of errors in coding stroke or demographic data, as well as because of misdiagnosis of transient ischemic attack or acute stroke rather than sequel of stroke. False negatives were attributed to coding errors (coding symptoms instead of reporting a stroke diagnosis, reporting stroke as the secondary and not the primary diagnosis, and errors in codification of demographic data [nonresident patients]). A limitation of the studies by Haesebaert et al[10] and Aboa-Eboulé et al[11] is that the premise that the reviewing stroke neurologist's diagnosis (heavily influenced by the diagnosis of the treating physician) represents the gold standard is not entirely accurate.

According to the results of the quoted studies,[10,11] administrative coding data of the French HDD cannot be considered as the sole means of conducting stroke surveillance because the stroke diagnosis reported in administrative coding data is unreliable for any research purpose. Despite that the results are not automatically generalizable to other administrative coding data systems, we think it is convenient to avoid use of any of those systems to track stroke, unless a prior preliminary exploratory study would prove their reliability. Furthermore, although administrative coding data may improve traditional methods, any application of administrative data to stroke research requires a thorough understanding of their limitations and the willingness to interpret results with caution. Indeed, the main problem, as confirmed by the quoted studies,[10,11] relies on the accuracy of the diagnosis, the quality of included data elements, and particularly on missing data elements. Furthermore, we should consider that the personnel who apply coding are not homogenous across all countries (administrative versus physician staff), and it is not always the same person who is in charge of the patient so that the coding needs to be extrapolated from medical records. Moreover, administrative coding data mostly exist for financial resources allocation, and those who code hospital data may be influenced by the need for resources. In countries with restrictive fiscal measures and hospital reduction plans, because coding personnel are aware of the fact that stroke codes are linked to a favorable cost-to-reimbursement ratio, increases in unfair incorrect diagnoses may also occur.

A further point to consider is that stroke research studies use advanced statistical methods, such as for risk adjustment. However, it is wellknown that in administrative coding data, information on race and ethnicity is either missing or unreliably reported because vascular risk factors are lacking or coded as secondary diagnoses. Furthermore, differences and lack of standardization in the definition of risk factors across coders can further affect the quality of the data.[16,17] In some instances, not all the diseases in the same patient receive a code. This may not occur for stroke that is associated with a high reimbursement but may occur for stroke risk factors and comorbidities that are not associated with high reimbursement.

Conclusions

Several new factors could greatly enhance future applications of administrative coding data in research. A periodic data auditing system might improve data quality and consistency as establishing computerized health networks among different health providers for the same patient or use of computerized medical record linkage systems. However, we should consider that concerns about data privacy and rising health care costs might also threaten these databases, especially if the value arising from having the data does not seem to outweigh the risks and costs of data collection.

If provider documentation, coding practice, and reimbursement incentives remain constant over time, the most appropriate use of discharge codes may be to track trends in national estimates of stroke. However, because there are factors that could jeopardize the consistency of these measures, we should consider the need to reappraise the appropriateness of the use of discharge data over time.

Disclosures

None.

References

1. Malmgren R, Warlow C, Bamford J, Sandercock P. Geographical and secular trends in stroke incidence. *Lancet*. 1987;2:1196–1200.
2. Feigin V, Hoorn SV. How to study stroke incidence. *Lancet*. 2004; 363:1920.
3. Sacco S, Stracci F, Cerone D, Ricci S, Carolei A. Epidemiology of stroke in Italy: a review. *Int J Stroke*. 2011;6:219–227.
4. Sudlow CL, Warlow CP. Comparing stroke incidence worldwide: what makes studies comparable? *Stroke*. 1996;27:550–558.
5. Sacco S, Marini C, Olivieri L, Pistoia F, Carolei A. Contribution of hematocrit to early mortality after ischemic stroke. *Eur Neurol*. 2007; 58:233–238.

6. Sacco S, Totaro R, Toni D, Marini C, Cerone D, Carolei A. Incidence, case-fatalities and 10-year survival of subarachnoid hemorrhage in a population-based registry. *Eur Neurol*. 2009;62:155–160.

7. Ferver K, Burton B, Jesilow P. The use of claims data in healthcare research. *The Open Public Health Journal*. 2009;2:11–24.

8. Kokotailo RA, Hill MD. Coding of stroke and stroke risk factors using international classification of diseases, revisions 9 and 10. *Stroke*. 2005;36:1776–1781.

9. Tirschwell DL, Longstreth WT Jr. Validating administrative data in stroke research. *Stroke*. 2002;33:2465–2470.

10. Haesebaert J, Termoz A, Polazzi S, Mouchoux C, Mechtouff L, Derex L, et al. Can hospital discharge databases be used to follow ischemic stroke incidence? *Stroke*. 2013;44:xxx–xxx.

11. Aboa-Eboulé C, Mengue D, Benzenine E, Hommel M, Giroud M, Béjot Y, et al. How accurate is the reporting of stroke in hospital discharge data? A pilot validation study using a population-based stroke registry as control. *J Neurol*. 2013;260:605–613.

12. Ellekjaer H, Holmen J, Krüger O, Terent A. Identification of incident stroke in Norway: hospital discharge data compared with a population-based stroke register. *Stroke*. 1999;30:56–60.

13. Leibson CL, Naessens JM, Brown RD, Whisnant JP. Accuracy of hospital discharge abstracts for identifying stroke. *Stroke*. 1994;25:2348–2355.

14. Stegmayr B, Asplund K. Measuring stroke in the population: quality of routine statistics in comparison with a population-based stroke registry. *Neuroepidemiology*. 1992;11:204–213.

15. Tolonen H, Salomaa V, Torppa J, Sivenius J, Immonen-Räihä P, Lehtonen A; FINSTROKE register. The validation of the Finnish Hospital Discharge Register and Causes of Death Register data on stroke diagnoses. *Eur J Cardiovasc Prev Rehabil*. 2007;14:380–385.

16. Blustein J. The reliability of racial classifications in hospital discharge abstract data. *Am J Public Health*. 1994;84:1018–1021.

17. Schoenman JA, Sutton JP, Elixhauser A, Love D. Understanding and enhancing the value of hospital discharge data. *Med Care Res Rev*. 2007;64:449–468.

Stroke

American
Heart
Association

American
Stroke
Association ®

Stroke Tracked by Administrative Coding Data: Is it Fair?
Simona Sacco, Francesca Pistoia and Antonio Carolei

The online version of this article, along with updated information and services, is located on the
World Wide Web at:
http://stroke.ahajournals.org/content/early/2013/06/04/STROKEAHA.113.001742.citation