Can Hospital Discharge Databases Be Used to Follow Ischemic Stroke Incidence?

Julie Haesebaert, MD, MPH; Anne Termoz, MPH; Stéphanie Polazzi, MPH; Christelle Mouchoux, PharmD, PhD; Laura Mechtouff, MD, MPH; Laurent Derex, MD, PhD; Norbert Nighoghossian, MD, PhD; Anne-Marie Schott, MD, PhD

Background and Purpose—Because acute ischemic strokes (ISs) are mainly hospitalized, hospital discharge data could be used to routinely follow their incidence management. We aimed to assess sensitivity and positive predictive value of the French hospital discharge database (HDD) to identify patients with acute IS using a prospective and exhaustive cohort (AVC69) of acute IS cases.

Methods—A selection algorithm based on IS diagnosis coded with the International Classification of Diseases (ICD-10) and cerebral imaging codes was used to identify all hospital stays with the primary diagnosis of IS in the HDD of the university hospitals of the Rhône area. Cases identified through HDD search were compared with IS cases identified through an exhaustive cohort study conducted in the Rhône district and confirmed on medical records review.

Results—There were 465 confirmed cases of IS hospitalized in 1 of the 4 university hospitals during the study period. The HDD search identified 313 among those (true-positive cases) but missed 152 cases (false-negative cases). The sensitivity of the HDD search was 67.3% (95% confidence interval, 63.1–71.5), and the positive predictive value was 95.1% (95% confidence interval, 92.8–97.4). Additionally, HDD search retrieved 16 cases, which were not eventually IS (false positives). Sensitivity was better when patients were hospitalized in neurological departments.

Conclusions—The lack of sensitivity to identify acute IS patients through HDD search does not seem to be accurate enough to validate the use of these data for incidence estimates. Efforts have to be made to improve the coding quality.

(Stroke. 2013;44:00-00.)

Key Words: hospital discharge database ■ ischemic stroke ■ positive predictive value ■ sensitivity

The door-to-needle time is the current challenge in acute ischemic stroke (IS) management. The 4.5-hour delay after symptom onset required for thrombolysis is difficult to achieve.1–4 In the United States, the thrombolysis rate has doubled between 2004 and 2009 but still remained low at ≈5% of all ISs.5 Delays in management can be reduced through optimal organization of health services for the management of stroke. In France, for instance, a national action plan has been set to globally improve stroke management with a notable increase in the number of stroke units.6,7 Impact of these changes in healthcare organization should be continuously assessed to evaluate the benefit of health policies on the outcome of patients with stroke. Routinely implemented indicators would help to follow improvements in management delays and thrombolysis rate for patients with IS.8

See related article, p XXX.

French national estimates of stroke incidence are currently mainly based on the French national hospital discharge database (HDD) named PMSI (Programme de Médicalisation des Systèmes d’Information). This HDD is the only national source of data on stroke. It routinely records all admissions in private and public hospitals. These data constitute a potential powerful source to be used in epidemiological surveys, notably for hospital-managed diseases such as stroke. If this database could be used to follow stroke incidence with reliability, the identification of IS patients would enable us to study their management and thrombolysis rate. However, this database has potential limitations because its primary aim is to produce medical information for hospital financing, and the appropriateness of coding depends on the willingness of physicians to code its activity. The coding quality, the exhaustivity, and the accuracy of the diagnosis are key elements limiting the reliability of estimates based on these data.9

We aimed to assess the accuracy of the French HDD to follow acute IS incidence. If the HDD could be considered reliable to assess IS incidence, this database could provide data useful to estimate the thrombolysis rate, an indicator of IS management quality.
Methods

Reference Population
We used an exhaustive observational cohort of patients with acute stroke named AVC69. The AVC69 study was a population-based multicentric prospective cohort study conducted in the Rhône area. This observational study aimed to assess the prehospital delays from symptom onset for patients with suspected stroke. Patients were identified by the emergency physician or a neurologist. All consecutive patients with a suspected acute stroke (symptoms onset <24 hours) who called the emergency medical services or admitted to 1 of the 17 emergency departments or to the stroke unit of the Rhône area between November 2006 and June 2007 were included in the study. The final diagnosis of stroke was ascertained by an experienced physician (neurologist or emergency physician) based on clinical criteria and cerebral imaging (computed tomography scan and MRI). We included in the present analysis only patients of the AVC69 cohort with a confirmed diagnosis of acute IS who were hospitalized in one of the emergency departments or stroke unit of the 4 University Hospitals of the Rhône area (1721999 inhabitants in 2011).

Hospital Discharge Database
The French HDD, named PMSI, is a medico-administrative database derived from the American Diagnosis-related Group. This database collects data on all stays of patients in all healthcare facilities using standardized discharge abstracts. Mandatory information on the patients, their diseases, and the procedures performed is collected. Diagnoses are coded by the physicians at the end of the stay using the tenth version of the International Classification of Diseases (ICD-10). The primary diagnosis is the condition that requires the highest use of resources. Secondary diagnoses are comorbidities. The procedures are coded according to the French Common Classification of Medical Procedures. For the present study, a computerized extraction from the HDD was performed. The algorithm we used in this study searched in the HDD all patients presenting with a primary diagnosis of IS according to the ICD-10 codes (I.63; Table 1) and hospitalized between November 2006 and June 2007. To be comparable to AVC69 inclusion criteria, we did not include patients with stroke symptoms above 24 hours (based on medical records analysis), those who did not have cerebral imaging (Table 1), and those <18 years of age.

Analysis

Agreement Between the HDD and the Reference
A comparison of the HDD and AVC69 cohort diagnoses was made by individual linkage between the 2 data sources. True positive cases of the HDD were cases which were confirmed by AVC69 and medical records review. False-positive cases of the HDD were cases which were confirmed by AVC69 and discarded through medical records review. False-negative results of the HDD were IS, which were not identified through the HDD search.

We defined sensitivity of the HDD as the proportion of patients identified as IS in the HDD among the whole population with confirmed IS. Positive predictive value (PPV) was defined as the proportion of patients with a confirmed IS among those identified as IS in the HDD. Sensitivity and PPV were presented as percentages with their 95% confidence intervals (95% CIs).

Factors Associated With Coding Quality
Univariate analyses were performed to identify factors associated with the probability of false negative among several factors: age, sex, and management (being hospitalized in neurological ward or stroke unit versus other wards, having had an MRI or a computed tomography scan) were analyzed. A P value <0.05 was considered significant. Statistical analysis was performed using the SAS software (SAS Institute Inc, Cary, NC, v9.1). This study was approved by the French data protection agency (CNIL) and the Institutional Review Board for the 4 French university hospitals.

Results

Study Sample
The inclusion process is described in the Figure. The gold standard (AVC69 cohort and medical record review) identified 465 cases with a confirmed diagnosis of IS. From the HDD extraction, 329 records were identified with a primary diagnosis code of IS.

Stroke Identification in the HDD and HDD Validity
Among the 465 confirmed cases of IS, 313 cases were correctly identified through HDD extraction, but 152 cases were missed. The false-negative rate, that is, the proportion of IS cases undetected by HDD search was, thus, 32.7% (95% CI, 28.4–37.0; 152 of 465 cases). Among the 329 cases identified through HDD search, 16 were false positives, which were discarded on medical record review, the false-positive rate was, thus, 4.9% (95% CI, 2.6–7.2; 16 of 329 cases).

The sensitivity of the HDD search was 67.3% (95% CI, 63.1–71.5), and the PPV was 95.1% (95% CI, 92.8–97.4; Table 2). Within the subgroup of patients hospitalized in the stroke unit (n=122), the sensitivity was 80.3% (95% CI, 75.7–84.8), and the PPV was 96.1% (95% CI, 93.3–99.9;
Table 3. In the group of patients who had a computed tomography scan (n=397), the sensitivity was 64.2% (95% CI, 59.5–69.0; 255 of 397 patients); this was 75.4% (95% CI, 68.0–82.8) among the 130 patients who received an MRI (98 of 130 patients). Among the subgroup of patients who were thrombolyzed (n=57), sensitivity of HDD search was 87.7% (95% CI, 79.2–96.2; 50 of 57 patients).

Table 3 describes false-positive and false-negative cases. Among the 152 false-negative cases not identified through HDD search as IS, nearly 30% were attributable to a lack of precision regarding the pathogenesis of stroke because they were coded with the I64 code (stroke without indication on ischemic or hemorrhagic pathogenesis). Approximately 20% were coded as transient ischemic attack (TIA) instead of stroke, and for 10%, only stroke symptoms were coded. Regarding false positives that were incorrectly coded as IS in the HDD, the actual diagnosis was mostly TIA in nearly 40% and hemorrhages in 25% of cases (hemorrhagic stroke, subarachnoid hemorrhage, or subdural hematomas).

Factors Associated With the Coding Quality
Being younger, being a man, being hospitalized in a stroke unit or in a neurological department versus any other type of department, having had an MRI versus computed tomography scan only, and having been thrombolyzed were factors significantly associated with a better coding of IS in the HDD (Table 4).

Table 2. Performances of HDD Search to Identify Ischemic Stroke Cases

<table>
<thead>
<tr>
<th>Cases Identified in HDD</th>
<th>Confirmed IS*</th>
<th>True-Positive IS</th>
<th>False-Negative IS</th>
<th>False-Positive IS</th>
<th>Sensitivity† (95% CI)</th>
<th>PPV‡ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All patients</td>
<td>329</td>
<td>465</td>
<td>313</td>
<td>152</td>
<td>67.3% (63.1–71.5)</td>
<td>95.1% (92.8–97.4)</td>
</tr>
<tr>
<td>Patients at stroke unit</td>
<td>102</td>
<td>122</td>
<td>98</td>
<td>24</td>
<td>80.3% (75.7–84.8)</td>
<td>96.1% (93.3–99.9)</td>
</tr>
</tbody>
</table>

CI indicates confidence interval; HDD, hospital discharge database; IS, ischemic stroke; and PPV, positive predictive value.
*Gold standard: AVC69 and medical record.
†True-positive cases/confirmed IS.
‡True-positive cases/confirmed IS.

Discussion
Our results showed that the main limit of using the French HDD for acute IS identification was a high proportion of false-negative cases: ≈33% of confirmed ISs were not coded as ISs in the HDD and would be missed if estimates were to be based on such databases, which are accessible at the national level. The sensitivity of acute IS ICD-10 codes was globally 65% in our university hospital, increasing to 80% in the stroke unit. PPV was satisfying, reaching 85%.

The French HDD was not exhaustive in identifying IS patients. We retrieved a relatively poor sensitivity compared with international data in North European countries or North America, which retrieved a sensitivity of up to 90% to 95%.11–15 However, the majority of these studies included all types of strokes,11–13 and analysis of subtypes revealed that HDD accuracy was better for hemorrhagic strokes than for ISs.11,14,15 We found a higher sensitivity among patients managed in the stroke unit and thrombolyzed, revealing a lower proportion of false-negative cases for these patients. At the time of the study, only 26.5% of acute ISs were managed in this specialized stroke unit, and the majority of patients were managed in other nonspecialized units. The more appropriate coding in the stroke unit and among patients who were thrombolyzed could induce a bias in evaluating efficiency of health policies based on the HDD with a risk of overestimation of the thrombolysis rate. Indeed, because patients who were thrombolyzed are more appropriately coded than others, if the...
The PPV, reflecting the probability of being treated for IS if coded as IS in the database, was high in our sample (85%), in concordance with results of previous studies, revealing a PPV ≈ 85% to 90%.11,13,17 Studies on all strokes revealed that the PPV was greater for IS than for hemorrhagic or all strokes (50%–80%). The lower PPV for nonischemic strokes reveals a risk of overestimation of global stroke incidence.11,13,16,18–20

The majority of previous studies were based on ICD-9 codes (430-438-9). The switch to the tenth version of the ICD (430-438-9). The switch to the tenth version of the ICD could have led to changes in coding practices and reliability. However, using ICD-10 codes to identify strokes in the Danish National Register of Patients, Krarup et al17 retrieved likewise a 97% PPV for IS but a lower value for intracranial hemorrhage (74%).

More recently, a study was published on the accuracy of the French HDD for the reporting of stroke.21 Comparison of their results with ours is interesting: they found a better sensitivity, 83.3% for IS from cardiac embolism and 77% for all stroke versus 67.3%. Their study is based on the data of the university hospital of Dijon (a city of 151,500 inhabitants). They probably had a better coding because it is a small hospital with homogenous coding procedures, whereas coding procedures in the 4 participating university hospitals were probably more heterogeneous, as they are expected to be when all French hospitals are considered in the French national HDD. Another reason to have a better sensitivity is that the only French stroke registry is in the city of Dijon. Hospital physicians are regularly contacted by the registry to provide their data and regularly receive advice on coding procedures for stroke. This is probably the reason why their coding quality improves over the years. Another difference is the PPV, which was better in our study than the study by Aboa-Eboule et al. Our algorithm selected only strokes with a code of cerebral imaging. This probably decreased the number of false-positive cases in our study but may, on the other hand, also decrease our sensitivity. Discrepancies between these 2 studies might be better understood by conducting a study at a national level.

The analysis of false-positive and false-negative cases of the HDD revealed 2 main situations leading to coding errors: the distinction between IS and TIA and between ischemic and hemorrhagic pathogenesis (I64 code). These coding errors may be linked to actual clinical difficulties in establishing a definite diagnosis. Studies on the reliability of the clinical diagnosis of TIA showed that this diagnosis was highly subjective, even for a neurologist and specialist of stroke.22 Ferro et al23 showed that interobserver reliability for the diagnosis of TIA was low, and that the main difficulty was the differential diagnosis with stroke. Regarding differential diagnosis between ischemic and hemorrhagic stroke, previous studies retrieved likewise a high proportion of unspecified stroke codes (I64) used instead of IS code (two thirds of the I64 codes for Krarup et al).13,17

Misleading codes attributable to clinical difficulties are not easy to avoid, whereas other errors in coding, such as the coding of stroke symptoms (20%) can be reduced by training physicians to the coding rules of the primary diagnosis.

Our results are limited to the university hospitals of the Rhône area. Thus, the results might be slightly different in nonuniversity public hospitals or in private hospitals. However, in those 4 hospitals, we could observe various practices and coding procedures. Furthermore, they receive a high volume of patients (second most important hospital volume in France) and especially a high number of strokes. The strength of our study is to be based on an exhaustive prospective cohort constituted over a 7-month period. A great effort had been made to identify and include every suspected case of stroke, and all included cases were validated by a neurologist. Only one stroke registry is available in France, and its coverage is restricted to 1 city (Dijon), which limits its generalizability. Available registry on stroke covering the Rhône area only records thrombolysed strokes; using this data source would also have introduced a major bias and restricted generalizability of results. However, comparison with an external exhaustive source of data, such as registry or cohort, is fundamental because this is the only way to measure the sensitivity and the 

---

**Table 3. Description of False-Positive and False-Negative Cases of the HDD Search**

<table>
<thead>
<tr>
<th>False-positive cases (n=16): actual diagnosis in medical record</th>
<th>True Positive (n=313)</th>
<th>False Negative (n=152)</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transient ischemic attack</td>
<td>6 (37.5)</td>
<td>10 (67.9)</td>
<td>0.763</td>
</tr>
<tr>
<td>Hemorrhagic stroke</td>
<td>2 (12.5)</td>
<td>3 (19.6)</td>
<td>0.835</td>
</tr>
<tr>
<td>Subarachnoid hemorrhage</td>
<td>1 (6.25)</td>
<td>0 (0)</td>
<td>1.000</td>
</tr>
<tr>
<td>Subdural hematomas</td>
<td>1 (6.25)</td>
<td>0 (0)</td>
<td>1.000</td>
</tr>
<tr>
<td>Other neurological condition leading to impaired motor function</td>
<td>2 (12.5)</td>
<td>2 (12.9)</td>
<td>0.904</td>
</tr>
<tr>
<td>Fall</td>
<td>2 (12.5)</td>
<td>2 (13.1)</td>
<td>0.981</td>
</tr>
<tr>
<td>Psychiatric disorder</td>
<td>2 (12.5)</td>
<td>2 (12.9)</td>
<td>1.000</td>
</tr>
<tr>
<td>Stroke, not specified as hemorrhage or infarction (I64)</td>
<td>43 (28.3)</td>
<td>5 (3.3)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Transient ischemic attack (I54)</td>
<td>26 (17.1)</td>
<td>18 (11.8)</td>
<td>0.038</td>
</tr>
<tr>
<td>Vascular syndromes of brain in cerebrovascular diseases (G46)</td>
<td>18 (11.8)</td>
<td>12 (7.9)</td>
<td>0.097</td>
</tr>
<tr>
<td>Stroke symptom (eg, aphasia, hemiplegia)</td>
<td>16 (10.5)</td>
<td>11 (7.2)</td>
<td>0.138</td>
</tr>
<tr>
<td>Sequelae of cerebrovascular disease (I69)</td>
<td>7 (4.6)</td>
<td>4 (2.6)</td>
<td>0.287</td>
</tr>
<tr>
<td>Nontraumatic intracerebral hemorrhage (I61)</td>
<td>7 (4.6)</td>
<td>3 (1.9)</td>
<td>0.210</td>
</tr>
<tr>
<td>Epilepsy (G40)</td>
<td>4 (2.6)</td>
<td>2 (1.3)</td>
<td>0.519</td>
</tr>
<tr>
<td>Traumatic lesion (S0600, S0650, S0691)</td>
<td>4 (2.6)</td>
<td>2 (1.3)</td>
<td>0.519</td>
</tr>
<tr>
<td>Other, miscellaneous</td>
<td>25 (16.4)</td>
<td>17 (11.2)</td>
<td>0.233</td>
</tr>
</tbody>
</table>

HDD indicates hospital discharge database.

---

**Table 4. Characteristics Associated With Coding Quality of Ischemic Stroke in the HDD**

<table>
<thead>
<tr>
<th>Ischemic Stroke (n=465)</th>
<th>True Positive (n=313)</th>
<th>False Negative (n=152)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, mean±SD</td>
<td>72.8±14.3</td>
<td>77.6±14.5</td>
<td>0.001</td>
</tr>
<tr>
<td>Male</td>
<td>160 (51.1)</td>
<td>62 (40.8)</td>
<td>0.037</td>
</tr>
<tr>
<td>CT</td>
<td>255 (81.5)</td>
<td>142 (93.4)</td>
<td>0.916</td>
</tr>
<tr>
<td>MRI</td>
<td>98 (31.3)</td>
<td>32 (21.1)</td>
<td>0.040</td>
</tr>
<tr>
<td>Neurological hospital stay</td>
<td>115 (36.7)</td>
<td>31 (20.4)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Stroke unit stay</td>
<td>98 (31.3)</td>
<td>24 (15.8)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Thrombolysis</td>
<td>50 (16.0)</td>
<td>7 (4.7)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

CT indicates computed tomography; and HDD, hospital discharge database.
false-negative rate and not just the PPV. This is very important because measuring only the false positives provides a biased estimate of HDD reliability.16,20

Analysis of each subgroup of stroke (TIA, IS, hemorrhagic stroke) with ICD-10 codes based on the HDD could not be considered reliable. Choice of the best algorithm remains difficult; an increase in sensitivity could lead to a decrease in PPV.11 Moreover, as shown by Tirschwell et al, the best algorithm depends on stroke subtype and research question.14 HDD accuracy for identifying stroke cases was very different according to the stroke subtypes.21

Conclusions

We cannot state that today, acute IS incidence can be based on HDD in France. Estimates of IS incidence based on HDD should be used with caution because of the high proportion of false-negative cases (32.7% of IS) leading to an underestimate of IS incidence. Conversely, our results showed few false-positive cases (4.9% of HDD identified cases). These results must be taken into account when interpreting and comparing IS incidence rates calculated based on these data. Routinely implemented indicators would be very useful in the context of stroke care reorganization to follow evolution in stroke management.

Acknowledgments

We acknowledge Nicole Berthoux and Alexia Henon for their help in data management and all physicians who participated in the AVC69 cohort study.

Sources of Funding

The AVC69 study was supported by a grant from the Program Hospitalier de Recherche Clinique 2006 of the French Ministry of Health (Ministère chargé de la Santé, Direction de l’Hospitalisation et de l’Organisation des Soins), Hospices Civils de Lyon, Lyon.

Disclosures

None.

References

Can Hospital Discharge Databases Be Used to Follow Ischemic Stroke Incidence?
Julie Haesebaert, Anne Termoz, Stéphanie Polazzi, Christelle Mouchoux, Laura Mechtouff, Laurent Derex, Norbert Nighoghossian and Anne-Marie Schott

*Stroke.* published online June 4, 2013;
*Stroke* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2013 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

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.113.001300

**Permissions:** Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Stroke* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

**Reprints:** Information about reprints can be found online at:
http://www.lww.com/reprints

**Subscriptions:** Information about subscribing to *Stroke* is online at:
http://stroke.ahajournals.org/subscriptions/