Measuring Accuracy of Discharge Diagnoses for a Region-Wide Surveillance of Hospitalized Strokes

Paolo Spolaore, MD; Stefano Brocco, MD; Ugo Fedeli, MD; Cristiano Visentin, ScD; Elena Schievano, ScD; Francesco Avossa, ScD; Giovanni Milan, MD; Vito Toso, MD; Diego Vanuzzo, MD, PhD; Lorenza Pilotto, MD; Achille Cesare Pessina, MD, PhD; Ruth Bonita, MPH, PhD

Background and Purpose—Measuring the impact of stroke through population-based stroke registers is complex and costly. The aim of the present study is to assess the validity of hospital discharge diagnoses (all ages) and to estimate the total number of hospitalized stroke events in the Veneto region (Northeastern Italy, 4 500 000 inhabitants).

Methods—All discharges covering a 1-year period (1999) from Veneto hospitals with International Classification of Diseases, 9th Revision codes 342, 430 to 434, and 436 to 438 were identified. A stratified sample was extracted and submitted to retrospective clinical record review according to the World Health Organization MONItoring trends and determinants in CArdiovascular disease stroke project. Using the positive predictive value (PPV) for validated acute stroke of each code to adjust for inaccuracy of discharge diagnoses, an estimate of hospital strokes was obtained.

Results—4015 admissions were reviewed. Codes 430, 431, 434, and 436 as primary diagnoses had the highest PPV, which sharply decreased in the other diagnostic levels. Code 342 also showed a high PPV. The probability of suspected events meeting the stroke definition increased with age and was highest for patients admitted to neurological wards and for fatal events. Overall 9400 strokes (first-ever and recurrent) were estimated to be hospitalized in 1999, with an attack rate of 208 per 100 000.

Conclusions—Our data indicate that once validation studies are undertaken on a sample of all hospitalized events, hospital discharge records can provide a valuable source of information on the actual burden of strokes on hospital services.

Key Words: epidemiology ■ International Classification of Diseases ■ stroke

Strokes represent a leading cause of death and disability; any effort to establish surveillance systems is therefore strongly warranted.1 The World Health Organization MONI-toring trends and determinants in CArdiovascular disease (WHO MONICA) project has provided an important source of information about stroke epidemiology by means of population-based registers accompanied by strict control of the quality of data collection.2 However, only subjects up to 65 or to 75 years of age were included, thus limiting an understanding of the total burden of stroke.3

The present study aims to assess the validity of stroke diagnosis on hospital discharge records (HDR) in the Veneto region (Northeastern Italy), to use the information to estimate the total burden of hospitalized stroke events using routinely collected HDR, as a part of a population-based register, and to verify the feasibility of this approach to monitor trends. This is the first region-wide study to our knowledge on accuracy of the International Classification of Diseases, 9th Revision (ICD-9) codes for stroke performed in Italy.

Materials and Methods

The total population of Veneto was 4 511 700 (ISTAT—National Institute of Statistics, 1999 data). The Regional Health System in Veneto, financed by the regional authority, is structured in local health units, with 65 hospitals for acute care, 28 for long-term care, providing ≈20 000 beds. There are ≈1 000 000 total discharges yearly.

All discharges from Veneto hospitals with ICD-9 codes related to stroke in all diagnostic categories were identified for the period January 1 to December 31, 1999. One primary and up to 5 secondary discharge diagnoses are registered in Veneto. Relevant codes (Table 1), were chosen because they were found to include most of the cases of stroke in pilot studies performed in Italy.4 In particular, the systematic evaluation of codes 410 to 414 and 798 to 799, as well as 250, 401 to 404, 420 to 429, and 440 to 447, showed negligible “false-negative” codes for real stroke events.4 Regarding those discharges with >1 diagnostic code related to stroke, we used only the first code reported to avoid double counting. Repeated hospital admissions referring to the same event were dropped from the population of selected discharges; they were defined as hospital admissions within 28 days from the previous admission or representa-ed patients transferred from one hospital to another.
Admissions suspected for stroke were first cross-classified into 4 groups: fatal/nonfatal events and ICD-9 code related to stroke reported as primary diagnosis/other diagnostic level. Then, for each group, a random sample stratified by gender, age class, and geographical area (local health unit) was extracted. To ensure comparative representation, ICD-9 codes were sampled at different rates; thus, the relative frequencies of ICD-9 codes in our sample do not reflect their relative frequencies among 1999 discharges.

The sample of suspected events was then submitted to a full retrospective clinical record review performed by trained physicians who applied the MONICA diagnostic criteria. Stroke was defined as rapidly developed signs of focal (or global) disturbance of cerebral function lasting >24 hours (unless interrupted by surgery or death), with no apparent cause other than a vascular origin. This definition does not include transient ischemic attacks or stroke events associated with trauma, blood disease, brain tumor, or brain metasteses. A stroke diagnosis was confirmed if onset of neurological symptoms occurred within 28 days before admission. Events were classified as “definite stroke,” “definite stroke associated with definite myocardial infarction,” “not stroke,” or “insufficient data.” If the event was included in the first 2 categories, it was again classified according to the MONICA Stroke protocol for the following subtypes: subarachnoid hemorrhage, intracerebral hemorrhage, cerebral ischemic infarction, and acute but ill-defined cerebrovascular disease. Data were also collected on health status before stroke, some measure of severity, patterns of care, and referral for rehabilitation.

We calculated the positive predictive value (PPV) for stroke of each identified hospital discharge code, distinguishing the first from all the other diagnostic positions. PPV was computed as the proportion of confirmed events (definite stroke associated or not with myocardial infarction) over all discharges with that code; the 95% confidence interval was obtained by means of approximation of the binomial to the normal distribution, taking into account the sampling fraction. Factors influencing the probability of a suspected event being confirmed as stroke at review were assessed by means of conditional logistic regression analysis (strata=ICD-9 discharge codes).

To obtain the estimated number of stroke cases, we multiplied the number of discharges with each code by its PPV; the total number of hospital strokes was then computed summing across codes. Stroke (first-ever) incidence rates were estimated by reducing the event rates (first-ever and recurrent) by the proportion of recurrent strokes found in the sample. The inpatient mortality was defined as the percentage of patients who died during hospitalization; in case of repeated admissions, events were defined as fatal or nonfatal on the basis of the discharge status in the last HDR.

Results

Of 40 796 discharges with the selected ICD-9 codes, 39 032 (95.7%) involved residents in the Veneto region. Repeated admissions referring to the same suspected event were 10.7% of all identified discharges; the percentage was higher among codes of hemorrhagic events (ICD 430=22.8%; ICD 431=17.7%; ICD 432=18.4%) or of late effects (ICD 342=22.8%; ICD 438=11.9%). The distribution of codes in the primary and in all other diagnostic levels in the remaining 34 869 hospital admissions is shown in Table 1. Considering that only 3% of all hospitalizations of Veneto residents occur outside the region, the number of suspected strokes missed because of treatment outside the study area is negligible.

Overall, 4015 of the admissions were sampled for review; PPV for stroke of ICD-9 codes is reported in Table 2. ICD-9 codes 430, 431, 434, and 436 as primary diagnoses had the highest values; however, a large proportion of discharges with a first code of hemiplegia (ICD-9 342) was found to meet the criteria for stroke at chart review. PPV sharply decreased when examining the other diagnostic levels, although the pattern is similar to that reported for primary diagnoses.

Table 3 shows the probability of a suspected event being confirmed as a stroke increased with increasing age, in patients admitted to neurology wards, and in those dying during hospitalization, whereas sex had a borderline influence.

From an extrapolation to the total Veneto population, overall 9400 strokes events (first-ever and recurrent) were estimated to be hospitalized in 1999: 4% of cases were classified as subarachnoid hemorrhage, 15% as intracerebral hemorrhage, and the remaining 81% as cerebral ischemic infarction or stroke of undetermined type.

Stroke attack rates increased exponentially with age for men and women (Figure). Men had higher age-specific rates than women; the lower overall crude rate in males is attributable to competing risk of mortality, leading to a decreased population in elderly men. The proportion of recurrent strokes found in the sample was 22%. These figures led to an estimate of an overall hospitalized attack rate of 208 and an incidence of 162 per 100 000 population. Almost one-quarter (23%) of the patients died during the hospitalization, with this percentage being higher for subarachnoid hemorrhage and an incidence of 162 per 100 000 population. Almost one-quarter (23%) of the patients died during the hospitalization, with this percentage being higher for subarachnoid hemorrhage.
hemorrhage (32%) and intracerebral hemorrhage (38%). Among fatal cases, 50% of patients died within 7 days, 37% from 8 to 28 days, and 13% after 28 days after admission. Case fatality increased from 14% to 30% in those aged younger than or older than 75 years, respectively, and was similar in males and females in each age class (data not shown).

Discussion

Several difficulties arise when estimating the burden of hospitalized stroke using routinely collected HDR: it is difficult to validate diagnoses and to distinguish multiple admissions for the same event.6 If the large proportion of patients with stroke readmitted soon after a previous discharge or transferred from one hospital to another is not taken into account, the number of events could be overestimated (especially hemorrhagic), leading also to a spurious reduction in the percentage of subjects dying during hospitalization.

Our study design is retrospective in nature and therefore does not allow computation of sensitivity or specificity of ICD-9 codes. However, using the PPV it is possible to adjust for inaccuracy of discharge codes and to reach an estimate of hospital strokes.5

As expected, ICD-9 diagnosis 433 included a low number of cases; the use of modifier codes (433.X0 and 433.X1) does not add further information.7 A study performed in Saskatchewan, Canada, reclassified as stroke a high proportion of 342 diagnoses, but the number of codes of hemiplegia was low.8 By contrast, in the Veneto region they accounted for almost 7% of the selected discharges and contributed to a substantial burden of estimated events. Codes 430, 431, 434, and 436 had the highest PPV, although figures were somewhat worse than those reported in other validation studies.6–14 However, our results represent average coding practices of several acute care facilities distributed on a large geographical area; therefore, they could better reflect widespread behaviors than the majority of previously published data, often collected in few hospitals, or in regions where the presence of population-based registries could have resulted in a local improvement of coding of stroke events.

Selection limited to primary diagnoses allows the identification of incident rather than prevalent strokes,7 and therefore increases the likelihood of agreement with the medical record review.9 Estimated hospital events are close to the number of discharges with codes 430, 431, 434, and 436 at the first diagnostic level. This is probably caused by false-positive and false-negative stroke cases counterbalancing each other. However, changes in diagnostic practices and in the provision of health care raise doubts that these errors are stable over time and question the extent to which trend might be derived from HDR.8,10

Results of logistic regression analysis indicate that the probability of an event to be classified as stroke is increased in patient admitted to neurology wards, aged 65 to 84 years, and who died during hospital stay. The higher probability found in specialist wards compared with general wards could be partially explained by the greater prevalence of admissions for stroke, a determinant of PPV besides sensitivity and specificity of coding. A similar explanation can be postulated for the increase with increasing age, whereas the leveling-off

<table>
<thead>
<tr>
<th>Terms</th>
<th>OR</th>
<th>CI</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;45 y</td>
<td>1.00</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>45–64 y</td>
<td>1.73</td>
<td>1.14–2.62</td>
<td>0.010</td>
</tr>
<tr>
<td>65–75 y</td>
<td>1.94</td>
<td>1.29–2.93</td>
<td>0.002</td>
</tr>
<tr>
<td>75–84 y</td>
<td>2.10</td>
<td>1.39–3.19</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>&gt;84 y</td>
<td>1.76</td>
<td>1.13–2.74</td>
<td>0.012</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.22</td>
<td>1.03–1.44</td>
<td>0.024</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ward</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical ward</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neurology</td>
<td>1.79</td>
<td>1.43–2.24</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Intensive care/neurosurgery</td>
<td>1.40</td>
<td>1.02–1.91</td>
<td>0.039</td>
</tr>
<tr>
<td>Other</td>
<td>0.31</td>
<td>0.24–0.41</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Discharge</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discharged alive</td>
<td>2.03</td>
<td>1.63–2.53</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Dead (fatal cases)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Odds Ratio With 95% Confidence Interval and 2-Tail Probability of Error Estimated by Means of Conditional Logistic Regression Analysis

Strata indicates ICD-9 discharge codes; dependent variable, confirmed stroke event; independent variables, age class, sex, death during hospitalization, and ward where patient was admitted; OR, odds ratio; CI, confidence interval.
in those aged 85 years or older could be related to diagnostic
uncertainties in elderly people with multiple comorbidities.\textsuperscript{1}
A study from an Italian hospital reported the wrong attribution
of stroke codes to be higher in patients with unexplained
sudden coma or death\textsuperscript{14}; by contrast, we found a larger
proportion of confirmed events among fatal with respect to
nonfatal suspected cases.

The percentage of recurrent stroke found in our sample was
very close to the figure reported in the literature. In fact, the
median proportion of recurrent stroke among MONICA
populations was 20%\textsuperscript{15}; it was 21% among patient discharged
from neurological wards in the Campania region (Southern
Italy).\textsuperscript{16}

Our data indicate that HDR can be regarded as a valuable
source of information on hospital strokes; therefore, the
importance of improving accuracy of ICD-9 diagnoses should
be further appraised. Certainly stroke surveillance has a key
role for health budget allocation, prevention, and outcome
evaluation in an aging population like that of the Veneto
region, and placing emphasis on improving the quality and
coverage of hospital diagnosis is likely to be a cheaper
alternative to population-based registers.

Acknowledgments
The study has been funded by the Veneto Region (D.G.R. December
28, 1998, 5089; and D.G.R. December 28, 1999, 4787) and by
Istituto Superiore di Sanità (convenzione 521C/5-VIII).

References
2. Asplund K, Bonita R, Kuulasmaa K, Rajakangas AM, Schaedlich H,
Suzuki K, Thorvaldsen P, Tuomilehto J. Multinational comparisons of
stroke epidemiology. Evaluation of case ascertainment in the WHO
MONICA Stroke Study. World Health Organization Monitoring Trends
3. Ferrario M, Giampoli S, Vancheri F, Vanuzzo D. Registro per gli eventi
coronarici e cerebrovascolari. Protocollo dello studio. Roma: Istituto
Superiore di Sanità; 2001. (Rapporti ISTISAN 01/8).
4. Menotti A, Giampoli S, Verdecchia A, Cesana GC, Feruglio GA,
Righetti F, Ferrario M, Vanuzzo D. Il Progetto MONICA (Monitoraggio
Malattie Cardiovascolari). Protocollo e manuale delle aree italiane.
Roma: Istituto Superiore di Sanità; 1989. (Rapporti ISTISAN 89/12).
5. Williams GR, Jiang JG, Matchar DB, Samsa GP. Incidence and
occurrence of total (first-ever and recurrent) stroke. \textit{Stroke.} 1999;30:
2523–2528.
7. Goldstein LB. Accuracy of ICD-9-CM coding for the identification of
patients with acute ischemic stroke: effect of modifier codes. \textit{Stroke.}
8. Liu L, Reeder B, Shuaib A, Mazagri R. Validity of stroke diagnosis on
hospital discharge records in Saskatchewan, Canada: implications for
DB. Inaccuracy of the International Classification of Diseases
(ICD-9-CM) in identifying the diagnosis of ischemic cerebrovascular
Mills D, Minneci L, Shukla R. The Greater Cincinnati/Northern Kentucky
Stroke Study: preliminary first-ever and total incidence rates of stroke
11. Leppala JM, Viitanen J, Heinonen OP. Validation of stroke diagnosis in
the National Hospital Discharge Register and the Register of Causes of
12. Rosamond WD, Folsom AR, Chambless LE, Wang CH, McGovern PG,
Howard G, Copper LS, Shahar E. Stroke incidence and survival among
middle-aged adults: 9-year follow-up of the Atherosclerosis Risk in
stroke in Norway: hospital discharge data compared with a
of ICD-9 codes in identifying ischemic stroke in the General Hospital of
15. Thorvaldsen P, Asplund K, Kuulasmaa K, Rajakangas AM, Schroll M.
Stroke incidence, case fatality, and mortality in the WHO MONICA
project. World Health Organization Monitoring Trends and Determinants
16. de Falco FA, Santangelo R, Majello L, Marasco G. Burden of acute stroke
and hospital resources in the Campania region of Italy. \textit{Cerebrovasc. Dts.}
2002;14:116–121.
Measuring Accuracy of Discharge Diagnoses for a Region-Wide Surveillance of Hospitalized Strokes
Paolo Spolaore, Stefano Brocco, Ugo Fedeli, Cristiana Visentin, Elena Schievano, Francesco Avossa, Giovanni Milan, Vito Tosó, Diego Vanuzzo, Lorenza Pilotto, Achille Cesare Pessina and Ruth Bonita

*Stroke*. 2005;36:1031-1034; originally published online March 24, 2005;
doi: 10.1161/01.STR.0000160755.94884.4a

*Stroke* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2005 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/36/5/1031