Identification of Incident Stroke in Norway
Hospital Discharge Data Compared With a Population-Based Stroke Register

Hanne Ellekjær, MD; Jostein Holmen, MD, PhD; Øystein Krüger, MD; Andreas Terent, MD, PhD

Background and Purpose—The validity of hospital discharge diagnoses is essential in improving stroke surveillance and estimating healthcare costs of stroke. The aim of this study was to assess sensitivity, positive predictive value, and accuracy of discharge diagnoses compared with a stroke register.

Methods—A record linkage was made between a population-based stroke register and the discharge records of the hospital serving the population of the stroke register (n=70,000). The stroke register (including patients aged 15 and older and with no upper age limit), applied here as a “gold standard,” was used to estimate sensitivity, positive predictive value, and accuracy of the discharge diagnoses classification. The length of stay in hospital by stroke patients was measured.

Results—Identifying cerebrovascular diseases by hospital discharge diagnoses (International Classification of Diseases, 9th Revision [ICD-9], codes 430 to 438.9, first admission) lead to a substantial overestimation of stroke in the target population. Restricting the retrieval to acute stroke diagnoses (ICD-9 codes 430, 431, 434, and 436) gave an incidence estimate closer to the “true” incidence rate in the stroke register. Selecting ICD-9 codes 430 to 438 of cerebrovascular diseases gave the highest sensitivity (86%). The highest positive predictive value (68%) was achieved by selecting acute stroke diagnoses (ICD-9 codes 430, 431, 434, and 436), at the expense of a lower sensitivity (81%). Accuracy of ICD codes 430 to 438.9 (n=678) revealed the highest proportion of incident strokes identified by the acute stroke diagnoses (ICD-9 codes 430, 431, 434, and 436). Seventy-four percent of hospital discharge diagnoses classified as first-ever stroke kept the original diagnosis. Only 4.6% of the discharge diagnoses were classified as nonstroke diagnoses after validation. The estimation of length of stay in the hospital was improved by selection of acute stroke diagnoses from hospital discharge data (ICD-9 codes 430, 431, 434, and 436), which gave the same estimate of length of stay, a median of 8 days (2.5 percentile=0 and 97.5 percentile=56), compared with a median of 8 days (2.5 percentile=0 and 97.5 percentile=51) based on the stroke register.

Conclusions—Hospital discharge data may overestimate stroke incidence and underestimate the length of stay in the hospital, unless selection routines of hospital discharge diagnoses are restricted to acute stroke diagnoses (ICD-9 codes 430, 431, 434, and 436). If supplemented by a validation procedure, including estimates of sensitivity, positive predictive value, and accuracy, hospital discharge data may provide valid information on hospital-based stroke incidence and lead to better allocation of health resources. Distinguishing subtypes of stroke from hospital discharge diagnoses should not be performed unless coding practices are improved. (Stroke. 1999;30:56-60.)

Key Words: stroke assessment ■ diagnosis ■ stroke classification ■ epidemiology

Until recently, hospital discharge data in Norway was the only source of information on stroke morbidity. Health service routines, technology, and resources available in the community, thus possessing potential biases,1-5 influence admission rates of stroke. Changes of interest in stroke care (eg, introduction of stroke units), implementation of new diagnostic tools such as CT scan, and changes in surveillance procedures (eg, coding practices) affect the discharge diagnoses of cerebrovascular diseases and thus the reflection of stroke events in the community. Several studies use hospital discharge diagnoses of cerebrovascular diseases as surrogate markers of incident cases of stroke, and most studies evaluating health care cost of stroke are based on hospital discharge data.6-8 Few studies, however, have validated hospital discharge data of cerebrovascular diseases in descriptive epidemiological studies. Because of changes in admission rates over time, discharge data may either underestimate9 or overestimate4,9 stroke incidence. Many population-based stroke registers are restricted to cases under the age of 75, and validation studies involving older patients are particularly scarce.1,3
A recently established stroke register in Norway (including patients aged 15 and older and with no upper age limit), used here as a “gold standard,” made it possible to assess the validity of discharge diagnoses. The aim of this article was to estimate positive predictive value and sensitivity of hospital discharge data of first-ever stroke and to study the accuracy of International Classification of Diseases, 9th Revision (ICD-9), codes of cerebrovascular diseases by comparing ICD-9 codes of hospital discharge diagnoses with the population-based stroke register. The median length of hospital stay is presented to demonstrate the potential impact on cost-of-illness analyses of various methods of retrieving stroke diagnoses from hospital discharge data.

Subjects and Methods

Identification of Incident Stroke in the Population-Based Stroke Register

The population-based stroke register was used as a gold standard. In the period September 1, 1994, through August 31, 1996, uniform information was collected on all suspected cases of first-ever and recurrent stroke occurring in people aged 15 years and older in the region of Innherred, Nord-Trøndelag county, in the central part of Norway (target population, 70,000). Case-finding routines included hospitalized and nonhospitalized cases. Every suspected stroke case was examined according to a standardized form within 48 hours after stroke onset. Discharge diagnoses from neighboring hospitals and death certificates with stroke mentioned were reviewed regularly. Altogether, 430 first-ever (72.8%) and 161 recurrent (27.2%) strokes were registered. Final diagnosis was made by one of the authors (H.E.), who applied the World Health Organization criteria: rapidly developing clinical signs of focal (or global) disturbance of cerebral function, lasting >24 hours or leading to death, with no apparent cause other than that of vascular origin. Classification into the subgroups of cerebral infarction, intracerebral hemorrhage, and subarachnoid hemorrhage (ICD-9 codes 430, 431, and 434) was based on CT or necropsy reports. Misclassification was examined by measuring interobserver agreement. Details of design and methods are published elsewhere.

Identification of Incident Stroke by Use of Hospital Discharge Data

Patients with discharge diagnoses of ICD-9 codes 430 to 438.9 (all positions) and ICD-9 codes 430, 431, 434, and 436 (all positions) at first admission to the hospital in the study period were selected. By using the personal identification number, a link was made between the first-ever strokes in the stroke register and those in the hospital discharge diagnoses. In this way, discharge diagnoses associated with first-ever stroke were identified. A surrogate estimate of stroke incidence based on hospital discharge data were compared with estimates from the stroke register, with and without hospitalized cases included.

Sensitivity and Positive Predictive Value

Sensitivity was defined as the proportion of first-ever stroke in the discharge data (identified by the stroke register) to first-ever stroke in the stroke register. Positive predictive value was defined as the proportion of first-ever stroke among discharge diagnoses.

Accuracy of Diagnoses

A comparison by cross-classification of discharge diagnoses versus hospitalized cases in the stroke register was made. A corresponding event was identified if the date of admission to the hospital and the date of the event in the stroke register coincided (±28 days).

Length of Stay in Hospital

The median length of stay in days was calculated for all hospitalized cases with discharge diagnosis ICD-9 codes 430 to 438.9. A comparison between validated stroke diagnoses in the stroke register and hospital discharge diagnoses was made.

Results

A total of 759 ICD-9 codes 430 to 438.9 and 508 acute stroke diagnoses (ICD-9 codes 430, 431, 434, and 436) were selected from the discharge records in the period of interest (Table 1). Identifying cerebrovascular diseases by hospital discharge diagnosis ICD-9 codes 430 to 438.9, first admission, gave an incidence estimate of 5.48 per 1000 (95% CI, 4.92 to 6.04) in the target population. Restricting the retrieval to acute stroke diagnoses (ICD-9 codes 430, 431, 434, and 436), first admission, gave the best estimate of stroke morbidity measured as incidence rate in the population, 3.67 per 1000 (95% CI, 3.21 to 4.13). In comparison, estimates from the stroke register were 2.78 per 1000 (95% CI, 2.39 to 3.17) and 3.10 per 1000 (95% CI, 2.84 to 3.42) when nonhospitalized cases were included.

After linkage to the stroke register, a total of 369 events among discharge diagnosis ICD-9 codes 430 to 438.9 (n = 759) were associated with a first-ever stroke, giving a positive predictive value of 49%. Discharge diagnosis ICD-9 codes 430 to 438.9 contained 369 first-ever stroke cases out of a total of 430 first-ever strokes in the population-based stroke register, giving a sensitivity of 86%. Restricting the analyses to hospitalized cases in the register increased to 95%

Table 1: Comparison of Different Approaches in Identification of Incidence of First-Ever Stroke by Use of Stroke-Coded Hospital Discharge Diagnosis With Stroke Register

<table>
<thead>
<tr>
<th>Variable</th>
<th>Discharge Diagnosis</th>
<th>Stroke Register</th>
<th>Stroke Register</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICD-9 Codes</td>
<td>ICD-9 Codes</td>
<td>(Hospitalized)</td>
</tr>
<tr>
<td></td>
<td>430-438, First</td>
<td>430, 431, and</td>
<td>369/389 (95%)</td>
</tr>
<tr>
<td></td>
<td>Admission</td>
<td>434, and 436,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>First Admission</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>347/389 (89)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>347/430 (81)</td>
</tr>
<tr>
<td>No. of cases</td>
<td>759</td>
<td>508</td>
<td>389</td>
</tr>
<tr>
<td>Incidence per 1000, n</td>
<td>5.48</td>
<td>3.67</td>
<td>2.78</td>
</tr>
<tr>
<td>95% CI</td>
<td>4.92–6.04</td>
<td>3.21–4.13</td>
<td>2.39–3.17</td>
</tr>
<tr>
<td>Hospitalized only</td>
<td>369/389 (95%)</td>
<td>347/389 (89)</td>
<td></td>
</tr>
<tr>
<td>Including nonhospitalized</td>
<td>369/430 (86%)</td>
<td>347/430 (81)</td>
<td></td>
</tr>
<tr>
<td>Positive predictive value, n (%)</td>
<td></td>
<td></td>
<td>347/508 (68)</td>
</tr>
<tr>
<td>Incidence of stroke</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
the sensitivity of discharge diagnosis ICD-9 codes 430 to 438.9. Selecting only acute stroke diagnosis ICD-9 codes 430, 431, 434, and 436 gave a total of 508 discharge diagnoses, and the proportion of first-ever strokes identified decreased to 81% (347 of 430) and 89% when compared with only hospitalized cases in the stroke register. However, the positive predictive value increased to 68% (Table 1).

The accuracy of discharge diagnoses assessed by cross-classification of ICD-9 codes 430 to 438 versus the stroke register categories of first-ever stroke, recurrent, other stroke diagnoses, and nonstroke is shown in Table 2. Ninety percent (678 of 759) had a validated diagnosis in the stroke register. A total of 81 events did not have a match in the stroke register: 32 events outside the accepted range of time (± 28 days) and 49 with no medical records available. Most first-ever incident strokes were confirmed by discharge diagnosis ICD-9 codes 430, 431, 434, and 436; the proportions identified as first-ever stroke were 69%, 77%, 68%, and 68%, respectively. Fifteen percent (73 of 471) of acute stroke diagnoses (ICD-9 codes 430, 431, 434, and 436) were classified as other stroke diagnoses (ICD-9 codes 432, 433, 435, 437, and 438) or nonstroke diagnoses in the stroke registry. Of the discharge diagnoses classified as first-ever stroke according to the stroke register, 74% (260 of 351) kept the original diagnosis. Most noteworthy, acute strokes cases with negative CT results were classified as ICD-9 code 436 (unspecified) at discharge; according to the diagnostic criteria in the stroke register, these should have been categorized as ischemic stroke events, thus leading to an underestimation of cerebral infarction (66.5% in the discharge data compared with 74.5% in the stroke register). Discharge data gave a good estimate of the proportion of subarachnoid and intracerebral hemorrhage in the population (2.8% and 12.0%, respectively) compared with 3.0% and 10.5% in the stroke register.

Thirty-six percent of the stroke cases were found to be other than first-ever or recurrent stroke. Nearly half of the

<table>
<thead>
<tr>
<th>Stroke Register</th>
<th>Discharge Diagnoses ICD-9 Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>All (n=678)</td>
<td>430 (n=13)</td>
</tr>
<tr>
<td>First-ever stroke, n</td>
<td>12</td>
</tr>
<tr>
<td>Recurrent, n</td>
<td>7</td>
</tr>
<tr>
<td>Nonacute stroke diagnoses, n</td>
<td>4</td>
</tr>
<tr>
<td>Incidence of stroke, %</td>
<td>69</td>
</tr>
<tr>
<td>Recurrent, %</td>
<td>0</td>
</tr>
<tr>
<td>Nonacute stroke diagnoses, %</td>
<td>8</td>
</tr>
<tr>
<td>Nonstroke, %</td>
<td>23</td>
</tr>
</tbody>
</table>
discharge diagnoses (327 of 678) were found to be other than first-ever stroke. Thirty-one of 678 (4.6%) were classified as nonstroke events.

The median length of hospital stay of patients with discharge diagnosis ICD-codes 430 to 438 (n=678) who had a match in the stroke register was 6 days (2.5 percentile=0 and 97.5 percentile=50). Retrieving only acute stroke diagnoses, first admission (ICD-9 codes 430, 431, 434, and 436) from hospital discharge data gave a median length of stay of 8 days (2.5 percentile=0 and 97.5 percentile=56). Of those who had a “true” stroke according to the stroke register (n=431, first-ever or recurrent hospitalized stroke), the median length of stay was 8 days (2.5 percentile=0 and 97.5 percentile=51).

**Discussion**

The present study demonstrates that restricting the hospital discharge diagnoses to acute stroke diagnoses (ICD-9 codes 430, 431, 434, and 436) gave an incidence estimate close to the true incidence rate of stroke in the stroke register. Selecting ICD-9 codes 430 to 438 gave the highest sensitivity of 86%, with only 14% of the incident strokes not identified. Identifying cerebrovascular diseases by hospital discharge diagnosis ICD-9 codes 430 to 438.9, first admission, led to a substantial overestimation of stroke when the stroke register was used as a gold standard. The estimation of length of hospital stay was improved by the selection of acute stroke diagnoses (ICD-9 codes 430, 431, 434, and 436) from hospital discharge data.

A strength of this study was the population-based stroke register. We made a great effort to trace every suspected case of stroke in the community, in the local hospital, and in the 2 neighboring hospitals. A validation of the registration procedures showed a case-finding rate of nearly 100% in the age group 15 to 74 years, 87% in the age group 75 to 84 years, and 76% in age group ≥85 years. A higher completion of case finding in the stroke register would lead to a reduction of the sensitivity of hospital discharge diagnoses.

Misclassification was measured by interobserver agreement between the research registrar, a neurologist, and a specialist in internal and physical medicine (the latter 2 from other hospitals). The validation of the interobserver agreement gave a $\kappa$ index of 0.68 between research registrar and neurologist, 0.61 between research registrar and specialist in internal medicine, and 0.65 between neurologist and specialist in internal medicine. Despite acceptable $\kappa$ indices, inaccuracy of diagnostic procedures in the stroke register may explain some of the discrepancies between the stroke register and hospital discharge diagnoses. When using the stroke register as a gold standard, these limitations should be taken into account.

In the hospital discharge register, 49 patient files could not be traced. This illustrates practical problems in maintaining high-quality hospital discharge files. There is no reason to assume that the missing files were selected, and they are therefore not likely to have any substantial impact on the results of the study.

A study from Rochester, Minn, reported 23% incident strokes missing from hospital discharge diagnoses. This study included persons aged ≥75 years but combined the data sets from 1980 and 1989, in which 19% of first-ever strokes were not hospitalized during the event. In the Innherred, Norway, stroke register, only 10% of first-ever stroke patients were not hospitalized; this may explain the differences in sensitivity between the 2 studies. The only Scandinavian study comparing discharge diagnoses versus a population based stroke register, the MONICA register in Northern Sweden, reported that 6% of incident stroke cases were not identified by hospital discharge diagnoses. However, this study reported results from a population <75 years of age, and the hospital discharge diagnoses did not include fatal events. This difference in study population and design may explain the lower proportion of unidentified incident stroke cases in the hospital discharge data in the MONICA register compared with that in the present study.

Limiting the number of diagnoses to acute stroke increased the positive predictive value of incident first-ever stroke to 68%, which is exactly the same as that reported from The MONICA register in Sweden and higher than that found in the Rochester study. Again, the differences in study design described above must be taken into account. Differences in age distribution in the study population and time of study may influence the probability of an acute stroke discharge diagnosis being a first-ever stroke event.

Accuracy of hospital discharge diagnoses have been validated in several studies performed at different times and in different populations. Additionally, the local diagnostic tools and coding practices may explain the discrepancies between studies. Leibson et al described the limitations of using discharge diagnoses in the classification of stroke by type. They found a lower proportion of unspecified stroke when a neurologist reviewed the medical records, indicating that knowledge about criteria and definitions of stroke subtypes are important in obtaining reliable classification. In the present study, 17% of incident strokes by discharge data were categorized as unspecified stroke compared with 12.0% in the stroke registry. The corresponding data in the Rochester study were 29% and 10%, respectively. Lindblad et al found that 73% of discharge diagnoses suggesting an acute stroke kept the original diagnosis after validation, compared with 74% in the present study. This Swedish study did not distinguish between first-ever and recurrent stroke, and therefore comparison of these data may be doubtful.

The validation of discharge diagnoses in the present study showed that only a few cases (4.6%) were classified as nonstroke events, which indicates that such data may be valid for analytical studies of predictors of stroke. Studies indicate also that applying more strict hospital routines, eg, in registering first-admission and acute stroke diagnoses (ICD-9 codes 430, 431, 434, and 436), might improve the quality of hospital stroke registers.

Calculations of costs for stroke are dependent on reliable incidence data. Ours study reveals that use of hospital discharge diagnoses of stroke must be used with caution to avoid an overestimation of first-ever stroke cases. The direct cost of providing medical care to stroke patients consists of, among other components, the length of hospital stay. Bed-days in the hospital are a major contributor to the cost of acute care of stroke, and the validity of such data are enhanced by assessment of the accuracy of the discharge diagnoses.
According to the findings of the present study, we conclude that hospital discharge diagnoses are valuable sources of stroke incidence data for both health service planning and epidemiological research. Acute stroke diagnoses are most valid, and better diagnostic routines would improve the validity of the other hospital discharge diagnoses.

Appendix

ICD-9 Codes for Cerebrovascular Disease (430–438)

430—Subarachnoid hemorrhage
431—Intracerebral hemorrhage
432—Other and unspecified intracranial hemorrhage
433—Occlusion and stenosis of precerebral arteries
434—Occlusion of cerebral arteries
435—Transient cerebral ischemia
436—Acute but ill-defined cerebrovascular disease
437—Other and ill-defined cerebrovascular disease
438—Late effects of cerebrovascular disease

Acknowledgment

This study was supported by the Norwegian Research Council.

References

Identification of Incident Stroke in Norway: Hospital Discharge Data Compared With a Population-Based Stroke Register
Hanne Ellekjær, Jostein Holmen, Øystein Krüger and Andreas Terent

Stroke. 1999;30:56-60
doi: 10.1161/01.STR.30.1.56

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/30/1/56

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/