Accuracy of Hospital Discharge Abstracts for Identifying Stroke

Cynthia L. Leibson, PhD; James M. Naessens, MPH; Robert D. Brown, MD; Jack P. Whisnant, MD

Background and Purpose  Much of the available data on stroke occurrence, service use, and cost of care originated with hospital discharge abstracts. This article uses the unique resources of the Rochester Epidemiology Project to estimate the sensitivity and positive predictive value of hospital discharge abstracts for incident stroke.

Methods  The Rochester Stroke Registry was used to identify all confirmed first strokes (hospitalized and nonhospitalized) among Rochester residents for 1970, 1980, 1984, and 1989 (n=364). The sensitivity of discharge abstracts was estimated by following these individuals for 12 months after stroke to determine the proportion assigned a discharge diagnosis of cerebrovascular disease (International Classification of Diseases [ICD] codes 430 through 438.9). The positive predictive value of discharge abstracts was assessed by identifying all hospitalizations of Rochester residents with an ICD code of 430-438.9 in 1970, 1980, and 1989 (n=377). Events were categorized as incident stroke, recurrent stroke, stroke sequelae, or nonstroke after review of the complete community-based medical record by a neurologist.

Much of the information on the epidemiology of stroke and stroke-associated service use in the United States has relied on hospital discharge abstracts with stroke listed as a diagnosis. Discharge abstracts are limited in that not all persons with stroke are hospitalized, not all hospitalized strokes are recorded, it is difficult to validate diagnoses and distinguish subtypes, and it is difficult to distinguish multiple admissions for the same event and first stroke from recurrent events. These biases are recognized, and studies of hospital-based stroke vary greatly in the amount and type of information available to identify and confirm incident stroke. Estimates of the potential bias, however, are essentially unavailable for the United States because of a shortage of community-based stroke registries in this country that can serve as a standard for comparison.

The present study provides estimates of the accuracy of hospital discharge abstracts for stroke by taking advantage of two previously identified, well-described data sets. The first data set is a population-based stroke registry consisting of all first strokes (both hospitalized and unhospitalized) occurring among residents of Rochester, Minn, for the years 1935 through 1989. For purposes of this study, all persons with confirmed first stroke (transient ischemic attack excluded) in 1970, 1980, 1984, and 1989 were considered. Using the resources of the Rochester Epidemiology Project medical records linkage system, we reviewed the complete community-based medical records of these individuals for all hospitalizations in the 12 months after stroke. Up to five diagnoses for each hospitalization were recorded and coded by trained nosologists. The proportion of hospitalized incidence cases for which there was a diagnosis code for stroke in the year after stroke afforded estimates of the sensitivity of discharge abstracts for incident stroke.

Estimates of positive predictive value of hospital discharge abstracts were obtained from a separate data set, the Olmsted County Hospital Utilization Data Set. This data set characterizes all hospitalizations among Olmsted County, Minn, residents, including Rochester, the county seat. For the present study, all hospitalizations with an International Classification of Diseases (ICD) discharge code of 430 through 438.9 (Cerebrovascular Disease; see "Appendix") were identified for the years 1970, 1980, and 1989 from the Olmsted County Hospital Utilization Data Set. Hospitalizations by non-Rochester/Olmsted County residents were excluded from the analysis. The remaining hospitalizations were first matched against the Rochester Stroke Registry to identify those associated with an incident stroke event. Those not identified in the registry were reviewed by a neurologist (R.D.B.) and categorized as either incident stroke, recurrent stroke, nonstroke event, or sequelae.
To compare the accuracy of alternative approaches using hospital discharge abstracts cited in the literature, the data were analyzed under various assumptions regarding the number of discharge diagnoses available and the diagnostic categories under consideration.

**Subjects and Methods**

The opportunity for longitudinal, population-based epidemiological and utilization studies in Rochester, Minn, is the result of a unique set of circumstances. Rochester is located 80 miles from other urban centers and is the home of one of the world's largest medical centers, the Mayo Clinic. Therefore, residents receive medical services from a limited number of providers, primarily the Mayo Clinic and its two affiliated hospitals and a second group practice, Olmsted Medical Group and its affiliated Olmsted Community Hospital. Since the turn of the century, every Mayo Clinic patient has been assigned a unique identifier, and information from any encounter (whether inpatient, outpatient, emergency room, home visit, nursing home, or autopsy) is contained within a unit medical record. In addition, the diagnoses for each encounter are coded and entered into a central index for ease of retrieval, part of the Rochester Epidemiology Project. This index was expanded to include diagnoses made at other medical facilities in and around Olmsted County (Olmsted Medical Group, Olmsted Community Hospital, University of Minnesota Hospitals, Minneapolis Veterans Administration Hospital, Rochester State Mental Hospital, and the very few private general practitioners in the area).

**Rochester Stroke Registry**

The Rochester Stroke Registry identifies all first strokes occurring among residents of Rochester, Minn, for the years 1935 through 1989. The registry was created using the resources of the Rochester Epidemiology Project records linkage system. Listings were generated of all residents of Rochester who were diagnosed as having stroke, or disorders that could be mistaken for stroke, during this time. The complete medical records for all potential cases were reviewed by a trained nurse abstractor, under supervision by a neurologist, to identify all first episodes of stroke during this period. All potential cases were evaluated using the same inclusion and exclusion criteria. Criteria for diagnosis have been reported elsewhere and include signs of focal neurological deficit due to vascular lesion of the central nervous system that were present for at least 24 hours and clinical characteristics suggesting that stroke was the cause of the lesion.7 Strokes diagnosed only at autopsy must have been recent to qualify as a case. To exclude persons who may have moved to Rochester for treatment of an existing disorder, cases were limited to persons who resided in Rochester for at least 1 year before the event date. The essentially complete ascertainment of stroke cases afforded by the medical records linkage system has been validated in an earlier study that compared stroke incidence rates using the method described here with rates obtained from prospective follow-up of an initially stroke-free cohort.13

**Olmsted County Hospital Utilization Data Set**

As a consequence of the geographical isolation and the small number of providers, over 95% of all hospitalizations among Olmsted County residents occur in the three area hospitals that are affiliated with the two practice groups, Mayo Clinic and Olmsted Medical Group (1988 MEDPAR File, Health Care Financing Administration). The utilization and billing data systems for the two practice groups and the three affiliated hospitals, Rochester Methodist Hospital, Saint Mary's Hospital, and Olmsted Community Hospital, were linked through a previous project to allow "capture" of inpatient activity for Olmsted County residents for the years 1970, 1980, 1989, and 1990.10,14 Inpatient charges were available for 1989 and 1990 hospitalizations. The maximum number of diagnoses available per discharge from this administrative data set was three in 1970, five in 1980, and 15 in 1989 through 1990.

**Sensitivity Analysis**

The analysis of sensitivity of discharge abstracts originated with the Rochester Stroke Registry. For the analysis, all episodes of first stroke, both hospitalized and nonhospitalized, were identified from the registry for the years 1970 (n = 64), 1980 (n = 81), 1984 (n = 111), and 1989 (n = 108). The selection of years was determined in part by the years for which data were available from the Olmsted County Hospital Utilization Data Set and by the fact that, when this study was initiated, the latest year for which incident strokes were identified was 1984. Simultaneous to our investigation, the update of the registry was completed, affording consideration of the most recent year, 1989. Data obtained from the registry included a unique patient identifier, sex, date of birth, type of stroke, whether the individual was hospitalized for the event, whether the event occurred during a hospitalization, date of last follow-up, and status at last follow-up. For 1989-incidence cases only, it was possible to use the computerized resources of the Olmsted County Hospital Utilization Data Set to identify all hospitalizations for the full 12 months after stroke (whether 1989 or 1990). Information obtained from the utilization data set included dates of admission and discharge, total hospital charges (ie, comparable to Medicare Part A charges), disposition at discharge (home and/or home health care, nursing home, hospital transfer, death), and discharge diagnoses. Olmsted County Hospital Utilization data were not available for 1971, 1981, 1984, or 1985. To obtain comparable data for 12 months after stroke for 1970-, 1980-, and 1984-incidence stroke cases, the complete community-based medical record for each individual was reviewed for all inpatient experience 12 months after stroke. Charge data were not available. Up to five discharge diagnoses for every hospitalization for 1970-, 1980-, 1984-, and 1989-incidence cases under study were coded by a trained nosologist, following recommendations outlined for the Uniform Hospital Discharge Data Set.15

**Analysis of Positive Predictive Value**

The analysis of positive predictive value originated with the Olmsted County Hospital Utilization Data Set. All hospital discharges among Olmsted County residents with an ICD diagnosis code of 430-438.9 for any of the first three diagnoses in 1970 and any of the first five diagnoses in 1980 and 1989 were identified from the utilization data set for each of those years. These were matched against the Rochester Stroke Registry, and any hospitalizations not associated with an incident stroke event in the registry were selected for residency review. Hospitalizations for which the individual did not satisfy the residency definition as applied in the stroke registry (ie, resident of Rochester for at least 1 year before the event) were excluded from the analysis. The complete medical records of remaining individuals were then reviewed by a neurologist (R.D.B.), who used the same criteria as was used in the stroke registry to categorize the hospitalization as associated with either incident stroke, recurrent stroke, stroke sequelae, or nonstroke event (eg, transient ischemic attack).

It is important to note that the analyses of sensitivity and positive predictive value were performed separately and should be interpreted separately. Although the stroke registry and the hospital utilization data sets both provide annual surveys of the Rochester population, the data sets were derived from different sources and constructed for different purposes. While many incident strokes and associated hospitalizations were found in both data sets, this was not always the case. A number of hospitalizations from the utilization data set with a discharge diagnosis code of 430-438.9, confirmed by the neurologist's medical record review as recurrent stroke or...
stroke sequelae, were individuals identified as prevalence nonincidence cases (ie, their first stroke occurred outside of Rochester). The 12-month periods surveyed for the two analyses were also not contemporaneous. The positive predictive analysis was based on a subset of the utilization data set consisting of stroke-coded hospital discharges during the calendar years 1970, 1980, or 1989. Some individuals identified in the utilization data set with a discharge diagnosis code for stroke in the calendar year of interest (eg, 1/15/70) were admitted in the previous year (eg, 12/25/89) and were confirmed by the neurologist’s review as hospitalized for incident stroke. By contrast, the sensitivity analysis, although beginning with incident events in 1970, 1980, 1984, and 1989, considered all hospitalizations in the entire 12-month period following the event.

### Results

The Rochester Stroke Registry revealed 364 first strokes among residents of Rochester, Minn, in the years 1970, 1980, 1984, and 1989 (Table 1). Forty-two percent of these individuals were male. The median age for both sexes combined was 75 years. Only 81% (n=296) of persons with first stroke were either hospitalized for the event or had the event during a hospitalization. There were 436 hospitalizations in the year after stroke for all 4 years combined. Of the 364 incidence cases, 313 (86%) were hospitalized at least once during the 12 months that followed stroke; the remaining 123 hospitalizations were for readmission.

Discharge diagnosis codes were obtained for all hospitalizations for each of the 313 individuals hospitalized at least once during the year. The proportion of hospitalized individuals that received a discharge diagnosis code of ICD 430-438.9 is shown as a function of the number of diagnoses considered in Table 2. When the analysis was limited to the principal discharge diagnosis (ie, the condition that, after review, led to the hospitalization), 76% of the hospitalized cases were assigned to codes 430-438.9.

The sensitivity was less for fatal events. Of the 364 first strokes, 73 (20%) were fatal, ie, death occurred within 28 days after stroke. Patients with fatal strokes were significantly less likely to be hospitalized than those with nonfatal strokes (55/73=76% versus 258/391=89%, P<.01) and therefore were less likely to receive a principal diagnosis code of 430-438.9 (40/73=55% versus 198/291=68%, P<.05). Ten of the 73 fatal strokes were events that occurred after admission to the hospital. Strokes that occurred after admission were also less likely to be identified with the principal discharge diagnosis. Of 40 individuals who experienced an incident stroke (either fatal or nonfatal) after hospital admission, only 25% were assigned a principal discharge diagnosis of 430-438.9 in the year after stroke.

As shown in Table 2, the sensitivity of discharge abstracts increased as the number of discharge diagnoses under consideration was expanded from one to three, with lesser gains thereafter. The detection of incidence strokes improved in 1989 relative to earlier years. A number of factors, including changes in technology, practice style, and reimbursement may have contributed to this finding. Of particular interest is the finding that differences over time were minimized as the number of diagnoses under consideration increased.

The second part of this study examined the positive predictive value of hospital discharge abstracts. Positive predictive value was defined as the proportion of individuals with an ICD discharge code of 430-438.9 for whom the hospitalization was confirmed to be associated with a stroke event. This part of the study used the Olmsted County Hospital Utilization database, with non-Rochester residents deleted. Data were available for the years 1970, 1980, and 1989.

Findings for positive predictive value are presented in Table 3. For all 3 years combined, there were 377 hospitalizations of 352 Rochester residents, for which the principal diagnosis code was 430-438.9. These hospitalizations were first matched against the Rochester Stroke Registry to identify first-stroke events. Hospitalizations not identified in this manner were categorized, based on the neurologist’s (R.D.B.) review of the individual’s complete medical record, as either incident stroke, recurrent stroke, stroke sequelae, or a nonstroke...
TABLE 2. Proportion of Persons With a First Stroke Event Who Were Hospitalized at Least Once in 12 Months After Stroke With Discharge Diagnosis ICD Codes 430-438.9, as a Function of the Number of Diagnoses Considered

<table>
<thead>
<tr>
<th>Year of Stroke</th>
<th>Persons hospitalized†</th>
<th>Events coded 430-438.9‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970</td>
<td>56</td>
<td>239 (76)</td>
</tr>
<tr>
<td>1980</td>
<td>74</td>
<td>271 (86)</td>
</tr>
<tr>
<td>1984</td>
<td>92</td>
<td>283 (90)</td>
</tr>
<tr>
<td>1989</td>
<td>91</td>
<td>290 (93)</td>
</tr>
</tbody>
</table>

ICD indicates International Classification of Diseases; Dx, diagnosis. Numbers in parentheses indicate percentage.

†Persons with a first stroke event were identified using the Rochester Stroke Registry.

‡Data on inpatient activity in the 12 months after stroke for 1970, 1980, and 1984 cases were abstracted from the community-based unit medical records. Data on inpatient activity in the 12 months after stroke for 1989 incidence cases were obtained from the complete 1989 and 1990 Olmsted County Hospital Utilization Data Set.

For individuals with more than one hospitalization, the analysis considered the first hospitalization with a code for stroke in the 12 months after stroke.

A number of studies using hospital discharge abstracts have recognized that the category 430-438.9 includes nonstroke codes and that the positive predictive power would be improved by eliminating these codes from consideration. The comparison with the stroke registry and the review by the neurologist afforded analysis of discrimination at the level of three-digit ICD codes ("Appendix"). The year 1970 was analyzed separately because of changes in ICD coding between 1970 and 1980 and differences in the number of diagnoses available. Data from 1980 and 1989 are shown in Table 4.

None of the events determined to be incident or recurrent stroke by the neurologist were assigned code 432 (other and unspecified intracranial hemorrhage, eg, nontraumatic subdural hematoma) or code 438 (late effects) on the discharge abstract. Of the 78 hospitalizations assigned code 435 (transient cerebral ischemia), only 12 were clinically consistent with a stroke event. When these three codes were excluded from the category 430-438.9, the positive predictive value for incident stroke rose from 47% to 63%. Data from 1970 (not shown) exhibit a similar pattern, showing higher positive predictive value after codes 432 (occlusion of precerebral arteries) and 435 (transient cerebral ischemia) were eliminated.

TABLE 3. Positive Predictive Value of Hospital Discharge Abstracts With Discharge Diagnosis ICD Codes 430-438.9 for Identifying Stroke Events

<table>
<thead>
<tr>
<th>Year of Stroke</th>
<th>1970</th>
<th>1980</th>
<th>1989</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>1 Dx</td>
<td>3 Dx</td>
<td>1 Dx</td>
<td>3 Dx</td>
</tr>
<tr>
<td>Hospitalizations</td>
<td>88</td>
<td>124</td>
<td>122</td>
<td>144</td>
</tr>
<tr>
<td>Individuals</td>
<td>81</td>
<td>110</td>
<td>117</td>
<td>133</td>
</tr>
<tr>
<td>Incident strokes</td>
<td>41</td>
<td>48</td>
<td>62</td>
<td>66</td>
</tr>
<tr>
<td>PPV, %</td>
<td>46</td>
<td>39</td>
<td>50</td>
<td>46</td>
</tr>
<tr>
<td>Incident or recurrent strokes</td>
<td>52</td>
<td>60</td>
<td>78</td>
<td>84</td>
</tr>
<tr>
<td>PPV, %</td>
<td>59</td>
<td>48</td>
<td>64</td>
<td>58</td>
</tr>
</tbody>
</table>

ICD indicates International Classification of Diseases; Dx, diagnosis; PPV, positive predictive values.

*Hospitalizations with a discharge diagnosis code ICD 430-438.9 were identified using the Olmsted County Hospital Utilization Data Set for the calendar years 1970, 1980, and 1989. Non-Rochester Olmsted County residents were excluded from the analysis.
TABLE 4. Cross-Classification of Individuals With a Discharge Diagnosis of Cerebrovascular Disease (ICD-9 Codes 430-438.9)* by Categorization of Stroke Versus Nonstroke Based on Neurologist's Review

<table>
<thead>
<tr>
<th>Neurologist's Determination</th>
<th>ICD-9 Code</th>
<th>430</th>
<th>431</th>
<th>432</th>
<th>433</th>
<th>434</th>
<th>435</th>
<th>436</th>
<th>437</th>
<th>438</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. incident stroke</td>
<td></td>
<td>11</td>
<td>17</td>
<td>0</td>
<td>3</td>
<td>70</td>
<td>9</td>
<td>43</td>
<td>1</td>
<td>0</td>
<td>154</td>
</tr>
<tr>
<td>No. recurrent stroke</td>
<td></td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>3</td>
<td>12</td>
<td>1</td>
<td>0</td>
<td>36</td>
</tr>
<tr>
<td>No. sequelae or nonstroke</td>
<td></td>
<td>0</td>
<td>3</td>
<td>7</td>
<td>17</td>
<td>15</td>
<td>66</td>
<td>9</td>
<td>7</td>
<td>15</td>
<td>139</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>11</td>
<td>23</td>
<td>7</td>
<td>20</td>
<td>102</td>
<td>78</td>
<td>64</td>
<td>9</td>
<td>15</td>
<td>329</td>
</tr>
<tr>
<td>Incident, %</td>
<td></td>
<td>100</td>
<td>74</td>
<td>0</td>
<td>15</td>
<td>69</td>
<td>12</td>
<td>67</td>
<td>11</td>
<td>0</td>
<td>47</td>
</tr>
<tr>
<td>Incident or recurrent, %</td>
<td></td>
<td>100</td>
<td>87</td>
<td>0</td>
<td>15</td>
<td>85</td>
<td>15</td>
<td>86</td>
<td>22</td>
<td>0</td>
<td>58</td>
</tr>
</tbody>
</table>

ICD indicates International Classification of Diseases.
*Hospitalizations with a discharge diagnosis code ICD 430-438.9 were identified using the Olmsted County Hospital Utilization Data Set for the calendar years 1970, 1980, and 1989. Non-Rochester Olmsted County residents were excluded from the analysis. Data are for 1980 and 1989 combined. Up to five diagnoses were considered. The first code for stroke was used for hospitalizations with >1 stroke code on the discharge abstract. The first hospitalization was used for individuals with >1 hospitalization.

excluded. There is no code for “late effects” in the Eighth Revision International Classification of Diseases Adapted for Use in the United States (ICDA) used in 1970.

Incidence rates obtained using the Rochester Stroke Registry were compared with those obtained using hospital discharge abstracts. Two different approaches to using discharge abstracts were selected. The first approach simulated data available and criteria used in studies that rely on the National Center for Health Statistics Hospital Discharge Survey. All hospitalizations during the year were considered (ie, multiple hospitalizations of the same individual were not distinguished); the investigation was limited to the principal discharge diagnosis, and all codes within the category 430-438.9 were included. The second approach identified multiple hospitalizations by the same individual. Based on findings reported above (Tables 2 through 4), the number of diagnoses was expanded to three, and codes that were poorly predictive of incident stroke (432, 435, and 438) were excluded. For individuals with more than one stroke-coded hospitalization or more than one stroke code per hospitalization, the analysis considered the code with the highest positive predictive value as shown in Table 3. The results of the comparison are presented in Table 5. It should be emphasized that this study is descriptive and covers the Rochester population completely (ie, no sampling was involved). The presentation of confidence limits is intended to reflect the natural variability inherent in the population.

Using the first approach, the incidence rate of stroke in Rochester, Minn, for 1980 and 1989 combined was calculated to be 231 per 100 000 person-years (95% confidence interval [CI], 205 to 260), 1.5 times the rate of 151 strokes per 100 000 person-years (95% CI, 131 to 174) calculated using Rochester Stroke Registry incidence cases for 1980 and 1989. In addition to overestimating incidence, there were serious problems of miscalculation with this approach. Only 69% of all incident strokes were identified, and only 47% of the discharges were associated with incident stroke. The incidence rate of 177 strokes per 100 000 person-years (95% CI, 155 to

TABLE 5. Comparison of Various Approaches Using Stroke-Coded Hospital Discharge Abstracts With Stroke Registry

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hospitalizations* Principal Dx Codes 430-438.9</th>
<th>Individuals† Three Dx Codes 430,431, 433,434,436,437</th>
<th>Stroke Registry</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. cases</td>
<td>Incidence/100 000 person-years</td>
<td>95% CI</td>
</tr>
<tr>
<td></td>
<td>289</td>
<td>231</td>
<td>205-260</td>
</tr>
<tr>
<td></td>
<td>Incidence/100 000 person-years</td>
<td>231</td>
<td>177</td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>205-260</td>
<td>155-202</td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>95% CI</td>
<td>131-174</td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Including nonhospitalized</td>
<td>130/189 (69)</td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Hospitalized only</td>
<td>130/165 (79)</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>Incident stroke</td>
<td>137/289 (47)</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>Incident and recurrent stroke</td>
<td>178/289 (62)</td>
</tr>
</tbody>
</table>

Dx indicates diagnosis; CI, confidence interval; and PPV, positive predictive value. Data are from 1980 and 1989 combined. Numbers in parentheses indicate percentage.
*This approach did not identify individuals. All stroke-coded hospitalizations during the year were considered.
†This approach identified individuals. For cases with more than one stroke-coded hospitalization or more than one stroke code per hospitalization, the analysis considered the code with the highest PPV.
202) using the second approach was closer to the Rochester Stroke Registry rate. This improvement is somewhat misleading. Discharge abstracts still failed to capture 23% of all cases of first stroke occurring among Rochester residents, and review of the medical records revealed that 36% of the cases identified using this approach were either recurrent stroke (15%) or non-stroke events (21%).

The problems of misclassification are apparent with comparisons of inpatient charges. Consideration of all hospitalizations with a principal discharge diagnosis of 430-438.9 in 1989 revealed an average charge of $8308 per hospitalization. This compares with an average charge of $10,598 for the first hospitalization for individuals with any of three discharge diagnoses coded 430, 431, 433, 434, 436, or 437. Both figures are less than the $11,615-per-event hospitalization charge for the 87 Rochester patients with stroke who were either hospitalized for the event or had their first stroke during the hospital stay. It is important to note that the 21 non-hospitalized cases were not considered in this comparison and cannot be considered in any analysis that relies solely on hospital discharge abstracts. These individuals will be included in an analysis of average annual inpatient, outpatient, and long-term costs of care after stroke (in preparation).

Discussion

The present study compares a community-based stroke registry (consisting of both hospitalized and nonhospitalized stroke patients) with stroke-coded hospital discharge abstracts for the same population from which the registry was drawn. The study examines multiple years, covering two decades of experience. The findings demonstrate that the use of all hospital discharge abstracts with a principal diagnosis code of 430-438.9 results in a significant overestimate of stroke incidence. Forty-one percent of all first strokes among Rochester, Minn, residents were missed using this approach, and over half of the hospitalizations coded as stroke were associated with an event other than an incident stroke. A marked improvement in accuracy was achieved by expanding the diagnoses under consideration to three, eliminating multiple hospitalizations for the same individual, and excluding codes for which the positive predictive value was poor. This approach resulted in a stroke incidence rate closer to that observed in the Rochester Stroke Registry. Problems of misclassification, however, still remained: 23% of incident strokes were unidentified, and 21% of stroke-coded diagnoses were found to be other than incident or recurrent stroke on review.

The “gold standard” for this analysis was the Rochester Stroke Registry. The medical records linkage system of the Rochester Epidemiology Project and the resulting diagnostic index serve as the major sources of potential cases that, after medical record review and confirmation, are included in the Rochester Stroke Registry. The index consists of diagnoses recorded from every encounter (inpatient, outpatient, emergency room, home visit, nursing home, autopsy, and death). Because the index includes all diagnoses recorded during the time an individual is residing in the area, diagnoses made after the acute episode are also captured for review. The completeness of ascertainment varies depending on the condition and the length of follow-up available. It is most complete for serious chronic conditions or life-threatening events accompanied by sequelae. The essentially complete ascertainment afforded by the medical records linkage system for stroke was validated in an earlier study that compared stroke incidence rates as estimated using the Rochester Stroke Registry with rates obtained from prospective follow-up of an initially stroke-free cohort. The cohort analysis included 1796 Rochester residents followed up for 13 years of observation. Five individuals were identified as having stroke who were not detected by the medical records linkage analysis; however, none of the five was a Rochester resident at the time of the stroke, so their exclusion from the registry was appropriate. Ultimately, of course, the extent to which any retrospectively obtained estimates, whether from hospital discharge abstracts, medical records–linked indices, or thorough record review, can serve as a “gold standard” is limited by clinical knowledge and documentation.

The scarcity of comparable community-based registries in the United States makes it difficult to assess the generalizability of the sensitivity estimates presented here. A study from Sweden compared discharge abstracts with a WHO MONICA (World Health Organization Monitoring of Trends and Determinants in Cardiovascular Disease) registry and found only 6% of true strokes were missing from hospital discharge abstracts compared with the 23% reported here (Table 5). Although the definitions of confirmed stroke and the exclusion of nonstroke codes were similar between the two studies, the MONICA registry was limited to persons 25 through 74 years of age. As shown in Table 1, approximately half the incident strokes in Rochester occurred among persons aged 75 years and older. The Swedish analysis considered all listed diagnoses compared with a maximum of five diagnoses in this study. The most likely explanation for the lower proportion of unidentified strokes in the Swedish study, however, is the relatively small proportion of nonhospitalized stroke events in that country. Only 3.2% of all patients with nonfatal stroke in the MONICA registry were not hospitalized for the event. By contrast, 11% (33/291) of patients with nonfatal first strokes and 14% (51/364) of all patients with first strokes in Rochester were not hospitalized in the 12 months after stroke.

The proportion of stroke-coded hospitalizations found to be nonstroke or stroke sequelae in the Swedish study was greater than that reported here (32% versus 21%). The most frequently occurring false-positive events in the Swedish study were those without documented focal neurological deficit or those for whom deficits lasted <24 hours. Again, some of the discrepancy in study findings may be attributable to the greater number of diagnoses considered in the Swedish study. As demonstrated here, improvements in sensitivity achieved by increasing the number of diagnoses are partially offset by declines in positive predictive value. The Swedish data do not include hospitalizations that ended in death. The extent to which these differences, in addition to variations in coding practices between the two countries, might contribute to a higher false-positive finding in Sweden compared with Rochester is not known.
Low positive predictive values have also been reported for studies of hospitalized stroke in the United States. In the Minnesota Heart Survey study, McGovern et al. found that the occurrence of acute stroke-coded discharges that failed to meet their criteria for definite stroke ranged from 69% in 1970 to 50% in 1985. Again, there were differences in study design: all listed diagnoses were considered, the subjects were less than 75 years of age, and the criteria for definite stroke were more rigorous than those applied here. However, these factors are unlikely to account completely for the differences in study findings. Rochester, Minn, has 70,000 residents and is serviced by one of the world's largest medical centers. One explanation for the low rate of false-positive codes reported for Rochester compared with those for Minneapolis/St Paul or Sweden is a relatively higher likelihood that a resident of Rochester who is hospitalized with a diagnosis of stroke will be examined by a neurologist.

This explanation is supported when the distribution of events by three-digit code for Rochester residents (Table 4) is compared with that for all listed diagnoses for inpatients discharged from short-stay hospitals in the United States.17,18 For 1980 and 1989 combined, approximately 45% of US stroke-coded discharges were assigned codes of either 433, 437, or 438 (ie, "other cerebrovascular disease"). Only 13% of Rochester residents with a stroke-coded hospitalization were assigned codes 433, 437, or 438. Table 4 also shows that 29% of confirmed stroke events (5% of the 190 incident or recurrent strokes) were assigned code 436 (acute but ill-defined cerebrovascular disease) on the discharge abstract. This contrasts with fewer than 10% of strokes in the Rochester Stroke Registry for which the neurologist's review determined that there was insufficient information to assign a specific type.19 These findings confirm that discharge abstracts are of limited value for distinguishing stroke by type.1,20 They suggest as well that the clinically meaningful categories of "infarction," "hemorrhage," and "not otherwise specified" may not translate easily to the current ICD categorization system.

In an effort to increase the specificity of discharge abstracts, some authors have elected to exclude "nonstroke codes" from their analyses. Modan and Wagener excluded "late effects" (438 in ICD-9) in their analysis of secular trends in stroke incidence. The Minnesota Heart Survey excluded codes 433 and 438 and, in some studies, code 435.13 Studies using diagnosis related groups (DRGs) have focused on DRG 14. DRG 14 relies on principal diagnosis, excludes codes 433 and 435 (occlusion of precerebral arteries and transient ischemic attack), and includes code 784.3 (aphasia).4 Studies also differ with regard to the number of diagnoses that were examined. In the present study, it was possible to compare the accuracy of alternative approaches by varying both the number of discharge diagnoses and the specific codes under consideration.

The generalizability of the conclusions is limited by the very few events in each ICD category and by the fact that coding practices vary between institutions. The number of diagnoses and specific codes presented here should not be interpreted as a recommendation but rather as a model for how the sensitivity and positive predictive value of discharge abstracts can be increased.

Despite the capacity for improvement, the problems of identification and misclassification inherent with hospital discharge abstracts are not easily resolved. As this study shows, fatal events and events that occur in-hospital are less likely to be identified. The average inpatient charge is also underestimated. As revealed in an earlier study using the Rochester Stroke Registry, the proportion of persons with first stroke hospitalized for the event and the proportion of persons with first stroke occurring in-hospital have changed over time.21 Therefore, studies of temporal trends in stroke incidence, severity, and cost of care that rely solely on hospital discharge abstracts may be misleading.

The distinction between incident and recurrent events using discharge abstracts is also problematic. The present study, using three discharge diagnoses, found that 20% (51/249) of confirmed stroke events were recurrent stroke (Table 3). Limiting the analysis to the principal diagnosis only reduced the number of recurrent events by three. Some studies using hospital discharge abstracts have had access to the event-hospitalization record for all or a sample1 of cases. Confirmation of "first stroke" is facilitated with this additional information. The determination of incidence in the Rochester Stroke Registry is based on review of the medical record. The record is essentially complete for the time that the person was a resident of the community; the average residency for persons in the registry is longer than 35 years. In the absence of these resources, there is evidence to suggest that the precision of identifying incident events may be increased by eliminating from the analysis all persons with a hospitalization with a principal diagnosis of stroke during the 5 years preceding the index stroke.22

Even after nonstroke codes were excluded from the analysis, 21% of stroke-coded hospitalizations were determined to be nonstroke events (Table 5). This figure is perhaps optimistic, given the low frequency of false-positive stroke diagnoses relative to studies in other communities where use of diagnostic technology and neurological expertise is likely to be less than in Rochester, Minn. Misdiagnosis, inadequate documentation, and poor discrimination afforded by ICD-coded discharge abstracts contribute to the misclassification. At the very least, studies that rely on hospital discharge abstracts should obtain for review the hospital records for a sample of discharges to determine which codes from the category 430-438.9 can be eliminated from the analysis.

Clinical practice, public policy, and reimbursement decisions are increasingly informed by data on disease prevalence, associated service use, and costs that are obtained from administrative data sets. Unfortunately, there are very few community-based disease registries in the United States with utilization data that can serve as a standard against which to measure the accuracy of this information. The present study provides previously unavailable estimates of the sensitivity of stroke-coded hospitalizations for a US community. It presents a model for increasing the sensitivity and positive predictive value of discharge abstracts and demonstrates that the improvements can be substantial. In this way, it helps contribute to our understanding of the validity and usefulness of administrative data sets.
Appendix

Reference for 1970 Hospitalizations: ICDA (Eighth Revision International Classification of Diseases Adapted for Use in the United States)

Cerebrovascular Disease (430-438)

430 Subarachnoid hemorrhage
431 Cerebral hemorrhage
432 Occlusion of precerebral arteries
433 Cerebral thrombosis
434 Cerebral embolism
435 Transient cerebral ischemia
436 Acute but ill-defined cerebrovascular disease
437 Generalized ischemic cerebrovascular disease
438 Other and ill-defined cerebrovascular disease


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

Acknowledgments

This study was supported in part by National Institutes of Health research grants NS-0663 and AR-30582 and by a FIRST Award from the National Institute on Aging (Dr Leibson, AG-08729). The authors thank Dr Walter Rocca and the reviewers of this manuscript for their helpful comments. We are grateful to LaVonne Gates, Ann Ryan, and Carol Siverling for abstracting assistance.

References

Accuracy of hospital discharge abstracts for identifying stroke.
C L Leibson, J M Naessens, R D Brown and J P Whisnant

Stroke. 1994;25:2348-2355
doi: 10.1161/01.STR.25.12.2348
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 1994 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/25/12/2348

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/