Epidemiology of Recurrent Cerebral Infarction
A Medicare Claims–Based Comparison of First and Recurrent Strokes on 2-Year Survival and Cost

Gregory P. Samsa, PhD; John Bian, MS; Joseph Lipscomb, PhD; David B. Matchar, MD

Background and Purpose—Because recurrent strokes will tend to leave patients with greater disability than first strokes, patients with recurrent strokes should have poorer outcomes on average than those with first strokes. The extent of this difference has, however, not yet been estimated with precision.

Methods—Using a random 20% sample of Medicare patients aged 65 years and older admitted with a primary diagnosis of cerebral infarction during calendar year 1991, we used historical data from the previous 4 years to classify patients as having either first or recurrent stroke and followed survival and direct medical costs for 24 months after stroke. First and recurrent stroke groups were compared with the log-rank test (survival) and t test (cost) and also multivariate modeling.

Results—Survival from first stroke is consistently better than that for recurrent stroke: 24-month survival was 56.7% versus 48.3%, respectively. Costs were similar for the initial hospital stay and in months 1 to 3 after stroke. During months 4 to 24 after stroke, total costs were higher among those with recurrent stroke by approximately $375/mo across all patients, with this difference being greatest for younger patients and least for patients aged 80 years or older. Most of the difference in total monthly cost was attributable to nursing home utilization (averaging approximately $150/mo) and acute hospitalization (averaging approximately $120/mo).

Conclusions—Patients with recurrent stroke have, on average, poorer outcomes than those with first stroke. To be as accurate as possible, clinical policy analyses should use different estimates of health and cost outcomes for first and recurrent stroke. (Stroke. 1999;30:338-349.)

Key Words: cerebral infarction ■ costs and cost analysis ■ epidemiology ■ mortality

From the clinical and public health perspectives, recurrent strokes are important. The risk of subsequent stroke is highest in the period immediately after a stroke,1–4 and thus 1 of the clinical goals of acute and long-term management of stroke is to prevent its recurrence. Also, because a recurrent stroke will tend to leave patients with greater disability than a first stroke, from the public health perspective it follows that one would expect that patients with recurrent strokes will have poorer health and economic outcomes (on average) than those with first strokes. However, few studies have compared first and recurrent strokes, and thus the extent of the differences between these populations remains a topic for examination.

From a public health perspective, another reason for giving careful consideration to recurrent strokes is their effect on estimates of the overall national burden of stroke. For example, Broderick et al5 argue that the epidemiology of total stroke (ie, first stroke plus recurrent stroke) is a better index of the overall burden of stroke than whites in Rochester, Minn, then inflating this baseline by figures reflecting both the recurrent stroke rate and the higher rate of stroke among African Americans, the estimated annual burden of stroke in the United States is not the commonly cited 500 000 cases per year6 but instead is at least 731 000 cases.5 Apart from its immediate implications for the annual incidence of stroke, such a difference would also have a profound impact on estimates (for example, the total medical costs among a cohort of patients with incident strokes during a given year), which are derived from incidence. In addition, if outcomes of first and recurrent strokes differ markedly, it would be preferable that policy analyses (eg, decision and cost-effectiveness models) reflect these differences (for example, by using different estimates of health and cost outcomes for first and recurrent stroke).

With the use of a large nationally representative cohort of patients obtained from Medicare claims files, the primary goal of this study was to contrast outcomes (24-month
survival and direct medical costs) of patients with first and recurrent stroke.

Subjects and Methods

As described elsewhere, we selected from the Medicare Provider Analysis and Review files a random 20% sample of Medicare patients aged 65 years and older admitted with a primary diagnosis of cerebral infarction during calendar year 1991. (Medicare hospital insurance covers approximately 95% of Americans aged 65 years and older. Medicare Part A covers services in acute care hospitals, rehabilitation facilities, skilled nursing facilities, hospital outpatient care, and home health care. Medicare Part B covers services provided by physicians and various nonphysician specialists (eg, physical therapists), outpatient diagnostic and laboratory services, and certain medical equipment. Coverage under Part A and Part B is subject to patient cost-sharing and other limitations. Not covered at all are long-term nursing home care and outpatient drugs. Patients were excluded from the analysis if enrolled in a health maintenance organization or not covered by Part B, which is optional; in either case, the claims data submitted to Medicare are likely to be incomplete.)

Cerebral infarction (hereafter referred to simply as stroke) was operationalized to include occlusion of cerebral arteries (International Classification of Diseases, 9th Revision (ICD-9) code 434) and acute but ill-defined cerebrovascular disease (ICD-9 code 436). All Medicare claims for these patients were then extracted until December 31, 1993. We also extracted any acute hospital records during the preceding 4 years with a principal diagnosis of cerebral infarction, transient ischemic attack (TIA; ICD-9 code 435), or hemorrhagic stroke (ICD-9 codes 430 to 432). Data on date of death (if applicable) were obtained from denominator files provided by the Health Care Financing Administration.

We used the 4-year historical data described above to classify stroke patients into 2 groups: first strokes and recurrent strokes. All of the patients in the sample had at least 1 hospitalization for cerebral infarction during 1991 (the first such hospitalization being denoted as the “index hospitalization”). Patients in the recurrent stroke group also had 1 or more hospital stays during the period 1987 to 1990 with a primary diagnosis of cerebral infarction or hemorrhagic stroke, whereas patients in the first stroke group did not. All other patients, including those with a history of TIA but not stroke, were placed in the first stroke group. Patients were followed for 24 months beyond the index hospitalization for utilization and survival, even if subsequent strokes were observed during follow-up.

The primary outcome variables were mortality and direct medical costs. Mortality was based on date of death as described above. Costs included the costs of the initial hospitalization and were then grouped into monthly intervals. The first month extended for 30 days, whereas a patient dying during the follow-up period only contribute person-months of follow-up during the time they are alive. For example, a patient dying during month 2 would contribute 2 person-months of follow-up to the interval of 1 to 3 months, whereas a patient surviving beyond month 3 would contribute 3 person-months. Costs for the first and recurrent stroke groups were compared using a t test applied to the log-transformed data (with patients with no costs assigned a cost of $1 to avoid taking the logarithm of zero). A nonparametric analysis with the Wilcoxon test produced similar results (data not shown).

It should be noted that the survival and cost calculations use slightly different time frames. For survival, we are using the traditional Kaplan-Meier framework, which takes as time 0 the date of the stroke. For costs, the index hospitalization is treated separately, and time 0 is the date of discharge from this hospital stay. This distinction was made to most clearly separate costs of the initial hospitalization from subsequent costs.

The first and recurrent stroke groups were also compared with the use of multivariate analyses, which accounted (to the degree possible with Medicare claims files) for differences in case-mix between the 2 groups. Variables used in the case-mix adjustment included age, race, gender, various diagnoses (hypertension, congestive heart failure, myocardial infarction, chronic obstructive pulmonary disease, and valvular heart disease), history of TIA, history of previous stroke, and Charlson comorbidity score (with the latter score being the weighted sum of selected conditions shown to be associated with poor outcomes). Survival was compared using the Cox proportional hazards model and cost was compared with linear ANCOVA.

The software package SAS, version 6.12, was used for all analyses.

Results

The 1991 Medicare sample contained 49,333 patients hospitalized with a primary diagnosis of cerebral infarction. Of these, 44,386 (90.0%) had no record of being hospitalized for a previous stroke during the preceding 4 years, whereas 4947 (10.0%) had a record of at least 1 previous stroke. Of those patients with previous stroke, 4225 (85.4%) had a record of exactly 1 previous stroke, 621 (12.5%) had a record of exactly 2 previous strokes, 79 (1.6%) had a record of exactly 3 previous strokes, and the remaining 22 patients (0.5%) had a record of 4 or more previous strokes. All subsequent analyses
compared the group with no previous strokes (ie, “first strokes”) with those in the group with 1 or more previous strokes (ie, “recurrent strokes”). Among patients with recurrent stroke, the last such recorded stroke occurred in 1987 for 13.9% of patients, in 1988 for 17.7% of patients, in 1989 for 24.9% of patients, and in 1990 for 43.6% of patients.

As illustrated in Table 1, in the group of patients with first stroke, the average age was 78.8 years (±7.5), 40.4% were male, 85.0% were white, and 3.6% had a recorded history of TIA during the preceding 4 years. In the group of patients with recurrent stroke, the average age was 79.0 years (±6.9), 38.2% were male, 82.7% were white, and 9.7% had a recorded history of TIA. The first and recurrent stroke groups had similar distributions of hypertension, congestive heart failure, diabetes mellitus, myocardial infarction, chronic obstructive pulmonary disease, and valvular heart disease.

Figures 1 to 5 compare the survival of patients with first and recurrent stroke. The probability of surviving for at least 2 years after stroke is higher for females and for younger patients. In absolute terms, survival from first stroke is consistently better than that for recurrent stroke (for example, the 2-year survival from first stroke for males aged 65 to 69 years was 66.5% compared with 51.1% for recurrent stroke; Figure 1; \( P < 0.001 \) by log-rank test). These figures were 64.0% versus 53.6% (Figure 2; \( P < 0.001 \)) for males aged 70 to 74 years, 58.2% versus 52.3% (Figure 3; \( P = 0.019 \)) for males aged 75 to 79 years, 51.7% versus 45.0% (Figure 4; \( P = 0.005 \)) for males aged 80 to 84 years, 38.6% versus 32.7% (Figure 5; \( P = 0.257 \)) for males aged 85 years or older, 68.0% versus 53.5% (Figure 1; \( P < 0.001 \)) for females aged 65 to 69 years, 68.9% versus 58.3% (Figure 2; \( P < 0.001 \)) for females aged 70 to 74 years, 63.5% versus 50.4% (Figure 3; \( P < 0.001 \)) for females aged 75 to 79 years, 54.1% versus 45.4% (Figure 4; \( P < 0.001 \)) for females aged 80 to 84 years, and 42.0% versus 40.6% (Figure 5; \( P = 0.370 \)) for females aged 85 years or older. Survival differences for first and recurrent strokes are typically manifest within 1 to 3 months after stroke and then widen (in absolute terms) as patients are followed over time. The relationship between history of stroke and 2-year mortality is consistent across age and gender, with the exception of patients aged 85 years or older (for whom the effect of stroke history is weakest).

Table 2 presents data on cost. For the initial hospitalization, patients with recurrent stroke had longer stays than patients with first stroke by 0.7 days on average, but total costs were similar (\( P > 0.05 \) for 7 of 10 age-gender strata with a \( t \) test on log-transformed costs). Costs were also similar during months 1 to 3 after stroke (\( P > 0.05 \) for 4 of 10 age-gender strata). During months 4 to 12 and 13 to 24 after stroke, total monthly costs for patients with recurrent stroke were higher (\( P < 0.001 \) for all age-gender strata), with this difference in cost greatest for younger patients and least for patients 80 years of age or older. The average difference in cost during months 4 to 24 after stroke was approximately $375 per patient per month, most of which was attributable to the categories of nursing home (approximately $150/mo) and acute hospitalization (approximately $120/mo; see Figure 6). We do not know how much the observed difference in nursing home costs was attributable to patients who were previously living at home versus patients who were previously living in nursing homes (with the latter presumably more common in the recurrent stroke group).

Table 3 presents the results of multivariate modeling, which examines the effect of history of previous stroke after accounting for case-mix. This analysis confirms the previous results. The additional sample size available to the multivariate modeling in comparison with the stratum-specific analyses also allowed the effect of history of previous stroke to be examined in greater detail. For survival and costs, a dose-response was observed, in that (after adjusting for case mix) patients with a history of more than 1 previous stroke had on average lower survival and higher costs than patients with a history of exactly 1 previous stroke, who had on average lower survival and higher costs than patients without a history of previous stroke.

### Discussion

We found that patients with recurrent stroke had poorer survival than those with first stroke: at 24 months, the figures were 56.7% and 48.3%, respectively. In addition, patients with recurrent strokes had higher costs in months 4 to 24 after stroke, with this difference amounting to approximately $375 per patient per month on average. Cost differences between patients with first and those with recurrent stroke were especially noticeable for nursing home stays and rehospitalizations and were less pronounced for patients 80 years of age and older. The observation that patients with recurrent stroke have (on average) poorer survival and higher costs than those with first stroke was confirmed by multivariate modeling, accounting for case-mix. These multivariate analyses also indicated a dose-response in that patients with a history of more than 1 previous stroke had (on average) poorer outcomes than patients with a history of exactly 1 previous stroke, who in turn had (on average) poorer outcomes than patients with first stroke.

Our basic approach was very similar to that of May and colleagues,\textsuperscript{12} who also used Medicare claims files and found that in 1985 the (unadjusted) 24-month survival for patients with and without previous stroke was 51.6% and 57.9%,
Figure 1. Survival with first and recurrent strokes by gender, 65 to 69 years of age. The x axis indicates the time since stroke onset and the y axis indicates the proportion of the overall cohort surviving. For recurrent stroke, the Kaplan-Meier estimate of the survival curve (lines) and upper (open squares) and lower 95% (closed circles) confidence limits are plotted. For first stroke, the Kaplan-Meier estimate of the survival curve is plotted (open triangles). Confidence limits are not plotted for first strokes, because these survival curves are based on large sample sizes, and the confidence limits are virtually indistinguishable from the estimated survival curve. Note that the lower boundary for the y axis is 30%.
Figure 2. Survival with first and recurrent strokes by gender, 70 to 74 years of age. The x axis indicates the time since stroke onset and the y axis indicates the proportion of the overall cohort surviving. For recurrent stroke, the Kaplan-Meier estimate of the survival curve (lines) and upper (open squares) and lower 95% (closed circles) confidence limits are plotted. For first stroke, the Kaplan-Meier estimate of the survival curve is plotted (open triangles). Confidence limits are not plotted for first strokes, because these survival curves are based on large sample sizes, and the confidence limits are virtually indistinguishable from the estimated survival curve. Note that the lower boundary for the y axis is 30%.
Figure 3. Survival with first and recurrent stroke by gender, 75 to 79 years of age. The x axis indicates the time since stroke onset and the y axis indicates the proportion of the overall cohort surviving. For recurrent stroke, the Kaplan-Meier estimate of the survival curve (lines) and upper (open squares) and lower 95% (closed circles) confidence limits are plotted. For first stroke, the Kaplan-Meier estimate of the survival curve is plotted (open triangles). Confidence limits are not plotted for first strokes, because these survival curves are based on large sample sizes, and the confidence limits are virtually indistinguishable from the estimated survival curve. Note that the lower boundary for the y axis is 30%.
Survival with First and Recurrent Strokes by Gender, Age 80-84

Figure 4. Survival with first and recurrent stroke by gender, 80 to 84 years of age. The x axis indicates the time since stroke onset and the y axis indicates the proportion of the overall cohort surviving. For recurrent stroke, the Kaplan-Meier estimate of the survival curve (lines) and upper (open squares) and lower 95% (closed circles) confidence limits are plotted. For first stroke, the Kaplan-Meier estimate of the survival curve is plotted (open triangles). Confidence limits are not plotted for first strokes, because these survival curves are based on large sample sizes, and the confidence limits are virtually indistinguishable from the estimated survival curve. Note that the lower boundary for the y axis is 30%.
Figure 5. Survival with first and recurrent stroke by gender, 85 years of age or older. The x axis indicates the time since stroke onset and the y axis indicates the proportion of the overall cohort surviving. For recurrent stroke, the Kaplan-Meier estimate of the survival curve (lines) and upper (open squares) and lower 95% (closed circles) confidence limits are plotted. For first stroke, the Kaplan-Meier estimate of the survival curve is plotted (open triangles). Confidence limits are not plotted for first strokes, because these survival curves are based on large sample sizes, and the confidence limits are virtually indistinguishable from the estimated survival curve. Note that the lower boundary for the y axis is 30%. 
respectively. In 1989, these figures were 50.3% and 59.3%, respectively. Thus, our overall conclusions about survival were quite similar to those of previous reports. May and colleagues did not analyze costs.

The pattern of outcomes described above is consistent with patients with recurrent stroke having, on average, greater levels of disability than those with first strokes. If so, one might not necessarily expect dramatic differences in acute treatment but would expect more nursing home stays, more rehospitalizations, and poorer survival (eg, due to the sequelae of inactivity). The attenuated differences for patients 80 years of age and older are consistent with this hypothesis, assuming that very elderly patients with first stroke tend to leave the hospital with relatively high levels of disability (either from the stroke itself or other sources). Lower rehabilitation costs during months 1 to 3 after stroke for patients with recurrent stroke are also consistent with this hypothesis, because rehabilitation efforts tend to be most concentrated on the “middle band” of patients with moderate functional deficits rather than those with very severe deficits.

Both this study and that of May et al12 rely on Medicare files, which trade off large sample sizes and excellent generalizability against the amount and quality of the information per patient. Ideally, one would like to compare claims-based results with those from various population-based cohorts (which, although small and of limited generalizability, tend to have extensive and high-quality information on individual patients). Unfortunately, most of the literature (in particular, reports from the Framingham and Rochester, Minn, cohorts on which much of our detailed knowledge of the epidemiology of stroke in the United States is based) tend to present data for either all strokes or first strokes only but do not follow outcomes for recurrent strokes separately.3,4 Much of the difficulty can be attributed to sample size, because less than half of all strokes represent recurrent strokes.13

This study has a number of limitations that are generic to the analysis of claims data. First, the level of reliable clinical detail is slight. In particular, Medicare files do not include crucial information about stroke severity or subsequent disability. Also, the variables used for case-mix adjustment can be both inconsistent and incomplete. For example, because of limitations on the number of positions for ICD-9 codes, diagnoses such as hypertension may be less likely to be coded.

### Table 2. Costs of Patients With First and Recurrent Strokes, by Age and Sex Gender Categories

<table>
<thead>
<tr>
<th>Category</th>
<th>n</th>
<th>Age Initial LOS, d</th>
<th>Initial Hospital Cost, $</th>
<th>Month 1–3, $</th>
<th>Month 4–12, $</th>
<th>Month 13–24, $</th>
</tr>
</thead>
<tbody>
<tr>
<td>First strokes overall</td>
<td>44</td>
<td>78.8 (7.5)</td>
<td>10.9 (10.4)</td>
<td>7091 (8599)</td>
<td>3368 (5953)</td>
<td>1361 (2948)</td>
</tr>
<tr>
<td>Males, 65–69 y</td>
<td>2804</td>
<td>67.4 (1.3)</td>
<td>10.2 (10.9)</td>
<td>7752 (11 063)</td>
<td>3095 (6112)</td>
<td>1087 (3258)</td>
</tr>
<tr>
<td>Males, 70–74 y</td>
<td>4121</td>
<td>72.0 (1.4)</td>
<td>10.4 (11.1)</td>
<td>7339 (10 807)</td>
<td>3221 (5752)</td>
<td>1136 (3166)</td>
</tr>
<tr>
<td>Males, 75–79 y</td>
<td>4480</td>
<td>77.0 (1.4)</td>
<td>10.6 (10.0)</td>
<td>7505 (9679)</td>
<td>3501 (6261)</td>
<td>1257 (3198)</td>
</tr>
<tr>
<td>Males, 80–84 y</td>
<td>3591</td>
<td>81.9 (1.4)</td>
<td>11.0 (10.6)</td>
<td>7142 (7559)</td>
<td>3376 (5481)</td>
<td>1384 (3055)</td>
</tr>
<tr>
<td>Males, ≥85 y</td>
<td>2916</td>
<td>88.4 (3.1)</td>
<td>10.8 (9.2)</td>
<td>6643 (6926)</td>
<td>3448 (5555)</td>
<td>1599 (2918)</td>
</tr>
<tr>
<td>Females, 65–69 y</td>
<td>2724</td>
<td>67.5 (1.2)</td>
<td>10.5 (10.0)</td>
<td>7504 (9735)</td>
<td>3428 (6007)</td>
<td>1190 (3093)</td>
</tr>
<tr>
<td>Females, 70–74 y</td>
<td>4293</td>
<td>72.1 (1.4)</td>
<td>10.7 (10.2)</td>
<td>7129 (8807)</td>
<td>3440 (5832)</td>
<td>1237 (2840)</td>
</tr>
<tr>
<td>Females, 75–79 y</td>
<td>5754</td>
<td>77.1 (1.4)</td>
<td>10.9 (10.4)</td>
<td>7162 (8918)</td>
<td>3569 (5892)</td>
<td>1273 (2840)</td>
</tr>
<tr>
<td>Females, 80–84 y</td>
<td>6018</td>
<td>82.0 (1.4)</td>
<td>11.3 (10.0)</td>
<td>6972 (7034)</td>
<td>3455 (5224)</td>
<td>1357 (2610)</td>
</tr>
<tr>
<td>Females, ≥85 y</td>
<td>7685</td>
<td>89.1 (3.4)</td>
<td>11.3 (10.9)</td>
<td>6495 (6662)</td>
<td>3134 (4459)</td>
<td>1656 (2395)</td>
</tr>
<tr>
<td>Recurrent strokes overall</td>
<td>4947</td>
<td>79.0 (6.9)</td>
<td>11.6 (11.0)</td>
<td>6939 (7883)</td>
<td>3315 (4955)</td>
<td>1700 (3039)</td>
</tr>
<tr>
<td>Males, 65–69 y</td>
<td>225</td>
<td>67.8 (1.2)</td>
<td>11.1 (9.7)</td>
<td>7464 (9418)</td>
<td>3356 (5361)</td>
<td>1693 (3376)</td>
</tr>
<tr>
<td>Males, 70–74 y</td>
<td>459</td>
<td>72.1 (1.4)</td>
<td>11.7 (12.2)</td>
<td>7593 (9291)</td>
<td>3623 (5834)</td>
<td>1440 (3030)</td>
</tr>
<tr>
<td>Males, 75–79 y</td>
<td>518</td>
<td>77.0 (1.4)</td>
<td>10.5 (7.9)</td>
<td>6524 (5864)</td>
<td>3344 (4862)</td>
<td>1688 (3278)</td>
</tr>
<tr>
<td>Males, 80–84 y</td>
<td>418</td>
<td>81.9 (1.4)</td>
<td>11.8 (12.3)</td>
<td>6666 (8352)</td>
<td>2872 (4235)</td>
<td>1518 (2932)</td>
</tr>
<tr>
<td>Males, ≥85 y</td>
<td>269</td>
<td>88.1 (3.0)</td>
<td>11.6 (13.3)</td>
<td>6785 (8629)</td>
<td>3234 (4574)</td>
<td>1685 (2629)</td>
</tr>
<tr>
<td>Females, 65–69 y</td>
<td>217</td>
<td>67.9 (1.1)</td>
<td>12.2 (10.1)</td>
<td>8132 (9209)</td>
<td>3890 (5334)</td>
<td>1678 (2907)</td>
</tr>
<tr>
<td>Females, 70–74 y</td>
<td>513</td>
<td>72.1 (1.4)</td>
<td>11.4 (9.2)</td>
<td>7081 (6615)</td>
<td>3594 (5168)</td>
<td>1652 (2901)</td>
</tr>
<tr>
<td>Females, 75–79 y</td>
<td>722</td>
<td>77.2 (1.4)</td>
<td>12.0 (11.5)</td>
<td>7298 (8956)</td>
<td>3513 (5205)</td>
<td>1856 (3210)</td>
</tr>
<tr>
<td>Females, 80–84 y</td>
<td>768</td>
<td>81.9 (1.4)</td>
<td>11.5 (9.7)</td>
<td>6509 (8547)</td>
<td>3094 (4373)</td>
<td>1660 (2418)</td>
</tr>
<tr>
<td>Females, ≥85 y</td>
<td>838</td>
<td>88.7 (3.2)</td>
<td>12.1 (12.8)</td>
<td>6570 (7623)</td>
<td>3023 (3818)</td>
<td>1915 (3400)</td>
</tr>
</tbody>
</table>

Values in parentheses represent SDs. For months 1–3, 4–12, and 13–24, costs are reported as dollars per patient month.
suggest that 25% to 35% of strokes are recurrent,6,13 although (The limited data available from population-based cohorts underestimate the proportion of strokes that are recurrent. Thus, an unknown percentage of patients that we classified as having first stroke actually had recurrent strokes, and a (presumably much smaller) percentage of patients that we classified as having recurrent stroke actually had first stroke or no stroke at all. The crucial assumption underlying our analysis is that the pattern of outcomes among the strokes that we classify as recurrent is representative of the pattern of outcomes among all recurrent strokes.

As described in the “Methods,” most categories of cost are based directly on Medicare charges, whereas others amalgamate estimates from other sources. We particularly note that nursing home costs, which were a key component of the difference in cost between first and recurrent strokes, were not directly based on observed nursing home utilization but rather on an algorithm that in turn was primarily based on the overall pattern of Part B physician claims (for which nursing home was the recorded site of service). Even though the above algorithm is unvalidated, we have no reason to believe that it should perform differently for patients with recurrent strokes than for patients with first stroke (thus, even if the algorithm over- or under-counts costs, it should do so similarly in both groups). Stroke-related drug costs were also based on a multistep algorithm, but these costs were relatively small and similar for first and recurrent strokes. Our cost data were skewed and quite variable (as is often the case for costs); thus, although the conclusions hold at a group level, individual patients may have costs that are quite different from those reported here.

With respect to policy modeling, the primary implication of these findings is that decision and cost-effectiveness models should use different estimates of survival and cost outcomes depending on whether the patient has a history of stroke. Survival should be greatest for patients with first stroke, least for patients with recurrent stroke, and in between if a general population of stroke patients (ie, regardless of history of previous stroke) is considered. Because survival for stroke patients overall is essentially a weighted average of the survival for first and recurrent stroke patients, the survival curve for the general stroke population will be very similar to that of first strokes if the proportion of patients with recurrent stroke is small, and the curves will increasingly diverge as the proportion of recurrent strokes increases. A similar set of relationships holds for per diem costs; these will be lowest for patients with first stroke, highest for patients with recurrent stroke, and in between for the general population of stroke patients. The magnitude of the error that results from the use of the incorrect base population will be greatest when the survival and cost experience of first strokes (or all strokes) are applied to patients with recurrent stroke.

Among patients with many other diagnoses; and, indeed, the effects of the hypertension variable in the multivariate analyses were opposite to clinical intuition. More generally, we do not have clinical confirmation regarding the presence or absence of various diagnoses (for example, history of TIA is likely underreported). Finally, some additional variables such as presence of atrial fibrillation would have been helpful to include in the case-mix adjustment process. It should be recognized that even in the best case case-mix adjustment from administrative claims files will be incomplete, and caution in interpretation is indicated.

A second limitation involves the potential misclassification of patients with stroke. Our strategy was designed to limit the number of false positives (ie, nonstroke patients incorrectly classified as having a stroke) by the using ICD-9 codes shown to have a good yield in correctly identifying stroke cases (ie, ICD-9 codes 434 and 436 but not 433) and also by basing patient classification on the primary diagnosis only.6,14,15 Although this approach will tend to limit the number of false positives, it may also tend to disproportionately exclude patients with complicated and severe strokes for whom other conditions may have been used as the primary diagnosis.

One implication of the above strategy is that it will tend to underestimate the proportion of strokes that are recurrent. (The limited data available from population-based cohorts suggest that 25% to 35% of strokes are recurrent,6,13 although we classified only 10% of strokes as such.) In particular, we will miss all strokes that occurred more than 4 years before the beginning of follow-up. (This interval is less for patients who are close to 65 years of age, because Medicare claims would typically not be available before 65 years of age.) Also, we will miss any strokes that were coded as anything other than the primary diagnosis, any instances of multiple strokes within the same hospitalization, and any strokes that do not lead to hospitalization (ie, in the United States approximately 15% to 25% of strokes).16 Thus, an unknown percentage of the incorrect base population will be greatest when the assumption underlying our analysis is that the pattern of outcomes among the strokes that we classify as recurrent is representative of the pattern of outcomes among all recurrent strokes.

### Table 3. Multivariate Modeling of Survival and Cost

<table>
<thead>
<tr>
<th>Survival HR</th>
<th>P</th>
<th>Cost Beta</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>1.07</td>
<td>&lt;0.001</td>
<td>-52</td>
</tr>
<tr>
<td>Nonwhite</td>
<td>1.08</td>
<td>&lt;0.001</td>
<td>306</td>
</tr>
<tr>
<td>Age, y</td>
<td>1.07</td>
<td>&lt;0.001</td>
<td>-37</td>
</tr>
<tr>
<td>70–74</td>
<td>1.00</td>
<td>&lt;0.001</td>
<td>58</td>
</tr>
<tr>
<td>75–79</td>
<td>1.19</td>
<td>&lt;0.001</td>
<td>193</td>
</tr>
<tr>
<td>8–84</td>
<td>1.47</td>
<td>&lt;0.001</td>
<td>93</td>
</tr>
<tr>
<td>≥85</td>
<td>1.99</td>
<td>&lt;0.001</td>
<td>361</td>
</tr>
<tr>
<td>Charlson comorbidity</td>
<td>0.91</td>
<td>&lt;0.001</td>
<td>226</td>
</tr>
<tr>
<td>Hypertension</td>
<td>0.61</td>
<td>&lt;0.001</td>
<td>-126</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>1.76</td>
<td>&lt;0.001</td>
<td>158</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>1.14</td>
<td>&lt;0.001</td>
<td>57</td>
</tr>
<tr>
<td>COPD</td>
<td>0.74</td>
<td>&lt;0.001</td>
<td>95</td>
</tr>
<tr>
<td>History of TIA</td>
<td>0.05</td>
<td>0.163</td>
<td>319</td>
</tr>
<tr>
<td>History of previous stroke</td>
<td>1.22</td>
<td>&lt;0.001</td>
<td>57</td>
</tr>
</tbody>
</table>

COPD indicates chronic obstructive pulmonary disease; HR, hazard ratio. Survival rate to 24 month survival and is modeled using the Cox proportional hazards model. Cost refers to mean monthly cost during months 13–24 and is modeled using linear ANCOVA. Results for log-transformed cost were substantively similar and are not presented. For categorical predictors the reference categories are as follows: female gender, white race/ethnicity, age 65–69 years, hypertension absent, congestive heart failure absent, diabetes mellitus absent, myocardial infarction absent, COPD absent, valvular heart disease absent, no history of TIA, and no history of previous stroke. These regressions included all of the above variables, regardless of statistical significance.

Among patients with many other diagnoses; and, indeed, the effects of the hypertension variable in the multivariate analyses were opposite to clinical intuition. More generally, we do not have clinical confirmation regarding the presence or absence of various diagnoses (for example, history of TIA is likely underreported). Finally, some additional variables such as presence of atrial fibrillation would have been helpful to include in the case-mix adjustment process. It should be recognized that even in the best case case-mix adjustment from administrative claims files will be incomplete, and caution in interpretation is indicated.

A second limitation involves the potential misclassification of patients with stroke. Our strategy was designed to limit the number of false positives (ie, nonstroke patients incorrectly classified as having a stroke) by the using ICD-9 codes shown to have a good yield in correctly identifying stroke cases (ie, ICD-9 codes 434 and 436 but not 433) and also by basing patient classification on the primary diagnosis only. Although this approach will tend to limit the number of false positives, it may also tend to disproportionately exclude patients with complicated and severe strokes for whom other conditions may have been used as the primary diagnosis.

One implication of the above strategy is that it will tend to underestimate the proportion of strokes that are recurrent. (The limited data available from population-based cohorts suggest that 25% to 35% of strokes are recurrent, although we classified only 10% of strokes as such.) In particular, we will miss all strokes that occurred more than 4 years before the beginning of follow-up. (This interval is less for patients who are close to 65 years of age, because Medicare claims would typically not be available before 65 years of age.) Also, we will miss any strokes that were coded as anything other than the primary diagnosis, any instances of multiple strokes within the same hospitalization, and any strokes that do not lead to hospitalization (ie, in the United States approximately 15% to 25% of strokes). Thus, an unknown percentage of patients that we classified as having first stroke actually had recurrent strokes, and a (presumably much smaller) percentage of patients that we classified as having recurrent stroke actually had first stroke or no stroke at all. The crucial assumption underlying our analysis is that the pattern of outcomes among the strokes that we classify as recurrent is representative of the pattern of outcomes among all recurrent strokes.

As described in the “Methods,” most categories of cost are based directly on Medicare charges, whereas others amalgamate estimates from other sources. We particularly note that nursing home costs, which were a key component of the difference in cost between first and recurrent strokes, were not directly based on observed nursing home utilization but rather on an algorithm that in turn was primarily based on the overall pattern of Part B physician claims (for which nursing home was the recorded site of service). Even though the above algorithm is unvalidated, we have no reason to believe that it should perform differently for patients with recurrent strokes than for patients with first stroke (thus, even if the algorithm over- or under-counts costs, it should do so similarly in both groups). Stroke-related drug costs were also based on a multistep algorithm, but these costs were relatively small and similar for first and recurrent strokes. Our cost data were skewed and quite variable (as is often the case for costs); thus, although the conclusions hold at a group level, individual patients may have costs that are quite different from those reported here.

With respect to policy modeling, the primary implication of these findings is that decision and cost-effectiveness models should use different estimates of survival and cost outcomes depending on whether the patient has a history of stroke. Survival should be greatest for patients with first stroke, least for patients with recurrent stroke, and in between if a general population of stroke patients (ie, regardless of history of previous stroke) is considered. Because survival for stroke patients overall is essentially a weighted average of the survival for first and recurrent stroke patients, the survival curve for the general stroke population will be very similar to that of first strokes if the proportion of patients with recurrent stroke is small, and the curves will increasingly diverge as the proportion of recurrent strokes increases. A similar set of relationships holds for per diem costs; these will be lowest for patients with first stroke, highest for patients with recurrent stroke, and in between for the general population of stroke patients. The magnitude of the error that results from the use of the incorrect base population will be greatest when the survival and cost experience of first strokes (or all strokes) are applied to patients with recurrent stroke.
As an example of these calculations, suppose we assume that there are at least 100,000 Americans with recurrent stroke during a given year (ie, at least 500,000 new strokes, with at least 20% of them being recurrent). The difference in 2-year survival between first and recurrent strokes is 8.4% (ie, 56.7% versus 48.3%), so incorrectly applying the figures for first stroke rather than recurrent stroke implies an error in the number of persons surviving at least 2 years after stroke of approximately 100,000×0.084=8400. Making a similar calculation for costs, the difference in cost is approximately $375 per patient per month in months 4 to 24 after stroke. Recognizing that not all patients survive 24 months and using the figure of 15 months as the approximate expected survival during months 4 to 24 after stroke, the difference in cost is of the order of magnitude of 100,000×15×$375=$562 million.

These findings can also be applied to the task of estimating the overall national burden of stroke. To do this, one would first estimate the proportion of strokes that are recurrent, then apply separate cost- and survival-based functions for first and recurrent strokes, and finally combine these in a weighted-average calculation as described above. This procedure is illustrated in another manuscript under preparation.

In conclusion, patients with recurrent stroke have, on average, poorer outcomes than those with first stroke. To be as accurate as possible, policy and other analyses should use different patterns of health and cost outcomes for first and recurrent strokes.

Acknowledgments

Funding was provided by the Agency for Health Care Policy and Research, Contract No. 282-81-0028, Stroke Prevention Patient Outcomes Research Team (PORT), Washington DC. The original data file was provided by Health Economics Research Inc (Janet Mitchell, PhD, project director; Ann Larsen, programmer).

Appendix

Algorithm for Estimating Nursing Home Costs

Because Medicare pays for SNF days on a limited basis and does not cover non-SNF nursing home care at all, Medicare claims data will provide an incomplete accounting of the total nursing home costs for many stroke patients. In response, we assigned nursing home costs to each patient, for each month the patient was observed after discharge from the index hospital admission for stroke, on the basis of (1) the volume of Medicare-covered SNF costs generated each month as reported through Part A and (2) the algorithm outlined below. This algorithm uses Medicare Part A and Part B claims data in combination to impute the costs associated with SNF days not paid for by Medicare, as well as all non-SNF nursing home days. It builds on an earlier algorithm developed by Berg and colleagues17 for defining nursing home episodes; see especially step (3). But it differs from the approach of Berg et al in several respects, including the handling of “unknown” months, ie, those for which claims data are either absent or too sparse to permit a direct inference about nursing home use; see step (7).

Before presenting the algorithm, we must operationally define what is meant here by a Part B “nursing home” claim. It is any claim for which the site-of-care code indicates the service was rendered at a “nursing home” (code 31) or a “skilled nursing facility” (code 32), or for which the procedure code on the billing form indicates a nursing home (NH) or SNF physician service. For 1991, the relevant codes are 90300, 90315, 90340, 90350, 90360, and 90370. For 1992 and 1993, the relevant codes are 99001, 99030, 9903, 99311, 99312, and 99313.

The algorithm to assign nursing home costs by patient and month proceeds in 8 steps:

1. If there are 1 or more Part B NH claims during month t and no other Medicare encounter during month “t,” assign patient to nursing home for month “t.” The other encounter or encounters may be indicated by Part A expenditures for SNF; other institutional services, such as acute care or rehabilitation hospitalization; or noninstitutional services such as home health care or hospital outpatient visits (OPD).

2. If there is a Part B NH claim during month “t” and 1 or more Part A Medicare encounters during this month, assign patient to nursing home for a fraction of the month equal to N/N, where N is the number of distinct encounter types. For example, if the claims data pertaining to month “t” for a patient show both an acute care hospitalization and a NH visit (as inferred from the presence of a Part B NH record), there are 2 encounter types, leading to an assignment of 15 days of NH use.

3. If there is a Part B NH claim during month “t” and no Medicare encounter (Part A or Part B) during month t+1, assign patient to nursing home for month t+1. If, in addition, there is no Medicare encounter in month t+2, assign patient to nursing home for month t+2. If in addition there is no Medicare encounter in month t+3, assign patient to unknown (UKN). That is, a patient is assigned to NH status for at most 60 days (2 months) beyond the month when the NH stay starts, if there are no additional confirmatory data on NH admission. If another Medicare encounter does arise, use rule in step (4) below to apportion NH and non-NH costs over that month.

4. Suppose patient has a Part B NH claim in month “t” but no other Medicare encounter claim. Then, in month t+1, there is no Part B NH claim, but there is another type of claim, eg, home health visit. Then assign NH days in month t+1 according to step (2).

5. Assume that there is a Part B NH claim in month “t,” followed in month t+1 by a hospital outpatient claim (OPD) only (there are no other claim types in month t+1). Then ignore OPD here in allocating days in month t+1 to NH and UKN according to step (3). This is to allow for the not unusual case in which the NH patient receives outpatient services away from the NH.

6. Assume that there is a Part B NH claim in month “t,” followed by admission to an acute care hospital in month t+1. If UKN in month t+2, assign month t+2 to NH. If UKN in month t+3, assign t+3 to NH. If UKN in month t+4, assign t+4 to UKN, and so on.

7. Each UKN month is split into NH days and days at home by the following rule:

7a. For any unknown month “t,” define its “neighborhood” as the 3 immediately preceding months and the 3 immediately following months.

7b. Let X be the number of months in the neighborhood of “t” for which a Part B NH claim or SNF Part A claim provides direct evidence that the individual is in a NH.

7c. Let Y=(total months in the neighborhood of t—the number of other UKN months in the neighborhood of t—administratively censored months in the neighborhood).

7d. The probability of NH days imputed to UKN month “t” = prod(NH)=(X/Y).

7e. Apply prod(NH) to allocate the 30 days of UKN month “t” into 2 parts, NH days and at-home days. For example, if prob(NH)=0.12 for a particular UKN month, then 3.6 days are assigned to NH and 26.4 days are assumed to be at home. Because each UKN has its own estimate of prob(NH), the latter varies not only across patients but within each patient according to month.

8. Let NHit be the number of nursing home days assigned to month “t” for individual “i” according to steps (1) through (7) Let CPD be the estimated average per diem cost for non-Medicare-covered NH days, computed here to be $75 (1991 dollars). This figure emerges as a weighted average of the per diem cost estimate for a SNF ($115) and non-SNF NH ($70), with the ratio of non-SNF to SNF days of 9:1.

The SNF per diem cost was computed from Medicare data available to the Stroke Prevention PORT. It is based on SNF cost report data reported to the Health Care Financing Administration and is designed to reflect economic costs, not charges or Medicare’s...
reimbursement levels. The non-SNF nursing home per diem is roughly equal to Medicaid’s national average per diem reimbursement to non-SNF nursing homes in 1991; it is assumed here that Medicaid payments are an adequate approximation of per diem economic costs. The 9:1 ratio was derived by combining data on Medicare and Medicaid per diem reimbursement rates in 1995, from the Health Care Financing Administration, with an estimate of the national average per diem payment across all nursing beds in 1995, as reported by the Health Insurance Association of America. It was assumed that this 1995-based ratio was applicable in 1991.

One general caveat with this algorithm should be noted. There is the possibility that some SNF nursing home care reflected in Part A payments will be separately picked up by Part B claims, leading to double counting. However, the most straightforward way to correct this, removing SNF Part B claims from the algorithm, would likely lead to a more serious problem: failure to detect all SNF days not covered by Medicare.

References
Epidemiology of Recurrent Cerebral Infarction: A Medicare Claims–Based Comparison of First and Recurrent Strokes on 2-Year Survival and Cost
Gregory P. Samsa, John Bian, Joseph Lipscomb and David B. Matchar

Stroke. 1999;30:338-349
doi: 10.1161/01.STR.30.2.338

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/30/2/338

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/