Validity of Claims-Based Stroke Algorithms in Contemporary Medicare Data

Reasons for Geographic and Racial Differences in Stroke (REGARDS) Study Linked With Medicare Claims

Hiraku Kumamaru, MD, MPH; Suzanne E. Judd, PhD, MPH; Jeffrey R. Curtis, MD, MS, MPH; Rekha Ramachandran, MS; N. Chantelle Hardy, MPH; J. David Rhodes, BSN, MPH; Monika M. Safford, MD; Brett M. Kissela, MD, MS; George Howard, DrPH; Jessica J. Jalbert, PhD; Thomas G. Brott, MD; Soko Setoguchi, MD, DrPH

Background—The accuracy of stroke diagnosis in administrative claims for a contemporary population of Medicare enrollees has not been studied. We assessed the validity of diagnostic coding algorithms for identifying stroke in the Medicare population by linking data from the REasons for Geographic And Racial Differences in Stroke (REGARDS) Study to Medicare claims.

Methods and Results—The REGARDS Study enrolled 30,239 participants ≥45 years in the United States between 2003 and 2007. Stroke experts adjudicated suspected strokes, using medical records. We linked data for 15,089 participants, among whom 422 participants had adjudicated strokes during follow-up. An algorithm using primary discharge diagnosis codes for acute ischemic or hemorrhagic stroke (International Classification of Diseases, Ninth Revision, Clinical Modification codes: 430, 431, 433.x1, 434.x1, 436) had a positive predictive value of 92.6% (95% confidence interval, 88.8%–96.4%), a specificity of 99.8% (99.6%–99.9%), and a sensitivity of 59.5% (53.8%–65.1%). An algorithm using only acute ischemic stroke codes (433.x1, 434.x1, 436) had a positive predictive value of 91.1% (95% confidence interval, 86.6%–95.5%), a specificity of 99.8% (99.7%–99.9%), and a sensitivity of 58.6% (52.4%–64.7%).

Conclusions—Claims-based algorithms to identify stroke in a contemporary Medicare cohort had high positive predictive value and specificity, supporting their use as outcomes for etiologic and comparative effectiveness studies in similar populations. These inpatient algorithms are unsuitable for estimating stroke incidence because of low sensitivity.

Key Words: cohort studies • comparative effectiveness research, diagnosis • health services research • Medicare

Previous studies have assessed the accuracy of diagnoses of stroke in administrative claims databases.1–12 Relatively high positive predictive values (PPVs) ranging from 70% to 96% have been reported, especially in recent studies using fifth-digit International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes, supporting the use of these algorithms for the identification of stroke in the databases. The most recent studies in Medicare patients include one in a population of kidney transplant recipients in 200413 and another in patients with atrial fibrillation in 1998–1999.8 Improvements in modalities for diagnosis of stroke have changed clinical practice during the last 15 years,14,15 and changes to ICD-9-CM codes have also occurred.16 Thus, the validity of stroke ascertainment algorithms in a contemporary population might differ from that suggested by older studies. High specificity and PPVs are crucial for the validity of estimates of etiologic and comparative effectiveness studies when these claims-based algorithms are used to identify outcome events.17–19

Investigators for the REasons for Geographic And Racial Differences in Stroke (REGARDS) Study, a nationwide epidemiological study of 30,239 participants, recently linked the cohort to Medicare claims data. This linkage provided...
WHAT IS KNOWN
- Administrative data are being increasingly used for etiologic and comparative effectiveness studies to capture longitudinal clinical events.
- High specificity is essential for the result of these studies to be valid, but no evidence exists for their validity in the contemporary Medicare population.

WHAT THE STUDY ADDS
- Medicare data can be linked to a cohort study data with high accuracy and can capture additional outcomes to supplement the cohort study.
- The algorithms using inpatient discharge diagnoses had high specificity and positive predictive value, supporting their use for outcome identification in etiologic and comparative effectiveness studies.
- Further research is needed to identify optimal algorithms with higher sensitivity for the use of these databases for incidence rate and utilization studies.

a unique opportunity to assess the validity of claims-based algorithms for stroke diagnosis in Medicare beneficiaries. The purpose of the current study is to assess the validity of the claims-based algorithms and illustrate the types of analyses in which these algorithms are useful in the contemporary Medicare population.

Methods

Data Sources
We used linked REGARDS-Medicare claims data for the study. The REGARDS data contained information on 30,239 community-dwelling participants ≥45 years recruited throughout the United States between 2003 and 2007, with oversampling of black participants and those living in the Stroke Belt, a region with a particularly higher incidence rate of stroke compared with the rest of the country (Alabama, Arkansas, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, and Tennessee). The study methods have been described previously by Howard et al.20 Participants who agreed to be enrolled completed a 45-minute telephone interview to collect demographic, socioeconomic, risk factor, and medical history information at baseline. A health professional then completed a visit to collect blood and urine samples, blood pressure measurements, ECG, and other key study variables. Participants were followed through phone interviews performed every 6 months. Medical records for self-reported, suspected stroke events and stroke symptoms were retrieved whenever possible and centrally adjudicated for stroke events by a team of trained stroke experts.

Medicare is the primary health insurer of the US population ≥65 years,21 and administrative data for fee-for-service enrollees, who constitute >80% of the total Medicare population,22 are collected and distributed by the Centers for Medicare & Medicaid Services for research use. The database contains final action claims data submitted by inpatient hospital providers, skilled nursing facility providers, institutional outpatient providers, noninstitutional providers, and durable medical equipment suppliers for reimbursement of services. The information contained in these files includes diagnoses and procedures, dates of service, reimbursement amounts, provider information, Medicare eligibility and enrollment information, and demographic and vital status characteristics.

For the current study, REGARDS data were linked to Medicare data for 2003 through 2009. In the entire REGARDS cohort, 26,659 (88.2%) participants provided their Social Security numbers and consented to linkage with Medicare data. REGARDS participants were linked based on exact agreement in Social Security number and sex and an exact match to ≥2 of 3 components of date of birth. Mismatch in the year of birth was allowed to differ by a maximum of 1 year. As a result of this procedure, a total of 17,942 participants were linked to Medicare data, resulting in a linkage rate of 91.2% (16,583/18,190) for those who were ≥65 years at any time during the study period.

Study Cohort
We restricted the study population in the linked data set to participants with ≥1 month of eligibility in fee-for-service Medicare Part A and Part B after REGARDS enrollment (therefore excluding Medicare Advantage enrollment periods). We constructed 2 cohorts for the study: an any-stroke cohort and a first-stroke cohort. The any-stroke cohort included all participants who met the criteria for the study population, and was used to conduct analyses on combined incident and recurrent stroke events. The first-stroke cohort was further restricted to those free of self-reported physician-diagnosed stroke in REGARDS before cohort entry to limit the analysis to incident cases (Figure 1).

Claims-Based Algorithms
We assessed the validity of the following inpatient claims-based algorithms using the Medicare institutional files: (1) the acute ischemic stroke (AIS) algorithm included ICD-9-CM code 433.x1, 434.x1, or 436 in the primary discharge diagnosis; (2) the intracranial hemorrhage (ICH) algorithm included code 430 or 431 in the primary discharge diagnosis; and (3) the AIS/ICH algorithm included codes 430, 431, 433.x1, 434.x1, or 436 in the primary discharge diagnosis. These algorithms using only primary discharge diagnosis codes have been reported to have high PPVs,23,24 and we expected high specificity and low sensitivity from them. Codes for transient ischemic attack were not included.
identified those with stroke claims in the Medicare database using the AIS/ICH algorithm and pursued for their medical records. After being reviewed by a stroke nurse for exclusion of obvious noncases, all retrieved medical records were reviewed by a committee of stroke experts. Stroke was defined according to the World Health Organization definition as rapidly developing clinical signs of focal, at times global, disturbance of cerebral function, lasting >24 hours or leading to death with no apparent cause other than vascular.22 When events did not meet this definition but had symptoms lasting <24 hours with neuroimaging consistent with acute ischemia or hemorrhage, they were classified as clinical strokes. Strokes were further classified as ischemic or hemorrhagic. Both World Health Organization–defined and clinical stroke cases were included as the gold standard. The event status was considered missing if medical chart could not be retrieved for adjudication.

Statistical Analysis
Baseline characteristics and age at cohort entry were summarized using the data collected in the REGARDS Study. For the calculation of the validity measures, each participant of the cohort contributed 1 suspected event (ie, a self-report or a claim). Those with multiple suspected events contributed their earliest of the suspected events. Because the proportion of unretrieved medical charts (ie, those of unknown gold standard status) differed between participants with and those without the claims, we calculated the validity measures in a stratified manner to avoid bias from informative missingness. All participants were categorized into one of the following 4 strata based on how the participant’s suspected stroke events were identified: (A) claim/self-report: participants with suspected events identified by both claims and self-report during a follow-up telephone interview; (B) claims only: participants with suspected events identified by Medicare claims but not self-reported; (C) self-report only: participants with suspected events self-reported but not identified by claims; and (D) no events: participants without suspected stroke events in either claims data or self-report. We first calculated the proportion of suspected events adjudicated as strokes among all events adjudicated in each of the 4 strata described above. When there was an adjudicated stroke occurring within ≥30 days of the admission date of the claim, that claim was considered as having correctly identified the stroke (ie, true positives). For the primary analysis, we assumed that the proportion of true strokes was the same among those with and those without retrieved medical records (ie, assumed missing completely at random within the strata defined by the presence of self-report and claim) and estimated the overall PPV, sensitivity, specificity, and negative predictive value (NPV) under this assumption, applying strata-specific imputed means (percentage of suspected events adjudicated as strokes among those with retrieved records in the strata) to those without medical records. We calculated 95% confidence intervals by normal approximation. We also assessed whether these validity measures varied among subgroups of participants defined by categories of age at cohort entry, sex, and race. Among those who had an event identified by the AIS/ICH algorithm, we compared the baseline characteristics of the participants with and those of participants without retrieved medical records.

Sensitivity Analysis
To assess how dependent the estimates of the validity measures were on the assumption of true stroke proportion among the unknown status events, we estimated the PPV, sensitivity, and specificity of the AIS/ICH algorithm in any-stroke cohort under different assumed proportions of true strokes. While the optimal method for evaluating the algorithms would be to test them against the gold standard obtained in the whole group or a random sample regardless of self-report or claims, we were not able to do so because of limited resources for the medical chart review. We assumed that there are no unidentified cases in the no event stratum in the primary analysis. We assessed the impact of this assumption in the second sensitivity analysis, by estimating the PPV, sensitivity, and specificity of the AIS/ICH algorithm under different prevalence of uncaptured true strokes in the stratum.

The institutional review boards of the Duke University Health System and the University of Alabama at Birmingham approved the study. We used SAS Version 9.3 (SAS Institute, NC) for all analyses.
Results
Among the 17,942 participants linked to Medicare claims data, 15,089 had ≥1 month of eligibility for fee-for-service Medicare Parts A and B during follow-up, and were eligible for the any-stroke cohort. In total, 422 of the participants had ≥1 stroke during the follow-up. Restricting the population to those without baseline report of stroke or transient ischemic attack, the first-stroke cohort consisted of 13,096 participants.

In the any-stroke cohort, the mean age was 69.3 years, and >87% of participants were ≥65. The cohort had equal numbers of men and women, and 37% of participants were black (Table 1). Except for the presence of prior stroke or prior transient ischemic attack at baseline, participants in the first-stroke cohort had similar demographic and comorbidity profiles, with slightly lower proportions of comorbid conditions.

Among participants with no previously self-reported events, we identified additional 120 events among 97 participants, using Table 2.

Table 2. Adjudication Results for Each Algorithm

<table>
<thead>
<tr>
<th>Algorithm and Strata</th>
<th>Total n</th>
<th>Stroke (A)</th>
<th>No Stroke (B)</th>
<th>Missing</th>
<th>Proportion of True Cases Among Those Adjudicated A/(A+B)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Any-stroke cohort</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Claim/report</td>
<td>161</td>
<td>126 (78.3)</td>
<td>17 (10.6)</td>
<td>18 (11.2)</td>
<td>0.88</td>
</tr>
<tr>
<td>Claims-only</td>
<td>83</td>
<td>26 (31.3)</td>
<td>3 (3.6)</td>
<td>54 (65.1)</td>
<td>0.90</td>
</tr>
<tr>
<td>Report-only</td>
<td>1773</td>
<td>140 (7.9)</td>
<td>1485 (83.8)</td>
<td>148 (8.3)</td>
<td>0.09</td>
</tr>
<tr>
<td>No event</td>
<td>13,072</td>
<td>0</td>
<td>13,072 (100.0)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>15,089</td>
<td>292 (1.9)</td>
<td>14,577 (96.6)</td>
<td>220 (1.5)</td>
<td></td>
</tr>
<tr>
<td>ICH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Claim/report</td>
<td>26</td>
<td>23 (88.5)</td>
<td>3 (11.5)</td>
<td>0</td>
<td>0.88</td>
</tr>
<tr>
<td>Claims-only</td>
<td>15</td>
<td>8 (53.3)</td>
<td>1 (6.7)</td>
<td>6 (40.0)</td>
<td>0.89</td>
</tr>
<tr>
<td>Report-only</td>
<td>1925</td>
<td>16 (0.8)</td>
<td>1733 (90.0)</td>
<td>176 (9.1)</td>
<td>0.01</td>
</tr>
<tr>
<td>No event</td>
<td>13,123</td>
<td>0</td>
<td>13,123 (100.0)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>15,089</td>
<td>47 (0.3)</td>
<td>14,860 (98.5)</td>
<td>182 (1.2)</td>
<td></td>
</tr>
<tr>
<td>AIS/ICH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Claim/report</td>
<td>185</td>
<td>150 (81.2)</td>
<td>17 (9.2)</td>
<td>18 (9.7)</td>
<td>0.90</td>
</tr>
<tr>
<td>Claims-only</td>
<td>97</td>
<td>34 (35.1)</td>
<td>3 (3.1)</td>
<td>60 (61.9)</td>
<td>0.92</td>
</tr>
<tr>
<td>Report-only</td>
<td>1745</td>
<td>163 (9.2)</td>
<td>1270 (75.7)</td>
<td>312 (18.1)</td>
<td>0.10</td>
</tr>
<tr>
<td>No event</td>
<td>13,059</td>
<td>0</td>
<td>13,059 (100.0)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>15,089</td>
<td>337 (2.2)</td>
<td>14,526 (96.3)</td>
<td>226 (1.5)</td>
<td></td>
</tr>
<tr>
<td><strong>First-stroke cohort</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Claim/report</td>
<td>110</td>
<td>89 (80.9)</td>
<td>8 (7.2)</td>
<td>13 (11.8)</td>
<td>0.92</td>
</tr>
<tr>
<td>Claims-only</td>
<td>49</td>
<td>17 (34.7)</td>
<td>2 (4.1)</td>
<td>30 (61.2)</td>
<td>0.89</td>
</tr>
<tr>
<td>Report-only</td>
<td>1307</td>
<td>94 (7.2)</td>
<td>1106 (84.6)</td>
<td>107 (8.2)</td>
<td>0.08</td>
</tr>
<tr>
<td>No event</td>
<td>11,630</td>
<td>0</td>
<td>11,630 (100.0)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>13,096</td>
<td>200 (1.5)</td>
<td>12,746 (97.3)</td>
<td>150 (1.1)</td>
<td></td>
</tr>
<tr>
<td>ICH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Claim/report</td>
<td>22</td>
<td>19 (86.4)</td>
<td>3 (13.6)</td>
<td>0</td>
<td>0.86</td>
</tr>
<tr>
<td>Claims-only</td>
<td>7</td>
<td>4 (57.1)</td>
<td>1 (14.3)</td>
<td>2 (28.6)</td>
<td>0.80</td>
</tr>
<tr>
<td>Report-only</td>
<td>1406</td>
<td>15 (1.2)</td>
<td>1265 (90.0)</td>
<td>126 (9.0)</td>
<td>0.01</td>
</tr>
<tr>
<td>No event</td>
<td>11,661</td>
<td>0</td>
<td>11,661(100.0)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>13,096</td>
<td>38 (3.0)</td>
<td>12,930 (98.7)</td>
<td>128 (1.0)</td>
<td></td>
</tr>
<tr>
<td>AIS/ICH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Claim/report</td>
<td>130</td>
<td>109 (83.8)</td>
<td>8 (6.2)</td>
<td>13 (10.0)</td>
<td>0.93</td>
</tr>
<tr>
<td>Claims-only</td>
<td>55</td>
<td>21 (38.2)</td>
<td>2 (3.6)</td>
<td>32 (58.2)</td>
<td>0.91</td>
</tr>
<tr>
<td>Report-only</td>
<td>1287</td>
<td>107 (8.3)</td>
<td>1073 (83.4)</td>
<td>107 (8.3)</td>
<td>0.09</td>
</tr>
<tr>
<td>No event</td>
<td>11,624</td>
<td>0</td>
<td>11,624 (100.0)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>13,096</td>
<td>237 (1.8)</td>
<td>12,707 (97.0)</td>
<td>152 (1.2)</td>
<td></td>
</tr>
</tbody>
</table>

AIS indicates acute ischemic stroke; and ICH, intracranial hemorrhage.
the AIS/ICH algorithm. Medical records for 65 of the 120 events were successfully retrieved, and 48 of them were adjudicated as strokes. Among the rest, 7 (11%) cases which were mostly asymptomatic carotid procedure hospitalizations were triaged by the stroke nurse, 1 (2%) case was adjudicated as transient ischemic attack, and 9 (14%) cases as nonstrokes. In Table 2, we present the numbers of adjudicated strokes and nonstrokes after medical record review, as well as the number of unretrieved (missing) medical records in each category of claims/self-report for the 3 algorithms, for the first suspected event of each participant. For example, for the AIS algorithm, there were 161 participants with their first suspected events identified via both claim and self-report. Among them, 143 medical records (88.9%) were retrieved, and 126 of them were adjudicated as strokes, resulting in a PPV of 88% among the adjudicated events. Among the 161, we were unable to retrieve the medical records for 18 participants (11.2% missing). The proportion of cases adjudicated as strokes among those with retrieved medical charts was similar across the 3 algorithms, but differed by strata. Approximately 90% of events were adjudicated as strokes in the claims/self-report and claims-only strata, while 10% or fewer cases were adjudicated as strokes in the self-report only stratum (Table 2). For the AIS and AIS/ICH algorithms, the proportion of suspected events with unretrieved medical charts in each stratum was ≈10% for the claim/self-report and self-report only strata, and was =60% in the claims-only stratum. Using the AIS/ICH algorithm, 282 participants had stroke (Table 2). Among these 282 patients, those with unretrieved medical charts were on average younger and had substantially higher proportions of black participants, current smokers, and those with deep vein thrombosis and diabetes mellitus compared with those whose medical charts were retrieved (Table 3).

The 3 algorithms had high specificity and NPVs (Table 4). The PPVs of the algorithms were also high, ranging from 85% to 93%. Sensitivities were lower and fairly consistent across the 3 algorithms, ranging from 58% to 68% in the any-stroke cohort and from 58% to 60% in the first-stroke cohort. Differences in age, race, and sex had limited influence on the specificity and NPVs, whereas the estimated PPV and sensitivity varied more by these demographic characteristics; however, the 95% confidence intervals were wide and overlapped among most of the subgroups (Tables 5 and 6). Figure 2 shows the results of the sensitivity analysis of different proportions of true cases assumed in the unknown status cases in the claims-only stratum in the any-stroke cohort. Sensitivity and specificity were robust to the change in the assumption, while the PPV showed a slight change in the altered proportion. Because the proportion of suspected events with unretrieved medical records was relatively small in strata A and C (≈10%), changing the assumed true stroke frequency among them had a much smaller impact (data not shown). Altering the assumption on frequency of unidentified true strokes in the no-event stratum did not affect the PPV and minutely affected specificity (Figure 3). On the other hand, sensitivity was meaningfully influenced by the change in this assumption.

**Discussion**

Using the adjudicated stroke cases in the REGARDS Study as the gold standard in a linked REGARDS-Medicare claims database, we calculated sensitivity, specificity, PPV, and NPV for 3 claims-based algorithms to capture incident and recurrent strokes among contemporary Medicare beneficiaries. The algorithms, which used ICD-9-CM codes in the primary position of discharge diagnosis, identified strokes with very high specificity and NPVs, high PPVs, but lower sensitivity. Differences in age, sex, and race had limited influence on specificity and NPVs, but sensitivity and PPVs varied by these factors, although most of the confidence intervals overlapped among subgroups.

The most recent validation study of the claims-based stroke algorithms in the Medicare population, aside from
one in a specific population of kidney transplant recipients.\textsuperscript{13} was conducted among hospitalized atrial fibrillation patients in 1998 and 1999. \textsuperscript{8} The study reported a PPV of 96\% and a sensitivity of 35\%, for more inclusive algorithms including 437 and 438 codes. As the authors explained, the inclusion of prior (prevalent) strokes in the gold standard most likely contributed to the low sensitivity, and reduces the applicability of these measures to etiologic studies where the aim is to capture incidence of strokes as opposed to prevalence. The PPVs in our study were more comparable with the PPV

| Table 4. Accuracy of International Classification of Diseases, Ninth Revision, Clinical Modification Codes in the Primary Diagnosis Position |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| Cohort          | Algorithm       | PPV (95% CI)    | Sensitivity (95% CI) | Specificity (95% CI) | NPV (95% CI) |
| Any-stroke cohort | AIS             | 88.6 (84.7–92.6) | 58.6 (53.6–63.6) | 99.8 (99.7–99.9) | 99.0 (98.8–99.1) |
|                 | ICH             | 88.6 (79.8–96.3) | 67.4 (54.8–79.9) | 100.0 (99.9–100.0) | 99.9 (99.8–99.9) |
|                 | AIS/ICH         | 90.5 (87.1–94.0) | 60.4 (55.8–65.1) | 99.8 (99.6–99.9) | 98.9 (98.7–99.0) |
| First-stroke cohort | AIS          | 91.1 (86.6–95.5) | 58.6 (52.4–64.7) | 99.9 (98.8–100) | 99.2 (99.1–99.4) |
|                 | ICH             | 84.8 (71.8–97.9) | 59.9 (44.9–74.9) | 100.0 (99.9–100.0) | 99.9 (99.8–99.9) |
|                 | AIS/ICH         | 92.6 (88.8–96.4) | 59.5 (53.8–65.1) | 99.9 (98.8–100) | 99.1 (98.9–99.3) |

AIS indicates acute ischemic stroke; CI, confidence interval; ICH, intracranial hemorrhage; NPV, negative predictive value; and PPV, positive predictive value.

Table 5. Accuracy Measures Stratified by Age, Sex, and Race in Any-Stroke Cohort

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Category</th>
<th>n</th>
<th>PPV (95% CI), %</th>
<th>Sensitivity (95% CI), %</th>
<th>Specificity (95% CI), %</th>
<th>NPV (95% CI), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIS</td>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;65 y</td>
<td>1915</td>
<td>85.4 (70.6–100)</td>
<td>44.2 (29.3–59.1)</td>
<td>99.8 (99.6–100)</td>
<td>98.7 (98.2–99.2)</td>
</tr>
<tr>
<td></td>
<td>65–74 y</td>
<td>9713</td>
<td>90.7 (85.3–96.0)</td>
<td>56.0 (48.8–63.2)</td>
<td>99.9 (99.8–100)</td>
<td>99.2 (99.0–99.3)</td>
</tr>
<tr>
<td></td>
<td>≥75 y</td>
<td>3461</td>
<td>87.8 (81.6–93.9)</td>
<td>66.4 (58.7–74.1)</td>
<td>99.6 (99.4–99.8)</td>
<td>98.6 (98.2–99.0)</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>7804</td>
<td>92.0 (87.1–96.9)</td>
<td>62.6 (55.4–69.8)</td>
<td>99.9 (99.8–100)</td>
<td>99.2 (98.9–99.4)</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>7285</td>
<td>85.8 (87.1–96.9)</td>
<td>55.2 (48.3–62.2)</td>
<td>99.7 (99.6–99.9)</td>
<td>98.8 (98.5–99.0)</td>
</tr>
<tr>
<td></td>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>5574</td>
<td>93.0 (88.3–97.6)</td>
<td>66.1 (58.8–73.3)</td>
<td>99.8 (99.7–100)</td>
<td>99.0 (98.7–99.2)</td>
</tr>
<tr>
<td></td>
<td>White</td>
<td>9515</td>
<td>86.5 (80.6–92.5)</td>
<td>53.3 (46.5–60.1)</td>
<td>99.8 (99.7–99.9)</td>
<td>99.0 (98.8–99.2)</td>
</tr>
<tr>
<td>ICH</td>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;65 y</td>
<td>1915</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>65–74 y</td>
<td>9713</td>
<td>85.4 (71.3–99.5)</td>
<td>67.3 (50.7–84.0)</td>
<td>100.0 (99.9–100)</td>
<td>99.9 (99.8–100)</td>
</tr>
<tr>
<td></td>
<td>≥75 y</td>
<td>3461</td>
<td>93.8 (81.0–100)</td>
<td>69.4 (49.9–88.8)</td>
<td>100.0 (99.9–100)</td>
<td>99.8 (99.7–100)</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>7804</td>
<td>86.7 (71.0–100)</td>
<td>77.8 (59.6–96.0)</td>
<td>100.0 (99.9–100)</td>
<td>99.9 (99.9–100)</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>7285</td>
<td>91.3 (79.8–100)</td>
<td>61.6 (45.3–78.0)</td>
<td>100.0 (99.9–100)</td>
<td>99.8 (99.7–99.9)</td>
</tr>
<tr>
<td></td>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>5574</td>
<td>81.4 (64.4–98.5)</td>
<td>82.5 (65.7–99.2)</td>
<td>99.9 (99.9–100)</td>
<td>99.9 (99.9–100)</td>
</tr>
<tr>
<td></td>
<td>White</td>
<td>9515</td>
<td>95.2 (86.1–100)</td>
<td>59.1 (42.5–75.6)</td>
<td>100.0 (100–100)</td>
<td>99.9 (99.8–99.9)</td>
</tr>
<tr>
<td>AIS/ICH</td>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;65 y</td>
<td>1915</td>
<td>86.1 (71.9–100)</td>
<td>45.5 (30.7–60.3)</td>
<td>99.8 (99.6–100)</td>
<td>98.7 (98.2–99.2)</td>
</tr>
<tr>
<td></td>
<td>65–74 y</td>
<td>9713</td>
<td>92.2 (87.7–96.8)</td>
<td>57.9 (51.3–64.5)</td>
<td>99.9 (99.8–100)</td>
<td>99.1 (98.9–99.2)</td>
</tr>
<tr>
<td></td>
<td>≥75 y</td>
<td>3461</td>
<td>90.1 (84.9–95.4)</td>
<td>67.9 (60.7–75.0)</td>
<td>99.6 (99.4–99.8)</td>
<td>98.4 (98.0–98.8)</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>7804</td>
<td>93.1 (88.8–97.4)</td>
<td>64.2 (57.5–71.0)</td>
<td>99.9 (99.8–100)</td>
<td>99.1 (98.9–99.3)</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>7285</td>
<td>88.3 (83.1–93.5)</td>
<td>57.3 (50.9–63.7)</td>
<td>99.8 (99.6–99.9)</td>
<td>98.6 (98.4–98.9)</td>
</tr>
<tr>
<td></td>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>5574</td>
<td>93.1 (88.9–97.4)</td>
<td>68.8 (62.1–75.5)</td>
<td>99.8 (99.7–99.9)</td>
<td>99.0 (98.7–99.2)</td>
</tr>
<tr>
<td></td>
<td>White</td>
<td>9515</td>
<td>89.3 (84.3–94.3)</td>
<td>54.6 (48.3–60.9)</td>
<td>99.8 (99.7–99.9)</td>
<td>98.8 (98.6–99.1)</td>
</tr>
</tbody>
</table>

AIS indicates acute ischemic stroke; CI, confidence interval; ICH, intracranial hemorrhage; NPV, negative predictive value; and PPV, positive predictive value.

*Calculations could not be completed as some cells had zero observations.
of 88% among Tennessee Medicaid enrollees from 1999 to 2003 and with the PPVs of 99% for AIS, 89% for ICH, and 94% for subarachnoid hemorrhage among Seattle residents from 1990 to 1996.

Compared with some disease-specific populations, our base population is more similar to the general US population, arising from a population-based community-dwelling sample that constitutes the study cohort of REGARDS. While the oversampling from the stroke belt resulting in high stroke prevalence in our cohort will influence the generalizability of PPV to those populations with lower or higher stroke prevalence, the sensitivity and specificity estimates should be directly applicable to other populations with varying prevalence. Also, our gold standard events included ≈10% of cases not resulting in hospitalizations, diagnosed only in outpatient settings. Inclusion of these cases led to lower sensitivity compared with some previously reported numbers, but again our estimate would be more relevant to most etiologic studies conducted using claims databases, where participants are community dwellers at risk of nonadmitted strokes.

We conducted sensitivity analyses to estimate the impact of assumed stroke frequency among the suspected events with unretrievable medical records, and observed limited impact on the validity measures, especially specificity. The original medical record retrieval in REGARDS was triggered by the participant’s self-report of a potential event. For this validation study, we retrieved additional charts for those with suspected strokes by inpatient diagnoses in Medicare data. The retrieval was less complete for events identified by Medicare claims only (43% missing) compared with those with self-reports (10% missing). One of the major reasons for nonretrieval was the hospital request to obtain an updated medical record release form from the participants, which was not always successful. This was particularly a problem for the additional chart retrieval for events identified by Medicare claims only, as some of these events had happened more than a year before

| Table 6. Accuracy Measures Stratified by Age, Sex, and Race in First-Stroke Cohort |
|--------------------------------|------|---------|-----------------|-----------------|-----------------|
| Algorithm | Category | n      | PPV (95% CI), % | Sensitivity (95% CI), % | Specificity (95% CI), % | NPV (95% CI), % |
| AIS | Age | | | | | |
| <65 y | 1551 | * | * | * | * |
| 65–74 y | 8702 | 96.9 (93.1–100) | 56.3 (47.9–64.6) | 100.0 (99.9–100) | 99.3 (99.1–99.5) |
| ≥75 y | 2843 | 89.4 (82.3–96.6) | 66.8 (57.3–76.3) | 99.7 (99.5–99.9) | 98.9 (98.5–99.3) |
| Sex | | | | | | |
| Female | 6782 | 95.7 (91.3–100) | 63.4 (54.9–71.9) | 99.9 (99.9–100) | 99.3 (99.1–99.5) |
| Male | 6314 | 87.2 (79.8–94.6) | 54.1 (45.4–62.8) | 99.8 (97.9–99.9) | 99.1 (98.8–99.3) |
| Race | | | | | | |
| Black | 4689 | 93.2 (87.4–99.0) | 62.7 (53.5–71.9) | 99.9 (98.8–100) | 99.1 (89.9–99.4) |
| White | 8407 | 90.6 (84.5–96.8) | 55.8 (47.6–64.0) | 99.9 (98.8–100) | 99.2 (99.1–99.4) |
| ICH | Age | | | | | |
| <65 y | 1551 | * | * | * | * |
| 65–74 y | 8702 | 76.7 (55.3–98.1) | 56.6 (35.1–78.2) | 100.0 (99.9–100) | 99.9 (98.8–100) |
| ≥75 y | 2843 | 92.3 (77.8–100) | 64.6 (42.9–86.4) | 100.0 (99.9–100) | 99.8 (98.6–99.9) |
| Sex | | | | | | |
| Female | 6782 | 79.2 (57.4–100) | 74.2 (50.3–98.2) | 100.0 (99.9–100) | 100.0 (99.9–100) |
| Male | 6314 | 88.2 (72.9–100) | 53.3 (34.8–71.7) | 100.0 (99.9–100) | 99.8 (99.7–99.9) |
| Race | | | | | | |
| Black | 4689 | 72.2 (46.9–97.6) | 71.5 (46.0–96.9) | 99.9 (99.9–100) | 99.9 (99.9–100) |
| White | 8407 | 94.1 (82.9–100) | 55.6 (37.6–73.8) | 100.0 (100–100) | 99.8 (98.6–99.9) |
| AIS/ICH | Age | | | | | |
| <65 y | 1551 | * | * | * | * |
| 65–74 y | 8702 | 96.2 (92.3–100) | 56.4 (48.6–64.2) | 100.0 (99.9–100) | 99.2 (99.0–99.4) |
| ≥75 y | 2843 | 92.4 (88.7–96.1) | 67.4 (58.8–76.0) | 99.8 (98.6–99.9) | 98.7 (98.2–99.1) |
| Sex | | | | | | |
| Female | 6782 | 96.3 (92.4–100) | 64.6 (56.5–72.6) | 99.9 (99.9–100) | 99.3 (99.1–99.5) |
| Male | 6314 | 89.9 (83.8–96.0) | 55.2 (47.3–63.1) | 99.8 (97.9–99.9) | 98.9 (98.6–99.2) |
| Race | | | | | | |
| Black | 4689 | 92.8 (87.2–98.4) | 64.3 (55.6–72.9) | 99.9 (98.8–100) | 99.1 (98.8–99.4) |
| White | 8407 | 93.3 (88.5–98.1) | 56.5 (49.0–63.9) | 99.9 (99.9–100) | 99.1 (98.8–99.3) |

AIS indicates acute ischemic stroke; CI, confidence interval; ICH, intracranial hemorrhage; NPV, negative predictive value; and PPV, positive predictive value. *Calculations could not be completed as some cells had zero observations.
from when the additional medical record retrieval was pursued for the current study. This informative missingness was accounted for, in part, by the stratified imputation of the missing gold standard stroke status.

To maximize the number of additionally captured strokes and obtain stable estimates for the validity measures in the presence of resource restrictions, we pursued additional medical records only for the events identified by the restrictive algorithm of AIS/ICH. Because of this, we have no reliable estimate of the true stroke rate among the events identified by more inclusive algorithms, such as those including secondary discharge diagnoses, outpatient diagnoses, or ICD-9-CM codes for transient ischemic attack (435), other and ill-defined cerebrovascular disease (437), or late effects of cerebrovascular disease (438) when not accompanied by self-report. We are thus unable to report the validity of these more expansive algorithms, which are expected to have greater specificity in exchange for specificity. In the Data Supplement, we present probable ranges of validity estimates for 2 such algorithms based on our data. These suggest that inclusion of secondary position discharge diagnoses for the same codes as the AIS/ICH algorithm will most likely result in a several-percentage-point higher sensitivity in exchange for a 0.1-to-0.3-percentage-point lower specificity, and can be an alternative for identifying strokes as outcomes for etiologic studies (Table II and Figure I in the Data Supplement). On the other hand, the algorithm using all 430 to 438 codes in primary and secondary discharge diagnoses could result in as high as an 80% sensitivity in exchange for substantially reduced specificity and PPV (Table III and Figure II in the Data Supplement). These algorithms should be directly validated in future studies.

Our study has several limitations. First, our primary analysis assumed that the failure of medical chart retrievals occurred dependent only on the event identification process. Other variables, observed or unobserved, associated with retrieval failure and the risk of stroke could have biased our estimates. However, the impact of the bias would be limited considering the results of sensitivity analyses. Second, not all participants ≥65 years in the REGARDS cohort were linked to Medicare claims data. The proportion of black participants was 13 percentage points higher in the nonlinked population, suggesting the possibility of systematic disapproval for linkage and incompleteness of linkage variables in this subpopulation (Table I in the Data Supplement). Third, for construction of the first-stroke cohort, we excluded patients with prior history or events suggestive of strokes, using the data from the REGARDS study only. It is possible that the identification of prior stroke is incomplete, but likely small. Lastly, because of the use of single imputed means to deal with the missingness, the widths of the reported confidence intervals are inevitably but not substantially underestimated.

Conclusion

In the REGARDS-Medicare claims linked data set of participants sampled from the general US population, claims-based algorithms using primary discharge diagnoses captured true stroke events among Medicare enrollees with a high PPV and a high specificity. This finding supports the validity of the relative risk estimates derived in etiologic or comparative effectiveness studies with stroke outcome under the assumption of nondifferential misclassification, as well as of stroke cohort identification. Because of their low sensitivity, however, the usefulness of these algorithms to accurately estimate population-level incidence rates of stroke or of related healthcare utilizations or costs is limited. Further studies are needed to evaluate more sensitive Medicare algorithms for these purposes.

Acknowledgments

We thank the investigators, the staff, and the participants of the REGARDS study for their valuable contributions. A full list of participating REGARDS investigators and institutions can be found at http://www.regardsstudy.org.

Sources of Funding

This project was sponsored by the Agency for Healthcare Research and Quality (AHRQ), US Department of Health and Human Services, Rockville, MD, as part of the Cardiovascular Consortium and funded under contract number HHSAA290201000007–Task Order 2, as part of the Research Consortia for Comparative Effectiveness Studies in Cardiovascular Disease. The REGARDS project is supported by
cooperative agreement U01NS041588 from the National Institute of Neurological Disorders and Stroke, National Institutes of Health, Department of Health and Human Services. Dr Kumamaru was supported by the Program in Pharmacoepidemiology at Harvard School of Public Health through training grants from Pfizer, Millennium, Pharma, and ASISA and Honjo International Scholarship Foundation. Dr Setoguchi was supported by midcareer development award K02HS017731 from AHRQ. Dr Curtis was supported by grant R01HS018517 from AHRQ and grant K23AR053351 from National Institute of Arthritis and Musculoskeletal and Skin Diseases.

Disclosures

Dr Setoguchi reported receiving research support from Johnson & Johnson. The other authors report no conflicts.

References
