Assessment of the Completeness and Accuracy of Case Ascertainment in the Michigan Stroke Registry

Mathew J. Reeves, PhD; Adrienne V. Nickles, MPH; Stacey Roberts, MS; Rochelle Hurst, RN; Sarah Lyon-Callo, MS

Background—Accurate case ascertainment is essential for clinical registries to be valid and representative. We assessed case ascertainment in the Michigan Stroke Registry by linking to a statewide hospital discharge database (Michigan Inpatient Database [MIDB]).

Methods and Results—In 2009, all ischemic stroke cases submitted by 30 registry hospitals were linked to ischemic stroke discharges (International Classification of Diseases, Ninth Revision code 433.x1, 434.x1, or 436) in the MIDB. Databases were linked using hospital, age, sex, and admission date. The MIDB was regarded as the gold standard. To assess completeness, we calculated the percent difference between the number of cases entered in the registry relative to the MIDB. To quantify accuracy, we defined sensitivity as the proportion of cases identified in the MIDB that were matched to the registry and positive predictive value as the proportion of cases identified in the registry that were matched to the MIDB. Before data linkage, 4 hospitals were known to be using a case sampling approach. The remaining 26 registry hospitals submitted 21% fewer cases (n=3403) than were found in the MIDB (n=4340). The overall sensitivity was 68.8% (95% confidence interval, 76.4%–79.3%), and positive predictive value was 87.7% (95% confidence interval, 87.4%–89.8%). The sensitivity of case ascertainment was significantly lower in teaching hospitals and primary stroke centers but was higher in the sites that used prospective case ascertainment methods.

Conclusions—Among registry hospitals, these results revealed relatively high levels of completeness and accuracy. Matching registry data to hospital discharge data identified hospitals that changed their case ascertainment method to a case sampling approach. This study illustrates the value of monitoring case ascertainment in stroke registries using external data sources. (Circ Cardiovasc Qual Outcomes. 2014;7:757-763.)

Key Words: registries • stroke • validity (epidemiology)

Accurate case ascertainment in clinical registries is essential to obtain valid and representative information. Although complete and accurate case ascertainment is recognized as essential, only a minority of cardiovascular disease–based registries report conducting case audits. The Paul Coverdell Acute Stroke Registry is a multistate hospital-based registry and quality improvement program that is overseen by the Centers for Disease Control and Prevention. Data on the completeness of case ascertainment in hospitals participating in the Paul Coverdell Registry are lacking. Barriers to completing case ascertainment audits include cost, time, and lack of a suitable external gold standard to identify eligible cases. Audits may also be complicated by the fact that some hospitals do not submit all their cases to the registry, relying instead on case sampling algorithms, such as those developed by the Joint Commission.

The aim of this study was to assess the completeness and accuracy of ascertainment of ischemic stroke cases in the Michigan Stroke Registry (MSR), a participant in the Centers for Disease Control and Prevention Paul Coverdell Acute Stroke Registry, by comparing MSR data with the Michigan Inpatient Database (MIDB), an administrative database of all hospital discharges in the state.

Methods

MSR: Design

The MSR collects information on the quality of acute stroke care as part of the Centers for Disease Control and Prevention Paul Coverdell National Acute Stroke Registry. Hospitals were selected through a stratified complex sampling scheme (based on geography, urban versus rural setting, and minority status [black versus white]) to obtain a representative statewide sample of hospitals providing acute stroke care. Human subjects’ approval was obtained from the Michigan Department of Community Health institutional review board. Each hospital was responsible for complying with its own institutional review board process before starting data collection. Because of the quality improvement focus of the project, the requirement for individual patient consent was waived. During 2007, 20 hospitals were selected as a representative statewide sample of hospitals providing acute stroke care.

Received October 28, 2013; accepted May 21, 2014.
From the Department of Epidemiology and Biostatistics, Michigan State University, East Lansing (M.J.R.); and Michigan Department of Community Health, Lansing (A.V.N., S.R., R.H., S.L.-C.).

The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention (CDC).

Correspondence to Mathew J. Reeves, PhD, Department of Epidemiology and Biostatistics, Michigan State University, 900 Fee Rd, B601, East Lansing, MI 48824. E-mail reevesm@msu.edu

© 2014 American Heart Association, Inc.

Circ Cardiovasc Qual Outcomes is available at http://circoutcomes.ahajournals.org

DOI: 10.1161/CIRCOUTCOMES.113.000706
WHAT IS KNOWN

• A nagging question about hospital-based registries is whether hospitals submit data on all eligible patients and whether incomplete case ascertainment leads to an important level of selection bias.
• Despite the importance of accurate case ascertainment, only a minority of cardiovascular disease–based registries report conducting case audits, in part because audits are time consuming and expensive to conduct and there is often no suitable gold standard to identify true eligible cases.

WHAT THE STUDY ADDS

• This study used an imperfect gold standard (a statewide inpatient database) to undertake a case ascertainment audit of hospitals participating in the Michigan Stroke Registry (a participant in the Paul Coverdell National Acute Stroke Registry).
• This study found that 26 registry hospitals reported ≈20% fewer stroke cases than were found in the inpatient database, but when 3 hospitals that had switched over to case sampling were excluded, this deficit decreased to only 12%.
• These data provide a strong rationale for conducting audits even when using an imperfect gold standard.

recruited as part of the first stage of the registry, and by the beginning of 2009, the registry had been expanded to 36 hospitals. The current analysis uses data from the 2009 calendar year.

MSR: Case Identification

Participating hospitals were instructed to identify all acute stroke admissions using prospective case ascertainment (ie, active surveillance) or retrospective case ascertainment (ie, passive surveillance using International Classification of Diseases, Ninth Revision [ICD-9] discharge codes) or a combination of both approaches. As part of an annual inventory survey, registry hospitals identified which of these 3 case ascertainment mechanisms they used in 2009. Four registry hospitals with large annual stroke caseloads (>300 cases) obtained permission from the registry to submit only a sample of their 2009 acute stroke cases. Three of these hospitals used a sampling approach designed by the registry that ensured sufficient precision of their estimated quality indicators, whereas the other hospital used a Joint Commission sampling scheme.

Six of the original 36 registry hospitals were excluded from this analysis: 5 were excluded because they did not consistently submit data during the 2009 calendar year and 1 was excluded because they did not correctly implement the registry’s case definition for acute stroke which resulted in invalid data (this hospital received additional training to improve coding practices after the issue was discovered). For the remaining 30 hospitals, all acute ischemic stroke cases entered into the registry between January 1 and December 31, 2009 were included in the analysis. To be eligible for the registry, these cases had to have a final clinical diagnosis of ischemic stroke and a principal ICD-9 discharge code of ischemic stroke defined as 433.x1, 434.x1, or 436. Patients aged <18 years, in-hospital stroke events, and cases admitted for elective stroke-related procedures (ie, any elective revascularization procedure including endarterectomy) were not included in the registry.

All registry hospitals entered data into the Get With The Guidelines—Stroke Patient Management Tool. The Patient Management Tool collects data on arrival and admission, initial symptoms, demographics, medical history, in-hospital procedures, treatments, and complications, as well as discharge instructions and treatments. Data were entered by staff involved directly in patient care or by chart abstractors—all of whom were required to participate in the training sessions provided by the data vendor on the use of the Patient Management Tool. To monitor data accuracy, regular site visits were conducted by registry personnel and data from a sample of cases were reabstracted and reviewed at each site. A reliability study undertaken in 2009 by an independent auditor reabstracted 153 data elements from 366 randomly selected charts. The overall agreement between hospital and auditor values was 87.6%, and 60% (n=90) of the elements had agreements of ≥90%. The final clinical stroke diagnosis had a 93% agreement.

Michigan Inpatient Database

The MIDB is a statewide database of all discharges from all acute care, non-Veterans Administration hospitals in the state. The 2009 database included 152 hospitals. All adult ischemic stroke discharges during the 2009 calendar year were identified using principal ICD-9 discharge codes: 433.x1, 434.x1, and 436. To identify only acute stroke admissions, we first excluded all admissions described as elective and then excluded any remaining cases with a length of stay of <2 days who had a principal procedure code that indicated a carotid revascularization procedure was performed (ie, 00.61, 00.63, 38.02, 38.12, 38.22, 38.3, 38.41, 38.42, 88.41).

Data Linkage and Analysis

The completeness of case ascertainment by registry hospitals was determined by comparing the number of ischemic stroke cases entered into the registry with the number of ischemic stroke cases found in the MIDB, thus the MIDB was regarded as the gold standard. To quantify completeness, we calculated the percent difference between the number of cases entered in the registry relative to the number of cases in the MIDB:

\[
\text{Percent difference (\%)} = \left(\frac{\text{registry} - \text{MIDB}}{\text{MIDB}}\right) \times 100.
\]

To determine accuracy of case ascertainment in the registry, case data in the registry and MIDB databases were then linked using the following 4 variables: hospital identification (ID) number, age (in years), sex, and admission date. Subjects were deemed matched only if they were concordant on all 4 variables (ie, deterministic matching). Data linkage was performed using the PROC SQL procedure in SAS Version 9.2. To quantify the accuracy of case ascertainment conducted by each registry hospital, we defined the following 2 measures using the linked data:

- Sensitivity = the proportion of ischemic cases in the MIDB that were matched to the respective registry hospital.
- Specificity = the proportion of ischemic cases in the registry that were matched to the MIDB.

<table>
<thead>
<tr>
<th>MIDB</th>
<th>Matched</th>
<th>Not matched</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matched</td>
<td>2986</td>
<td>417</td>
<td>3403</td>
</tr>
<tr>
<td>Not matched</td>
<td>1354</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4340</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Percent difference (PO) = (3403-4340)/4340 = -21.6%.
Sensitivity = 2986/4340 = 69.8%.
Positive Predictive Value (PPV) = 2986/3403 = 87.7%.

Figure. Summary results from linkage of the Michigan Stroke Registry (MSR) to the Michigan Inpatient Database (MIDB) among 26 registry hospitals not known to be using a case sampling approach: number of acute ischemic stroke cases.
Positive predictive value (PPV) = the proportion of ischemic cases identified by each registry hospital that were matched to the MIDB. Using the binomial approximation, we generated 95% confidence intervals (CIs) for all measures. Results are presented according to the hospital’s reported case ascertainment mechanism, that is, prospective, retrospective, or combined approach. We also determined whether the accuracy of case ascertainment (ie, sensitivity and PPV) varied significantly across different hospital characteristics, that is, teaching status, primary stroke center status, and hospital size (number of licensed beds).

Finally, to assess the potential for selection bias attributable to incomplete linkage that resulted from the deterministic matching algorithm, we compared patient characteristics between those cases that were successfully matched and those that were not within both the registry and MIDB databases.

### Results

Among the 30 registry hospitals included in this analysis, 57% (n=17) were primary stroke centers, 60% (n=18) were teaching hospitals, and 47% (n=14) were both primary stroke centers and teaching hospitals. Thirteen (43%) hospitals were classified as large (>300 beds), 6 (20%) medium (100–300 beds), and 11 (37%) small (<100 beds). Initially, 4 hospitals were known to be using a sampling approach and so were removed from the primary analysis. The remaining 26 hospitals submitted a total of 3403 acute ischemic stroke cases to the registry, which was 937 fewer than the 4340 total acute ischemic stroke discharges found in the MIDB (Figure); the overall percent difference was negative 21.6% ([3403–4340]/4340×100%). Following data linkage, a total of 2986 cases were matched between the 2 databases; thus, the overall sensitivity of the registry was 2986 of 4340 or 68.8% (95% CI, 67.4%–70.2%), and the PPV was 2986 of 3403 or 87.7% (95% CI, 86.6%–88.8%).

### Table 1. Completeness and Accuracy of Case Ascertainment in MSR Hospitals Not Known to be Using a Sampling Approach Who Used Either Prospective or Retrospective Case Ascertainment

<table>
<thead>
<tr>
<th>Hospital ID</th>
<th>Prospective Case Ascertainment (n=6)</th>
<th>Retrospective Case Ascertainment (n=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSR cases, n</td>
<td>a 137 b 587 c 110 d 204 e 163 f 9 g 1210</td>
<td>g 31 h 30 i 65 j 13 k 52 l 33 m 267 n 27 o 8 p 153 q 2 r 681</td>
</tr>
<tr>
<td>MIDB cases, n</td>
<td>a 150 b 649 c 129 d 251 e 225 f 34 g 1438</td>
<td>g 28 h 28 i 62 j 13 k 53 l 39 m 320 n 44 o 14 p 384 q 14 r 999</td>
</tr>
<tr>
<td>Absolute difference, n</td>
<td>a −13 b −62 c −19 d −47 e −62 f −25 g −228</td>
<td>g 3 h 2 i 3 j 0 k −1 l −6 m −53 n −17 o −6 p −231 q −12 r −318</td>
</tr>
<tr>
<td>Percent difference*</td>
<td>a −9 b −10 c −19 d −26 e −28 f −74 g −16</td>
<td>g 11 h 7 i 5 j 0 k −2 l −15 m −39 n −43 o −60 p −86 q −32</td>
</tr>
<tr>
<td>Matched cases, n</td>
<td>a 127 b 540 c 90 d 177 e 140 f 5 g 1079</td>
<td>g 26 h 17 i 58 j 11 k 51 l 31 m 245 n 25 o 6 p 147 q 2 r 619</td>
</tr>
<tr>
<td>Sensitivity, %†</td>
<td>a 64.7 b 83.2 c 69.8 d 70.5 e 62.2 f 14.7 g 75.0</td>
<td>g 92.9 h 60.7 i 93.5 j 84.6 k 96.2 l 79.5 m 76.6 n 56.8 o 42.9 p 38.3 q 14.3 r 62.0</td>
</tr>
<tr>
<td>PPV, %‡</td>
<td>a 92.7 b 92.0 c 81.8 d 86.8 e 85.9 f 55.6 g 89.2</td>
<td>g 83.9 h 56.7 i 89.2 j 84.6 k 98.1 l 93.9 m 91.8 n 92.6 o 75.0 p 96.1 q 100 r 90.9</td>
</tr>
</tbody>
</table>

Linkage results after matching to the MIDB. MIDB indicates Michigan Inpatient Database; MSR, Michigan Stroke Registry; and PPV, positive predictive value.

*Percent difference=((registry−MIDB)/MIDB)×100.
†Sensitivity=matched cases/MIDB cases×100.
‡PPV=matched cases/MSR cases×100.

### Table 2. Completeness and Accuracy of Case Ascertainment in MSR Hospitals Not Known to be Using a Sampling Approach Who Used a Combination Approach

<table>
<thead>
<tr>
<th>Hospital ID</th>
<th>Combination Approach (n=9)</th>
<th>Overall (n=26)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSR cases, n</td>
<td>r 8 s 371 t 262 u 151 v 59 w 36 x 103 y 401 z 121 Total 1512</td>
<td>Overall Grand Total 3403</td>
</tr>
<tr>
<td>MIDB cases, n</td>
<td>r 8 s 380 t 265 u 162 v 61 w 39 x 138 y 576 z 274 Total 1903</td>
<td>Overall Grand Total 4340</td>
</tr>
<tr>
<td>Absolute difference, n</td>
<td>r 0 s −9 t −3 u −11 v −2 w −3 x −35 y −175 z −153 Total −391</td>
<td>Overall Grand Total 937</td>
</tr>
<tr>
<td>Percent difference†</td>
<td>r 0 s −2 t −1 u −7 v −3 w −8 x 25 y −30 z −56 Total −21</td>
<td>Overall Grand Total 937</td>
</tr>
<tr>
<td>Matched cases, n</td>
<td>r 8 s 347 t 231 u 121 v 49 w 31 x 80 y 311 z 110 Total 1288</td>
<td>Overall Grand Total 2986</td>
</tr>
<tr>
<td>Sensitivity, %‡</td>
<td>r 100 s 91.3 t 87.2 u 74.7 v 80.3 w 79.5 x 58.0 y 54.0 z 40.2 Total 67.7</td>
<td>Overall Grand Total 68.8</td>
</tr>
<tr>
<td>PPV, %§</td>
<td>r 100 s 93.5 t 88.2 u 80.1 v 83.1 w 86.1 x 77.7 y 77.3 z 90.9 Total 85.1</td>
<td>Overall Grand Total 87.7</td>
</tr>
</tbody>
</table>

Linkage results after matching to the MIDB. MIDB indicates Michigan Inpatient Database; MSR, Michigan Stroke Registry; and PPV, positive predictive value.

*Overall, total combines data from all 3 case ascertainment methods (ie, Tables 1 and 2).
†Percent difference=((registry−MIDB)/MIDB)×100.
‡Sensitivity=matched cases/MIDB cases×100.
§PPV=matched cases/MSR cases×100.
be noted that these large percent discrepancies were associated with relatively small differences in absolute numbers (12 for hospital q, 35 for hospital x). Nine hospitals (b, d, e, f, m, p, x, y, z) entered ≥25 fewer cases than were identified in the MIDB, and 3 hospitals (p, y, z) had substantial discrepancies in case numbers of −231, −175, and −153, respectively. Hospital-specific sensitivity rates also varied widely from 14% (hospital q) to 100% (hospital r), as did hospital-specific PPVs from 55.6% (hospital f) to 100% (hospitals q and r). It should be noted, however, that these extreme values were associated with the smaller hospitals where modest absolute differences translated into large relative differences. We excluded 4 hospitals that had <25 stroke cases in the MIDB (ie, hospital j, o, q, and r) but found that the results in Tables 1 and 2 were essentially unchanged.

Table 3 summarizes the percent differences, sensitivities, and PPVs by different hospital characteristics (ie, teaching status, primary stroke centers, bed size, case ascertainment mechanism) for the 26 hospitals. The sensitivity of case ascertainment was significantly lower at teaching hospitals and primary stroke centers but was significantly higher at sites using prospective case ascertainment. There were no significant differences in PPVs, except by ascertainment mechanism.

Data on the percent differences, sensitivities, and PPVs for the 4 hospitals that were known to be using a case sampling approach before the evaluation was conducted are shown in Table 4. All 4 hospitals had a large number (between 279 and 420) of ischemic stroke discharges recorded in the MIDB. The overall sensitivity of 39.7% indicates that less than half of the cases discharged by these 4 hospitals were entered into the registry.

After further investigation of the 3 hospitals (p, y, z) with substantially lower registry case numbers compared with the number of cases in the MIDB (ie, 231, 175, and 153 fewer cases, respectively), it was determined that these sites had, unbeknownst to registry staff, begun using a case sampling approach during the 2009 calendar year. If these 3 hospitals were dropped from the overall analysis, then the overall percent difference was reduced to only −12.2% ([2728–3106]/3106×100), whereas the overall sensitivity increased to 77.8% (95% CI, 76.4%–79.3%) and the overall PPV to 88.6% (95% CI, 87.4%–89.8%).

In the MIDB, there were statistically significant differences between matched and unmatched cases in age and length of stay (Table 5). The proportion of unmatched cases was slightly lower in patients aged >80 years, and the length of stay was slightly longer in unmatched cases compared with matched cases. There were also some differences by ICD-9 code and discharge disposition, but these were only marginally significant (P=0.06). In the registry data, the only significant difference observed between matched and unmatched cases was for sex; the proportion of unmatched cases in the MSR was slightly higher in women compared with men (13.6% versus 10.9%), but this difference was not seen in the MIDB. Overall, these findings suggest that the matching process did not lead to any important selection bias.

**Discussion**

Our findings indicate that hospitals fully participating in the MSR entered ≥20% fewer acute ischemic stroke cases relative to the number found in the hospital discharge database. However, the matching process revealed wide hospital-level variability in both the completeness and accuracy of case ascertainment. The matching process also identified 3 hospitals

---

### Table 3. Accuracy of Case Ascertainment by Hospital Characteristics in 26 Hospitals Not Known to be Using a Sampling Approach

<table>
<thead>
<tr>
<th>Teaching Status</th>
<th>Primary Stroke Center</th>
<th>Bed Size</th>
<th>Ascertainment Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>No</td>
<td>&lt;100</td>
<td>100–300</td>
</tr>
<tr>
<td>Hospitals, n</td>
<td>14</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>MSR cases (n)</td>
<td>3017</td>
<td>386</td>
<td>2649</td>
</tr>
<tr>
<td>MIDB cases (n)</td>
<td>3876</td>
<td>464</td>
<td>3641</td>
</tr>
<tr>
<td>Absolute difference (n)</td>
<td>−859</td>
<td>−78</td>
<td>−812</td>
</tr>
<tr>
<td>Percent difference</td>
<td>−22</td>
<td>−17</td>
<td>−20</td>
</tr>
<tr>
<td>Matched cases (n)</td>
<td>2646</td>
<td>340</td>
<td>2325</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>68.3</td>
<td>73.3</td>
<td>67.2</td>
</tr>
<tr>
<td>PPV (%)‡</td>
<td>87.7</td>
<td>88.1</td>
<td>87.7</td>
</tr>
</tbody>
</table>

**Table 4. Completeness and Accuracy of Case Ascertainment in 4 MSR Hospitals Undertaking a Case Sampling Approach**

<table>
<thead>
<tr>
<th>Hospital ID</th>
<th>Case Sampling Hospitals (n=4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>aa bb cc dd Total</td>
</tr>
<tr>
<td>MSR cases, n</td>
<td>164 252 157 106 679</td>
</tr>
<tr>
<td>MIDB cases, n</td>
<td>415 420 362 279 1476</td>
</tr>
<tr>
<td>Absolute difference, n</td>
<td>−251</td>
</tr>
<tr>
<td>Percent difference</td>
<td>−61</td>
</tr>
<tr>
<td>Matched cases, n</td>
<td>156 221 144 65 586</td>
</tr>
<tr>
<td>Sensitivity, %†</td>
<td>37.6</td>
</tr>
<tr>
<td>PPV, %‡</td>
<td>94.5</td>
</tr>
</tbody>
</table>

---

**MIDB indicates Michigan Inpatient Database; MSR, Michigan Stroke Registry; PPV, positive predictive value; and %‡/ †/ ‡PPV matched cases/MSR cases >100.**
that had substantial discrepancies in the absolute number of ischemic stroke cases (>150) which on further investigation revealed that these hospitals had switched to using a sampling approach. When these 3 hospitals were dropped from the analysis, we found that the remaining 23 hospitals participating in the MSR entered ≈12% fewer acute ischemic stroke cases relative to the number in the discharge database. These findings illustrate the value of conducting periodic case ascertainment audits. The assessment of the completeness and accuracy of case ascertainment is an essential step in the evaluation of any clinical registry. It is advised that assessment be undertaken during the early development phases of a registry and repeated periodically thereafter to ensure that accuracy is maintained. This process is especially important in the context of quality improvement registries where the potential for bias in the selection of cases into the registry (whether occurring insidiously or otherwise) remains a distinct possibility.

The relative advantages and disadvantages of using prospective versus retrospective case ascertainment in acute stroke registries has been of central interest since their development. Whether the advantages of prospective case ascertainment, which potentially include more complete and
accurate case ascertainment, are sufficient to justify the added costs is unclear. A previous population-based epidemiological study compared active (ie, prospective) and passive (ie, retrospective) methods of surveillance and concluded that because active case ascertainment was more accurate it was the method of choice providing sufficient resources were available.14 Across the 3 case ascertainment mechanisms (prospective, retrospective, or combined approach) examined in our study, we found that completeness (ie, sensitivity) was higher at the sites that used prospective methods. We also confirmed that the retrospective method, which requires the least amount of additional clinical staffing, was relied on almost exclusively by small hospitals.

This current study builds on our prior work that assessed the accuracy of acute stroke case definitions used during the development phase of Michigan’s acute stroke registry.15 This prior study used an independent physician panel to determine whether cases entered into the registry met the case definition of acute stroke used by the Paul Coverdell Registry. We found that hospital teams were able to accurately identify suspect stroke admissions, successfully determine those that represented acute stroke cases, and assign an appropriate stroke subtype. Overall, we found that registry hospitals had a tendency to under-report rather than over-report acute stroke admissions.15 By using an external data source (ie, the MIDB), the current study extends this work by assessing the completeness of case ascertainment and in determining the concordance between case identification in the 2 databases. Of course, it is well recognized that hospital-based administrative data that use ICD-9 discharge codes to identify acute stroke discharges do not represent a perfect gold standard. Several studies have been undertaken to quantify the accuracy of hospital discharge data for stroke in the United States.16–18 These studies have generally found that the sensitivity and specificity of acute stroke codes are fair to good (ie, ≈75%–90%), and because the sensitivity and specificity estimates (which reflect false-negative and false-positive cases, respectively) tend to be about equal, the overall total number of cases identified also tends to equal out. However, discrepancies in the accuracy of ICD-9 coding likely contributed to the differences in case numbers observed between the 2 data sources in our study, as does the fact that even though the same ICD-9 codes (433.x1, 434.x1, or 436) were used to define ischemic stroke, the registry included the additional requirement that cases had to meet the Coverdell clinical case definition for ischemic stroke which required the treating physician to confirm that acute ischemic stroke was ultimately responsible for the admission.19 Other reasons for incomplete case ascertainment that contributed in the variability observed at the individual hospital level may have included staff turnover, temporary absence of staff (because of sickness, vacations, or change in duties), misunderstanding of the registry’s case definition for acute stroke, and incomplete active surveillance efforts.

There are some limitations to our approach. First, 6 of 36 registry hospitals (16%) were excluded because of incomplete data submission (n=5) or errors in the implementation of the registry’s case definition (n=1). Second, some of the hospitals participating in our registry were rural sites with small caseloads which resulted in imprecise estimates of case completeness and accuracy. Third, the deterministic matching algorithm is not completely accurate and likely contributes to the imperfect sensitivity and PPV estimates. However, such algorithms have been shown previously to produce high-quality linked data,4 and our own analysis indicated that the matching process did not lead to any meaningful level of selection bias in either database. Finally, 4 hospitals were excluded from the main analysis because they were known to be using a case sampling approach before the start of the evaluation, and an additional 3 hospitals were identified as using case sampling after large case discrepancies were investigated further. The goal of our registry is to obtain an accurate representation of all stroke admissions occurring in the participating hospitals. If the sampling fractions used at each of these hospitals were known, then it would be possible to weight the data (using the inverse of the sampling fraction or probability)20 so that it can be combined with the rest of the registry. Unfortunately, the hospitals undertaking case sampling in our study were not all able to define (with accuracy and confidence) their sampling fraction or the size of the initial patient population from which they selected cases. Thus, we were unable to include their data in the main analysis. It is obviously essential that the registry be able to accurately describe each individual hospital’s sampling method, so that it can be determined whether the specific case sample is representative of the hospital’s underlying patient population whose size should be known with certainty. These principles are well explained in sampling protocols, such as those used by the Joint Commission; however, the challenge is that the monitoring of the sampling approach used is left primarily to the hospitals themselves.

In the context of quality improvement registries where there is strong potential for selection bias, these data illustrate the value of monitoring case ascertainment using an external data source, even though the data source is not the ideal gold standard.

Matching registry data to hospital discharge data identified hospitals that had switched to using a case sampling approach, but in the remaining hospitals that continued to enroll all cases, the results revealed high levels of completeness and accuracy.

Sources of Funding
This publication was supported, in part, by the Michigan Department of Community Health Heart Disease and Stroke Prevention Unit, as well as through the grant No. U58 DP000854-01 from the Centers for Disease Control and Prevention.

Disclosures
None.

References


Assessment of the Completeness and Accuracy of Case Ascertainment in the Michigan Stroke Registry
Mathew J. Reeves, Adrienne V. Nickles, Stacey Roberts, Rochelle Hurst and Sarah Lyon-Callo

Circ Cardiovasc Qual Outcomes. 2014;7:757-763; originally published online August 26, 2014; doi: 10.1161/CIRCOUTCOMES.113.000706

Circulation: Cardiovascular Quality and Outcomes is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2014 American Heart Association, Inc. All rights reserved.
Print ISSN: 1941-7705. Online ISSN: 1941-7713

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circoutcomes.ahajournals.org/content/7/5/757

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Cardiovascular Quality and Outcomes 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 Circulation: Cardiovascular Quality and Outcomes is online at: http://circoutcomes.ahajournals.org//subscriptions/