Imaging Data Reveal a Higher Pediatric Stroke Incidence Than Prior US Estimates

Nidhi Agrawal, MD; S. Claiborne Johnston, MD, PhD; Yvonne W. Wu, MD, MPH; Stephen Sidney, MD, MPH; Heather J. Fullerton, MD, MAS

Background and Purpose—Prior annualized estimates of pediatric ischemic stroke incidence have ranged from 0.54 to 1.2 per 100 000 US children but relied purely on diagnostic code searches to identify cases. We sought to obtain a new estimate using both diagnostic code searches and searches of radiology reports and to assess the relative value of these 2 strategies.

Methods—Using the population of 2.3 million children (<20 years old) enrolled in a Northern Californian managed care plan (1993 to 2003), we performed electronic searches of (1) inpatient and outpatient diagnoses for International Classification of Diseases, 9th Revision codes suggestive of stroke and cerebral palsy; and (2) radiology reports for key words suggestive of infarction. Cases were confirmed through chart review. We calculated sensitivities and positive predictive values for the 2 search strategies.

Results—We identified 1307 potential cases from the International Classification of Diseases, 9th Revision code search and 510 from the radiology search. A total of 205 ischemic stroke cases were confirmed, yielding an ischemic stroke incidence of 2.4 per 100 000 person-years. The radiology search had a higher sensitivity (83%) than the International Classification of Diseases, 9th Revision code search (39%), although both had low positive predictive values. For perinatal stroke, the sensitivity of the stroke International Classification of Diseases, 9th Revision codes alone was 12% versus 57% for stroke and cerebral palsy codes combined; the radiology search was again the most sensitive (87%).

Conclusions—Our incidence estimate doubles that of prior US reports, a difference at least partially explained by our use of radiology searches for case identification. Studies relying purely on International Classification of Diseases, 9th Revision code searches may underestimate childhood ischemic stroke rates, particularly for neonates. (Stroke. 2009;40:3415-3421.)

Key Words: child ■ incidence ■ ischemic ■ neonatal ■ stroke

Pediatric stroke has been increasingly recognized as an important cause of childhood disability. Prior annualized incidence estimates in the US range from 0.54 to 1.2 ischemic strokes per 100 000 children.¹⁻³ However, these studies have typically relied on International Classification of Diseases, 9th Revision (ICD-9) codes for the identification of cases. Several reports have documented the limited and variable accuracy of stroke ICD-9 codes when applied to adults,⁶⁻⁷ and a single pediatric study has shown the same for the use of these codes in children.⁸ These data demonstrate the importance of case confirmation through chart review in the generation of stroke incidence estimates. However, because these reports did not address the sensitivity of ICD-9 code searches, they provide no insight into how many stroke cases are missed by epidemiological studies relying solely on this search strategy.

Kaiser Permanente Medical Care Program (KPMCP) is a large comprehensive healthcare plan in Northern California that maintains extensive electronic medical records, including radiology data, on all of its members. Using KPMCP as our setting, we sought to (1) obtain a more accurate estimate of the incidence of pediatric ischemic stroke in the United States by using both ICD-9 code searches of diagnosis databases and text string searches of radiology reports for case identification; and (2) to assess the relative value of these 2 search strategies.

Methods

We addressed these questions using the Kaiser Pediatric Stroke Study, a retrospective cohort study that attempted to identify all stroke cases within the population of children (0 to 19 years of age) enrolled in KPMCP from January 1993 through December 2003; its methods have been partially described in prior reports using Kaiser Pediatric Stroke Study data.⁹,¹⁰ The institutional review boards at KPMCP and the University of California, San Francisco approved this study.

Setting

KPMCP, the largest nonprofit managed care organization in the country, has 16 hospitals and 36 outpatient clinics that provide care to approximately 30% of the population of Northern California. During the 11-year study period, 2.3 million children were members...
and made up our cohort. The demographics of KPMCP are similar to those of California as a whole except for underrepresentation of socioeconomic extremes. KPMCP maintains extensive electronic databases that include demographics, all inpatient and outpatient diagnoses (comprehensive since 1993), and all radiology reports (comprehensive since 1997). Inpatient diagnoses are coded by a trained medical record abstractor, whereas outpatient diagnoses are coded by the medical provider: the physician, for medical visits, or the therapist, for rehabilitation visits. All care received at outside hospitals is captured within the electronic databases. The rare exception could occur if a child has dual forms of insurance and only the alternate insurance pays for the stroke hospitalization. However, because such children would likely return to KPMCP for follow-up, they might still be identified through an outpatient stroke diagnosis or follow-up brain imaging. Discharge summaries, and other records from outside hospitalizations, are placed in the traditional KPMCP medical charts.

Case Ascertainment
The Kaiser Pediatric Stroke Study used 2 search strategies to identify cases of pediatric stroke (ischemic and hemorrhagic) within the study cohort: (1) diagnostic (ICD-9) code searches; and (2) radiology text string searches. For the first, we searched the electronic medical records of all 2.3 million children for hospital discharge and outpatient diagnoses suggestive of a stroke. This included the following ICD-9 codes: 430 (subarachnoid hemorrhage), 431 (intracerebral hemorrhage), 433.x (occlusion/stenosis of precranial arteries), 434.xx (occlusion of cerebral arteries), 435.xx (transient cerebral ischemia), 436 (acute cerebrovascular disease), 437.x (“other and ill-defined cerebrovascular disease” including 437.4, cerebral arteries; 437.5, moyamoya disease; and 437.6, nonpyogenic thrombosis of intracranial venous sinus), and 438.x (late effects of cerebrovascular disease). Because some children with perinatal stroke might only receive a diagnosis of “hemiplegic cerebral palsy (CP),” we crossreferenced with a prior study that used a KPMCP birth cohort (all births from January 1991 through December 2002) and identified pediatric stroke cases through searches of the following CP codes: 342.x (hemiplegia), 343.x (infantile CP), and 344.xx (other paralytic syndromes).12,13

For the radiology search, we performed electronic text string searches of the formal reports of all head imaging, including CT, CT angiogram, MRI, MR angiogram, and conventional angiogram. Text strings included: stroke, infarct# (includes infarction, infarcted), thromb# (includes thrombus, thromboembolic, thromboembolism, thrombotic, thrombosis), ischemia# (includes ischemia, ischemic), lacun# (includes lacune, lacunar), vascular event, and porencephal# (includes porencephaly, porencephalic). We could not electronically distinguish between those reports affirming the presence of infarction and those negating it. Hence, the full imaging reports were then reviewed by a pediatric neurologist (H.J.F.) to classify the stroke type and determine the underlying etiology; strokes without an apparent etiology were defined as idiopathic.

Data Analysis
Stroke incidence rates were calculated for the study period when radiology records were complete (January 1997 to December 2003) as the number of strokes occurring during that study period divided by the total number of person-years at risk. We compared the usefulness of our 2 search strategies for the identification of pediatric ischemic stroke cases by calculating both sensitivities and positive predictive values (PPVs). Sensitivity of a search strategy was defined as the number of confirmed ischemic stroke cases identified through that particular search strategy divided by the total number of confirmed ischemic stroke cases. (The “gold standard” was therefore the combination of all search strategies used in the study.) Sensitivities were calculated using only cases that occurred during the study period when radiology records were complete (1997 to 2003). Because prior studies have typically relied on ICD-9 code searches of inpatient diagnoses only, we performed stratified analyses to determine the sensitivity of such an inpatient search and the added benefit of an outpatient search. We similarly assessed the added benefit of searching CP ICD-9 codes to identify ischemic stroke cases. We performed additional stratified analyses by age group, stroke subtype, and whether or not the stroke was idiopathic.

PPV was defined as the proportion of confirmed cases to potential cases for a particular search strategy. PPVs were calculated for the 2 search strategies. In addition, because the accuracy of a stroke code could differ when applied to an inpatient versus an outpatient, we calculated PPVs of stroke ICD-9 codes in the inpatient and outpatient settings. PPVs were also calculated for individual ICD-9 codes (including hemorrhagic stroke codes) to assess the accuracy of each code.

We used χ² tests (or Fisher exact, when appropriate) to compare proportions. Alpha was set at 0.05. All statistical comparisons were performed using Stata 9.0 (College Station, Texas).

Results
For the Kaiser Pediatric Stroke Study, we searched the records of 2,347,982 children enrolled in KPMCP from 1993 to 2003 (including 1,775,441 enrolled from 1997 to 2003, when electronic radiology records were complete). The diag-
nostic code search identified 441 cases with ischemic- or hemorrhagic stroke-related codes and an additional 866 cases with CP codes (Figure). The radiology search identified 9599 head imaging reports (regarding 8658 individual children) with stroke-related text strings; review of these reports yielded 510 cases with imaging consistent with stroke (infarction, intracerebral hemorrhage, or subarachnoid hemorrhage). Through independent chart review, a total of 205 cases of ischemic stroke was confirmed: 84 perinatal and 97 later childhood arterial ischemic strokes; and 9 perinatal and 15 later childhood venous sinus thromboses. Of these, only 42 were identified through both search strategies, whereas 45 were identified through the ICD-9 code search alone and 118 through the radiology search alone. We identified an additional 12 cases of transient ischemic attack and 153 cases of hemorrhagic stroke (21 in neonates, 93 in older children); 37 (24%) hemorrhagic strokes were incidentally identified through the radiology search and had no stroke ICD-9 code.

**Stroke Incidence (1997 to 2003)**
Of the 205 ischemic stroke cases, 132 occurred between 1997 to 2003, yielding an ischemic stroke incidence of 2.4 per 100,000 person-years (95% CI, 2.0 to 2.9) during that study period. The incidence of later-childhood ischemic stroke (excluding perinatal cases, n=60) was 1.3 per 100,000 person-years (95% CI, 1.0 to 1.7). Addition of hemorrhagic stroke cases (n=114) yielded a combined stroke incidence (all ages) of 4.6 per 100,000 person-years (4.0 to 5.2); inclusion of transient ischemic attack cases (n=7) increased the total stroke incidence to 4.7 per 100,000 person-years (95% CI, 4.1 to 5.3).

From 1997 to 2003, there were 208,876 live births within KPMCP and 60 ischemic strokes within that birth cohort. The prevalence of perinatal ischemic stroke was 29 per 100,000 live births (95% CI, 22 to 37), or one per 3500 live births. The overall prevalence of perinatal stroke (including 17 hemorrhagic stroke cases) was 37 per 100,000 live births (95% CI, 29 to 46), or one per 2700 live births.

**Figure.** Flow chart demonstrating methods for the ascertainment of stroke cases among 2.3 million children enrolled in KPMCP, 1993 to 2003. Because electronic radiology reports were not comprehensive until 1997, our incidence rate estimates and sensitivity analyses excluded cases from 1993 to 1996; totals for those analyses are therefore different than from those in this diagram.
Comparison of Search Strategies: Sensitivities and PPVs

Although the radiology search was sensitive for ischemic stroke cases, the stroke ICD-9 code search identified only one third (Table 1). Searches limited to inpatient stroke ICD-9 codes identified only one fifth of ischemic stroke cases. The addition of CP ICD-9 codes to the search increased the sensitivity to 59%. The radiology search was the most sensitive strategy regardless of age group, ischemic stroke subtype, or etiology (Table 2). The sensitivity of the stroke ICD-9 code search was particularly low for perinatal ischemic strokes: 12% compared with 50% for later childhood ischemic stroke. Conversely, the added benefit of CP ICD-9 code searches was greater in the younger age group, increasing the sensitivity by 45% for perinatal stroke compared with only 9% for later childhood stroke.

We compared later childhood ischemic stroke cases that never received a stroke ICD-9 code (as either an inpatient or outpatient) with those that received a stroke code (Table 3). They were younger, more likely to have had their stroke as an inpatient, and less likely to have presented with a hemiparesis or speech-related deficit. They were more likely to have a major infection (mostly meningitis) or trauma as their stroke etiology, suggesting that their deficits may have been directly attributed to those etiologies rather than to stroke as the intermediary.

**PPV of ICD-9 Codes**

In general, inpatient coding had a higher PPV than outpatient coding, and the hemorrhagic stroke codes had higher PPVs...
than the ischemic stroke codes (Table 4). Code 436 was the most frequently used code but had the lowest PPV.

Discussion

Using 2 search strategies for case identification—diagnostic code searches and radiology text string searches—we found an incidence of pediatric ischemic stroke of 2.4 per 100 000 person-years in a retrospective cohort of Northern Californian children. We found that radiology searches were more sensitive than diagnostic code searches, particularly for perinatal ischemic stroke, although both search strategies had limited accuracy with low PPVs.

Published pediatric stroke incidence rates are difficult to compare because of the multiple variables that could affect these estimates: population, demographics (higher rates in black children), study time period (before or after the MRI era), age range included, and multiple aspects of study design (eg, retrospective versus prospective, search strategies, methods of case confirmation). Our overall ischemic stroke incidence rate of 2.4 per 100 000 person-years is 2- to 4-fold higher than previously published estimates in US children that also included perinatal strokes (Table 5).1,2 Our estimate of 1.3 per 100 000 person-years also exceeds estimates from those US studies that excluded neonates.3-5 None of these prior reports used radiology searches to identify cases, and the majority of them relied only on inpatient stroke diagnoses.1-5,18-20 Our study demonstrates the low sensitivity of such diagnostic code searches, suggesting that our expanded case identification methods, with the addition of radiology searches, may explain the higher ischemic stroke rate we observed. Had we relied only on stroke and CP ICD-9 code searches, we would have obtained an overall ischemic stroke incidence rate of only 0.97 per 100 000 person-years (95% CI, 0.78 to 1.2), a rate comparable to those reported in prior retrospective studies in the United States that used ICD-9 code searches alone (Table 5).1,3-5

The highest published rate of childhood ischemic stroke (including perinatal stroke) was from a population-based study in Dijon, France (study period 1988 to 1989): 7.9 per 100 000 person-years.21 This was the only prospective study (Table 5) and therefore did not rely on any retrospective search strategies for case identification. Rather, it was a registry in which all stroke cases were identified prospectively in both the inpatient and outpatient setting as part of a larger epidemiological stroke study that included adults. Although the higher rate may also reflect a different patient population, it does suggest that incidence rates from retrospective studies are likely underestimates.

In our study, we found that ICD-9 code searches are not only insensitive for pediatric stroke, but also fairly inaccurate; the poor PPV of many stroke codes has been previously reported and was confirmed in our study (Supplemental Table I; available at http://stroke.ahajournals.org).6-8 These data have unfortunate implications on pediatric stroke research, which has depended largely on retrospective observational studies. Administrative data sets, for example, have been useful because they allow the identification of a large number of subjects with this relatively rare disease. However, these studies typically rely on ICD-9 codes alone for case identification and therefore are not only missing false-negative cases, but including false-positives. The latter issue can be overcome in studies in which cases can be confirmed through chart review, but this is typically not an option for studies using administrative data sets. When available, radiology text string searches, although time-consuming with a low yield, appear to be a better option for retrospectively detecting ischemic stroke cases. However, in evaluating and designing studies relying on diagnostic code searches alone, investigators should consider which subjects are most likely to be missed; among the later childhood stroke cases that did not receive an ICD-9 code, almost half had meningitis or sepsis as their stroke etiology and therefore might actually be excluded from certain stroke studies.

Our study has several limitations. Our conclusions regarding the relative advantages of the 2 search strategies are limited in that we studied only a single large managed care program. Because coding practices may differ in different institutions, our findings may not be generalizable to other settings. This may explain differences in the PPV of some

### Table 4. PPV of Diagnostic Codes for Ischemic and Hemorrhagic Stroke When Applied to Children in KPMCP, 1993 to 2003, in the Inpatient and Outpatient Settings

<table>
<thead>
<tr>
<th>ICD-9 Code</th>
<th>Inpatient PPV</th>
<th>Outpatient PPV</th>
<th>Overall PPV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Percent (95% CI)</td>
<td>N</td>
</tr>
<tr>
<td>Ischemic stroke</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>433.x (occlusion of precerebral arteries)</td>
<td>8</td>
<td>38 (9–76)</td>
<td>3</td>
</tr>
<tr>
<td>434.x (occlusion of cerebral arteries)</td>
<td>31</td>
<td>74 (55–88)</td>
<td>5</td>
</tr>
<tr>
<td>435.x (transient cerebral ischemia)</td>
<td>12</td>
<td>67 (35–90)</td>
<td>40</td>
</tr>
<tr>
<td>436 (acute, ill-defined cerebrovascular event)</td>
<td>13</td>
<td>46 (19–75)</td>
<td>164</td>
</tr>
<tr>
<td>437 (other ill-defined cerebrovascular event)</td>
<td>1</td>
<td>100 (3–100)</td>
<td>0</td>
</tr>
<tr>
<td>Hemorrhagic stroke</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>430 (subarachnoid hemorrhage)</td>
<td>39</td>
<td>82 (66–92)</td>
<td>2</td>
</tr>
<tr>
<td>431 (intracerebral hemorrhage)</td>
<td>52</td>
<td>79 (65–89)</td>
<td>71</td>
</tr>
<tr>
<td>All codes</td>
<td>156</td>
<td>73 (65–80)</td>
<td>285</td>
</tr>
</tbody>
</table>

N indicates total no. of potential stroke cases identified by ICD-9 codes search.
specific ICD-9 codes compared with a prior pediatric report; however, overall, PPVs were remarkably consistent between studies (Supplemental Table I). Another limitation is that we likely failed to detect some cases despite our use of 2 search strategies. Of particular concern are KPMCP patients who present acutely to a non-Kaiser hospital; head imaging studies performed outside of the KMPCP system would not have been included in our radiology text string search. However, diagnostic codes for all out-of-plan care are maintained in KMPCP databases and were searched for this study. In addition, these patients usually return to the KMPCP system for follow-up head imaging and follow-up clinical care. However, missed cases would impact both our stroke incidence estimates and estimates of the sensitivity of the different search strategies. For the latter, the “gold standard” was simply the combination of the different retrospective strategies. Of particular concern are KPMCP patients who present acutely to a non-Kaiser hospital; head imaging studies performed outside of the KMPCP system would not have been included in our radiology text string search. However, diagnostic codes for all out-of-plan care are maintained in KMPCP databases and were searched for this study. In addition, these patients usually return to the KMPCP system for follow-up head imaging and follow-up clinical care. However, missed cases would impact both our stroke incidence estimates and estimates of the sensitivity of the different search strategies. For the latter, the “gold standard” was simply the combination of the different retrospective strategies. Had our gold standard been a thorough and comprehensive prospective registry, as done in Dijon, France, the sensitivity estimates would have been even lower. Despite these limitations, our study is the first to use 2 search strategies to estimate the incidence of pediatric ischemic stroke in the United States and found a rate 2 to 4 times higher than prior reports. Although radiology searches appear to have a greater sensitivity for pediatric stroke, both search strategies have relatively low yield. These challenges in retrospective case identification support recent calls for the field to move toward prospective multicenter studies of pediatric stroke.5,22

Sources of Funding
Supported by the American Heart Association Scientific Development Grant, Neurological Sciences Academic Development Award (K12 NS01692), and Independent Scientist Award (K02 NS053883).

Disclosures
None.

References
Imaging Data Reveal a Higher Pediatric Stroke Incidence Than Prior US Estimates
Nidhi Agrawal, S. Claiborne Johnston, Yvonne W. Wu, Stephen Sidney and Heather J. Fullerton

Stroke. 2009;40:3415-3421; originally published online September 17, 2009;
doi: 10.1161/STROKEAHA.109.564633
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2009 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the
World Wide Web at:
http://stroke.ahajournals.org/content/40/11/3415

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/