Use of ICD-9 Coding for Estimating the Occurrence of Cerebrovascular Malformations

Mitchell F. Berman, Christian Stapf, Robert R. Sciacca and William L. Young

AJNR Am J Neuroradiol 2002, 23 (4) 700-705
http://www.ajnr.org/content/23/4/700

This information is current as of October 19, 2023.
Use of ICD-9 Coding for Estimating the Occurrence of Cerebrovascular Malformations

Mitchell F. Berman, Christian Stapf, Robert R. Sciacca, and William L. Young

BACKGROUND AND PURPOSE: Accurate epidemiologic data concerning cerebrovascular malformations are scarce. Our goals were to determine the distribution of lesions in the International Classification of Diseases, Ninth Revision, (ICD-9) code for cerebrovascular malformations and to evaluate the use of state discharge registries for estimating their detection rate.

METHODS: We reviewed records of all patients discharged from our center between January 1, 1992, and June 30, 1999, whose diagnoses included the ICD-9 code for cerebrovascular anomaly (code 747.81) to determine the accuracy of the coding. Hospital admission rates for cerebrovascular anomaly were calculated by using the 1995–1999 state discharge databases of California and New York.

RESULTS: Of 804 patients with this code, 706 (88%) had a lesion consistent with the diagnosis. Five lesions accounted for 99% of the diagnoses; the two most common were AVM (66%) and cavernous malformation (13%). The ratio of AVMs to all cerebrovascular anomalies was similar to that in a prior population-based study. The sensitivity of identifying a patient with cerebrovascular malformation by using ICD-9 coding was 94%; the false-positive rate was 1.7 cases per 100,000 person-years. For California and New York, rates of first hospital admission for cerebrovascular malformation were 1.5 and 1.8 cases per 100,000 person-years, respectively.

CONCLUSION: Rates of admission for cerebrovascular malformations calculated from state discharge databases are consistent with disease detection rates in the range of 1 case per 100,000 person-years. However, the false-positive rate for coding is in the same range as the disease detection rate. Thus, current state discharge registries cannot serve as sources of detailed epidemiologic data.

Cerebrovascular malformations are generally defined as developmental arterial and/or venous disorders of the brain, including arteriovenous malformations (AVMs), cavernous malformations, venous malformations, and telangiectasia (1). Among them, AVMs are most likely to be associated with intracranial hemorrhage and clinically important symptoms (2).

Accepted estimates for the prevalence of cerebrovascular malformations based on findings from pathology series have ranged from 360 to 4700 cases per 100,000 population (1, 3). For AVMs specifically, the prevalence has been estimated between 140 and 521 cases per 100,000 population (1, 4). Although these rates are often repeated in the literature, they are undoubtedly overestimates; the true rates are at least 10–100-fold lower (5). A population-based study (2) of cerebrovascular malformations revealed a prevalence of 19 cases per 100,000.

Given the paucity of reliable epidemiologic data concerning cerebrovascular malformations, we evaluated the use of general medical discharge registries as a source of epidemiologic data. States, regions (6), and national governments (7–10) maintain these discharge databases (also called administrative databases), and they have been used to provide epidemiologic and outcome data for other cerebrovascular diseases, including unruptured aneurysms and stroke (11, 12). We analyzed the distribution of cerebrovascular malformations that were evaluated at our tertiary medical center and estimated the accuracy of the diagnostic coding for these lesions.
Methods

Administrative databases generally use the *International Classification of Diseases, Ninth Revision*, (ICD-9) (13) to code medical diagnoses. In the ICD-9 scheme, all cerebrovascular malformations are coded together as “anomalies of cerebrovascular system” (code 747.81). This code includes AVMs of the brain. Rupture of a cerebrovascular malformation (most commonly an AVM) is coded as either a subarachnoid hemorrhage (code 430) or an intracerebral hemorrhage (code 431).

To estimate the accuracy of using administrative databases to estimate the detection rates of cerebrovascular malformations, we compared the diagnostic coding for patients at our hospital with the records of an AVM registry maintained by the AVM Study Group at our institution. The AVM Study Group is a multidisciplinary group of neurosurgeons, neurologists, anesthesiologists, and neuroradiologists that prospectively enroll consecutive patients admitted to our medical center for the evaluation and treatment of AVMs. In all patients followed up by the AVM Study Group, the diagnosis is confirmed with angiography, and the study group database was considered the criterion standard. For patients who were not part of the AVM Study Group, we examined the patient’s medical record, including the discharge and operative summaries, pathology and radiology procedure reports, to confirm the diagnosis.

Our analysis of the hospital’s administrative database involved creating a list of all patients with visits (both inpatient admissions and outpatient examinations) between January 1, 1992, and June 30, 1999, in whom the ICD-9 code for cerebrovascular anomaly (code 747.81) was recorded anywhere on their discharge list. These findings were then compared with the records of the AVM Study Group and the results of the detailed chart review described earlier to determine how many patients with a code for a cerebrovascular anomaly actually had one.

We also retrospectively evaluated the records of the patients in the Columbia AVM registry by assessing the diagnostic codes (ICD-9) for these patients in the hospital’s own administrative database; this was done to determine the percentage of patients with a known cerebrovascular anomaly for whom the correct ICD-9 code was used. The result of this evaluation is a measure of the sensitivity of an administrative database in identifying a patient with a cerebrovascular anomaly. Ideally, the sensitivity of ICD-9 coding should have been calculated with records in a prospectively collected database that included all types of cerebrovascular anomalies, but such a database does not exist.

Patient diagnoses were evaluated by using diagnostic lists involving all admissions for a particular patient, not on a per-admission basis. As many as 12 codes are recorded for each discharge at our hospital. Any or all of a patient’s admissions might have been unrelated to the diagnosis of cerebrovascular anomaly; the goal was to create a list of all patients who had been given this diagnosis at some point in their medical history. That is, if a patient had the diagnostic code for cerebrovascular anomaly in any admission, the patient was classified as having the diagnosis.

Detection rates are the most reliable and useful descriptor of disease occurrence (2, 5), and an estimate of the detection rate (eg, number of new cases per 100,000 person-years) is the goal in the analysis of regional medical registries. Prevalence rates, on the other hand, are difficult to define because cerebrovascular malformations may be present for years before they become symptomatic.

The accuracy of administrative databases can be formally described in terms of sensitivity and specificity. The sensitivity of ICD-9 coding for identifying a patient with cerebrovascular malformation (true-positive rate) was calculated as the ratio of the number of patients with a correct code for this type of lesion divided by the total number of patients in the studied population who had the lesion (14). Confidence intervals for the sensitivity were calculated by using the tails of the cumulative binomial distribution (15).

The specificity of ICD-9 coding was expressed as a false-positive error rate—the probability that someone with a code for a cerebrovascular malformation was never actually given that clinical diagnosis. The false-positive rate is inversely related to specificity, and it is equal to 1 minus the specificity (14). Use of the false-positive rate allows simple comparison with the underlying disease rate. To facilitate comparison with the detection rates, we expressed the false-positive error rate as a rate per unit time, that is, the number of false diagnoses per 100,000 person-years. Confidence intervals for the false-positive error rate were calculated by using the tails of the Poisson distribution (16).

The clinical definition of cerebrovascular malformation differs slightly from that used in ICD-9 coding, and it is limited to only arteriovenous malformation, venous malformation, cavernous malformation, and telangiectasia (1). Although our primary analysis concerned ICD-9 coding, we used the clinical definition for the comparison of our medical center data with the results of a population-based study of intracranial vascular malformations (2).

Discharge databases provided by the states of California and New York were used to calculate the number of admissions in 1995-1999 for treatment of cerebrovascular malformations. Data for all of 1999 were available for New York at the time of analysis, but California data were available only through June 30, 1999 (6, 17). Entries for 1995–1999 were searched for discharges in which cerebrovascular malformation (code 747.81) was one of the primary diagnoses; that is, it was listed as a principle diagnosis or one of first two other diagnoses for any admission. To approximate a detection rate, second and follow-up admissions for any particular patient during 1995–1999 were ignored. Because first admissions for any study period include those for patients in whom the lesion was detected prior to the period being studied (in addition to those for patients whose disease was detected in the study period), the rate of first admission is expected to be higher than the true detection rate. Confidence intervals were calculated by using the tails of the Poisson distribution (16).

In this article the term AVM, sometimes written as brain AVM for emphasis, denotes an AVM of the brain parenchyma. This is differentiated from dural arteriovenous fistulae (also referred to as dural AVMs in the literature) and other arteriovenous fistulae, such as vein of Galen malformations. These lesions have clinical characteristics, treatments, and prognoses that differ from those of brain AVMs (18–20), and they are generally separated in clinical research and epidemiologic studies (2, 21). We followed this convention to permit the comparison of our data concerning AVMs with the results of other studies. The distinction had no effect on the analysis of the detection rates for cerebrovascular malformations because all cerebrovascular malformations (brain AVMs, dural AVMs, vein of Galen malformations, and all other malformations) are grouped together in the ICD-9 code for anomalies of the cerebrovascular system.

Results

The administrative database of our medical center contained data of 804 patients (Fig 1) who had at least one discharge between January 1, 1992, and June 30, 1999, that included the ICD-9 code for cerebrovascular anomaly (code 747.81). In almost all cases (99%), the diagnosis was listed in one of the top three positions of importance. Of the 804, 706 patients (88%) had disease entities that were appropriate for the code for cerebrovascular anomaly. Of the
98 patients without evidence of a cerebrovascular anomaly, 67 had a lesion related to the nervous system (usually a vascular lesion outside the cerebral circulation or a nonvascular lesion such as a brain tumor), and in 31 patients, the coding was grossly erroneous (ie, they had a disease processes unrelated to the nervous system). These errors generally involved isolated vascular lesions, such as fistulas, malformations, or aneurysms, located outside the central nervous system, for which the source of the erroneous coding was readily apparent.

Table 1 shows the distribution of lesions correctly coded as cerebrovascular anomalies in our hospital database. Most were true brain AVMs, and the five most frequent diagnoses accounted for 99% of the group.

<table>
<thead>
<tr>
<th>Lesion</th>
<th>n</th>
<th>(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain AVM</td>
<td>464</td>
<td>65.7</td>
</tr>
<tr>
<td>Cavernous malformation</td>
<td>94</td>
<td>13.3</td>
</tr>
<tr>
<td>Unruptured cerebral aneurysm</td>
<td>71</td>
<td>10.1</td>
</tr>
<tr>
<td>Dural AVM (arteriovenous fistula)</td>
<td>55</td>
<td>7.8</td>
</tr>
<tr>
<td>Venous malformation</td>
<td>15</td>
<td>2.1</td>
</tr>
<tr>
<td>Other†</td>
<td>7</td>
<td>1.0</td>
</tr>
<tr>
<td>Total</td>
<td>706</td>
<td>(100)</td>
</tr>
</tbody>
</table>

* Distribution of lesions correctly coded with the ICD-9 diagnosis of 747.81, “anomalies of cerebrovascular system.”† Including vein of Galen malformation (n = 3), Sturge-Weber syndrome (n = 1), anomalous venous drainage (n = 1), venous varix (n = 1), and von Hippel-Lindau disease (n = 1).

98 patients without evidence of a cerebrovascular anomaly, 67 had a lesion related to the nervous system (usually a vascular lesion outside the cerebral circulation or a nonvascular lesion such as a brain tumor), and in 31 patients, the coding was grossly erroneous (ie, they had a disease processes unrelated to the nervous system). These errors generally involved isolated vascular lesions, such as fistulas, malformations, or aneurysms, located outside the central nervous system, for which the source of the erroneous coding was readily apparent.

Table 1 shows the distribution of lesions correctly coded as cerebrovascular anomalies in our hospital database. Most were true brain AVMs, and the five most frequent diagnoses accounted for 99% of the group.

To determine the sensitivity of ICD-9 coding for the identification of patients with cerebrovascular malformations, we analyzed the discharge diagnoses of the 458 patients with AVMs in the local AVM registry described in Methods. A total of 429 had the appropriate ICD-9 code (code 747.81), leading to a sensitivity of 94% (95% CI: 91%, 96%) for ICD-9 coding. Of the 29 patients (6%) with known AVMs who did not have the appropriate ICD-9 code, 12 (3%) had the code for cerebral aneurysm (code 437.3), although only four of these actually had an aneurysm in addition to the AVM. Of the remainder, five had a combination of codes for intracranial hemorrhage, subarachnoid hemorrhage, and/or seizures, and the rest had codes unrelated to the diagnosis of cerebral AVM.

As an external control, we calculated the ratio of AVMs to all cerebrovascular malformations for comparison with data from the only existing population-based study of intracranial vascular malformations (2) known to us. The investigators in this study retrospectively analyzed population-based data about the residents of Olmsted County, Minnesota (United States), during a 27-year period ending in 1992. The definition of intracranial vascular malformations used in the study more closely resembled the medical definition than the ICD-9 code grouping (1, 22), and it included brain AVMs, dural AVMs, venous malformations, and cavernous malformations. With this definition of vascular malformation, brain AVMs accounted for 464 (73.9%) of 628 intracranial vascular malformations in our hospital database. In the Olmsted County data, the detection rate of brain AVMs relative to all intracranial malformations is 1.11 cases per 100,000 person-years compared with 1.82 cases per 100,000 person-years, or 61% of the rate for all detected intracranial vascular malformations.

We calculated the false-positive rate for the use of ICD-9 coding to identify patients with cerebrovascular anomalies as the number of patients inappropriately given the code for cerebrovascular anomaly divided by the total number of patients (753,405 unique patients, not patient visits) who came to our medical center during the study period. Thus, the rate was 98 of 753,405 unique patients in 7.5 years, or 1.7 cases per 100,000 person-years (95% CI: 1.4 per 100,000 person-years, 2.1 cases per 100,000 person-years).

Table 2 shows the results for the analysis of the discharge data from New York and California. During 1995–1999, 3534 and 2667 admissions occurred in California and New York, respectively, with a major diagnosis of cerebrovascular malformation (ie, cerebrovascular malformation was listed as one of the three most important diagnoses, as described in Methods). Based on the average population for each
of these two states during the study period, rates of first hospital admission for cerebrovascular malformation during the study period were between 1 and 2 cases per 100,000 person-years.

**Discussion**

We found that the use of administrative databases for studying the epidemiologic features of cerebrovascular malformations is primarily limited by the low but notable false-positive rate of detection. The rate of first hospital admission for cerebrovascular malformation in the California and New York datasets were remarkably similar—1.5 and 1.8 cases per 100,000 person-years, respectively. The results were in the same range as the presumed true detection rate of between 1 and 2 cases per 100,000 person-years (2, 5). However, the false-positive rate of coding for a cerebrovascular malformation, 1.7 cases per 100,000 person-years, was also in the same range. Thus, the error rate and the coded admission rate are too close to derive an independently validated estimate of the detection rate.

Still, the New York and California datasets add important information to the question of the actual prevalence of cerebrovascular malformations. As discussed in the introduction, pathology series have historically been used to determine the prevalence rates, but they are biased toward overestimates (5, 20). McCormick, the pathologist who elaborated the current classification system for cerebrovascular malformations, published one of the most carefully documented and frequently cited series reports (1) during the 20-year period ending in 1984. In his series, the prevalence of cerebrovascular malformations is 4700 cases per 100,000 population, and that of arteriovenous malformations is 521 per 100,000.

However, the prevalence of detected disease must mathematically equal the disease duration multiplied by the detection rate (23). Even if one allows for an average disease duration of 50 years, McCormick’s estimates require incidence rates of 94 cases per 100,000 patient-years for cerebrovascular malformations and 10 cases per 100,000 patient-years for arteriovenous malformations. In McCormick’s series, 12% of the vascular malformations overall and more than two thirds of the arteriovenous malformations involved considerable intracranial hemorrhage (5, 24). If the results of the pathology series were indicative of occurrence rates in the general population, hospital admission rates for cerebrovascular malformations would have been much higher than the 1–2 per 100,000 patient-years that we found in the California and New York administrative datasets.

Results of the analysis of the New York and California data were consistent with the lower rates described in reports of two existing population-based studies involving cerebrovascular malformations. Brown’s analysis from Olmsted County, Minnesota (data from the Mayo Clinic, 1965–1992), revealed a detection rate of 1.8 cases per 100,000 person-years for cerebrovascular malformations (2). Nogueira’s data from Qatar is consistent with a detection rate for AVMs of 0.9 case per 100,000 person-years (25, 26). These two studies involved a small numbers of lesions—48 malformations in 27 years in Minnesota and 14 AVMs in 6 years in Qatar. The large number of patients identified in the state datasets and the overall size of the samples are important in confirming the results of the smaller population-based studies.

Our analysis demonstrates some of the methods and limitations involved in using hospital discharge databases for epidemiologic studies. In theory, they are ideal for studying infrequent but potentially life-threatening lesions such as cerebrovascular malformations. However, errors in medical coding and deficiencies in the ICD-9 coding system, primarily diagnostic groups that are not specific enough, are recognized problems (27, 28).

To use a discharge database to approximate a detection rate, one must have estimates of the sensitivity of coding (the probability that the diagnosis of interest receives the correct ICD-9 code) and of the relative magnitude of the false-positive rate (spurious occurrences of the ICD-9 code being studied) compared with the disease rate. Individual hospitals compile the data for state discharge databases and electronically send the data to state health agencies; thus, the coding errors occur at the level of the treating hospital.

In our analysis, we assumed that the accuracy rates for coding that we measured at our medical center were comparable to those at other medical centers and, thus, applicable to the administrative databases maintained by our own state and other states. As at other medical centers, coding at our hospital is performed by general medical coders who abstract the entire medical record. Although our hospital is a tertiary care center, treatment of cerebrovascular disease represents a tiny fraction of treatments in the patient population. No special input is received from members of the institution’s AVM Study Group. The sensitivity of using ICD-9 coding in the identification of patients with malformations was high—94% received the correct code. This result is not surprising because the description in the ICD-9 manual is quite specific, and it contains no similar code that might be mistakenly entered for these lesions (13).

---

**TABLE 2: Rates of first hospital admissions for cerebrovascular malformation determined by using state discharge data for 1995–1999**

<table>
<thead>
<tr>
<th>State</th>
<th>Hospital Admissions</th>
<th>Unique Patients</th>
<th>State Population (millions)</th>
<th>Rate of First Hospital Admission (cases per 100,000 person-years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>California</td>
<td>3534</td>
<td>2157</td>
<td>32.2</td>
<td>1.5 (95% CI: 1.4, 1.6)</td>
</tr>
<tr>
<td>New York</td>
<td>2667</td>
<td>1664</td>
<td>18.2</td>
<td>1.8 (95% CI: 1.7, 1.9)</td>
</tr>
</tbody>
</table>

Small errors in estimating the false-positive coding rate can obscure the measured detection rate of rare diseases. The false-positive rate of coding for cerebrovascular malformations derived from our local hospital data was only slightly less than the overall admission rate for the disease in our own state (New York), and it was slightly higher than the overall hospital admission rate in California. Still, unless one hypothesizes that the great majority of cerebrovascular malformations in 1995–1999 were clinically misdiagnosed or administratively miscoded, the discharge datasets can be used to confirm the upper limit of a range for the detection rate.

Other aspects of the analysis deserve comment. To arrive at a conservative estimate for a detection rate for cerebrovascular malformations, we limited our analysis of state data to admissions for which cerebrovascular malformation was one of the three main diagnoses. The goal was to eliminate the inclusion of follow-up admissions, for which the vascular lesion was not the main concern and which probably did not represent first detection of the lesion. To calculate error rates based on our own hospital data, we sought to estimate the maximum possible error and took the opposite approach. A diagnosis of cerebrovascular malformation included anywhere in the list of 12 diagnoses was sufficient for us to consider the patient as having a code for a cerebrovascular malformation. In fact, in 99% of the patients in our hospital, the code for the lesion appeared in one of the top three positions of importance; therefore, the difference in approach is only theoretically important.

Finally, in an administrative database, unique patient identifiers (generally, encrypted combinations of the patient’s social security number, birth date, and other demographic data) are used to identify and eliminate readmissions of the same patient. Any period selected for study includes the first admissions for newly detected lesions and the follow-up admissions for patients who received a diagnosis in an earlier period. Thus, the rate of first hospital admission causes overestimation of the detection rate of a disease. As the length of the study period increases, the percentage of patients who first received a diagnosis in an earlier time period decreases, and the rate of first admission in the time period being studied approaches the true detection rate.

Conclusion

The two most common types of cerebrovascular malformation in the general patient population at our medical center, a tertiary referral center for the evaluation of cerebrovascular malformations, were AVMs (66%) and cavernous malformations (13%). The percentage of AVM lesions was similar to that of the single population-based study (2) from Olmsted, Minnesota, conducted some 20 years ago. ICD-9 coding may have adequate sensitivity for the identification of patients with cerebrovascular malformations in regional discharge databases, but the rate of false-positive errors (coding for a malformation when it was never diagnosed), though low, is in the range of the disease detection rate itself. Data from New York and California state discharge databases are consistent with a detection rate of fewer than 2 cases per 100,000 person-years, but state discharge databases cannot provide more detailed epidemiologic data concerning cerebrovascular malformations.

References

17. Public Patient Discharge Data. Sacramento, Calif: California Office of Statewide Health Planning and Development; 2001
28. Chewning SJ, Nussman DS, Griffo ML, Kiebzak GM. Health care information processing: how accurate are the data? J South Orthop Assoc 1997;6:8–16