Journal of Rehabilitation Research and Development Vol. 38 No. 2, March/April 2001
Pages 281–289

CLINICAL REPORT

Stroke: Who's counting what?

Dean M. Reker, PhD, RN; Byron B. Hamilton, MD, PhD; Pamela W. Duncan, PhD, PT1; Shu-Chuan Jennifer Yeh, PhD, RN; Amy Rosen, PhD
Department of Veterans Affairs Medical Center, Kansas City, MO; Center on Aging, University of Kansas Medical Center, Kansas City, KS; Center on Aging, Duke University Medical Center, Durham, NC; Institute of Human Resources Management, National Sun Yat-Sen University, Kaohsiung, Taiwan; Center for Health Quality, Outcomes, and Economic Research, Bedford VAMC, Bedford, MA

Abstract—Introduction: Patients with stroke are often selected for epidemiological reporting and research using ICD-9-CM (ICD-9) diagnostic codes. This study addresses the accuracy of these codes in identifying patients with stroke. Methods: A sample of 279 patients with new stroke and 392 non-stroke (no evidence of new stroke) patients were identified by medical record review from 11 Veterans Affairs Medical Centers. Administrative records containing ICD-9-CM diagnoses were matched with this sample. Coding sensitivity and specificity were determined using individual ICD-9 codes and two coding algorithms. Results: Significant variation was found in the accuracy of cerebrovascular ICD-9-CM codes in identifying patients diagnosed with stroke. Two coding algorithms were identified with the following performance statistics based on the sampled populations: 1) 91-percent sensitivity, 40-percent specificity; and 2) 54-percent sensitivity, 87-percent specificity. Discussion/Conclusions: Variability in identifying patients with stroke using ICD-9 codes has been reported in the literature and confirmed. Two coding algorithms for maximizing sensitivity or specificity are proposed. Caution is urged when using ICD-9-coded administrative data to identify patients with stroke.

Key words: *diagnosis, epidemiology studies, stroke, stroke classification.*

This material is based on work supported by the Health Services Research and Development Service, Rehabilitation Research and Development Service and QUERI Stroke, STR-98-000-1, Office of Research and Development, Veterans Health Administration, Department of Veterans Affairs.

Address all correspondence and requests for reprints to: Dean M. Reker, PhD, Veterans Administration Medical Center, Research (151), 4801 Linwood Boulevard, Kansas City, MO 64128; email: dean.reker@med.va.gov or dreker@kumc.edu.

INTRODUCTION

What is a stroke? From a clinical perspective, the diagnosis of stroke is fairly unambiguous, particularly with the assistance of brain imaging. Following clinical diagnosis, however, the classification of patients with stroke begins to become less clear. The ambiguity starts with the assignment of diagnostic coding in the form of International Classification of Disease, 9th Revision, Clinical Modification (ICD-9-CM, or ICD-9) codes.

A review of the general ICD-9 codes for cerebrovascular disease delineates seemingly clear boundaries for diagnostic subgroups:

430.xx Subarachnoid hemorrhage

431.xx Intracerebral hemorrhage

432.xx Other and unspecified intracranial hemorrhage

433.xx Occlusion and stenosis of precerebral arteries

434.xx Occlusion of cerebral arteries

435.xx Transient cerebral ischemia

436.xx Acute, but ill-defined, cerebrovascular disease

437.xx Other and ill-defined cerebrovascular disease

438.xx Late effects of cerebrovascular disease

Each of the above general codes have many subclassifications (using the fourth and fifth digits) that are too numerous to replicate in this paper, but further add to the boundary delineations (1). Patients are tagged with these diagnoses in outpatient clinics, on hospital discharge summaries, on fiscal billing forms, and on death certificates. Are these codes assigned accurately? Can patients with stroke be identified using these codes? The purpose of this study was to address these questions and to validate the use of ICD-9 coding algorithms in the identification of patients with stroke.

The Centers for Disease Control and Prevention and the National Center for Health Statistics group patients with stroke as "Cerebrovascular Disease" and use all cerebrovascular codes from 430–438, inclusive (2,3). When the National Center for Health Statistics reports mortality, patients, and number of discharges for stroke, they use all cerebrovascular codes 430–438, inclusive (4). Similarly, the American Heart Association defines stroke as ICD-9 codes 430 through 438, inclusive (5). Many, if not all, stroke authors cite these national sources when estimating the relative impact of stroke on the United States population and the health care system. As a result, the impact of stroke may be overestimated (6). For additional discussion of this topic, see Williams et al. (7).

When researchers select patients with stroke for a study sample, they will often selectively choose subsets of ICD-9 codes from the cerebrovascular array of ICD-9 codes 430–438. In a non-random, convenience sample of 10 stroke studies, 7 different ICD-9 coding strategies were used to identify patients with stroke (6,8–16).

There are two critical issues to consider when using differing ICD-9 coding criteria to identify patients with stroke: volume and accuracy. In order to understand how stroke volume varies with differing ICD-9 definitions for stroke, **Table 1** displays patient volume for individual ICD-9 codes and differing combinations of aggregate codes for a 1 y (1998) period of hospital admissions in the Veterans Health Administration.

As evidenced in **Table 1**, the "count" of patients with stroke is highly dependent upon which codes are selected. As an example, a stroke sample may start with a discharge diagnosis code between 430.xx and 438.xx. In the above data, there are 17,729 discharges meeting this

Table 1.
"Stroke" discharges over 12 months using differing ICD-9 definitions

ICD-9 codes		Diagnostic field	Number of discharges (% of all)
430-438	All cerebrovascular codes	Any (includes secondary)	46,011 (7.6%)
430-438	All cerebrovascular codes	Admitting or discharge	18,080 (3.0%)
430-438	All cerebrovascular codes	Admitting	17,735 (3.0%)
430-438	All cerebrovascular codes	Discharge	17,729 (3.0%)
430	Subarachnoid hemorrhage	Discharge	135 (0.02%)
431	Intracerebral hemorrhage	Discharge	664 (0.11%)
432	Other intracranial hemorrhage	Discharge	392 (0.06%)
433.x0	Occlusion and stenosis of precebral arteries without mention of cerebral infarction	Discharge	4,198 (0.7%)
433.x1	Occlusion and stenosis of precebral arteries with cerebral infarction	Discharge	630 (0.1%)
434.x0	Occlusion of cerebral arteries without mention of cerebral infarction	Discharge	200 (0.03%)
434.x1	Occlusion of cerebral arteries with cerebral infarction	Discharge	3,699 (0.6%)
435	Transient cerebral ischemia	Discharge	2,941 (0.5%)
436	Acute, but ill-defined, cerebrovas disease	Discharge	3,302 (0.5%)
437	Other and ill-defined cerebrovas disease	Discharge	468 (0.08%)
438	Late effects of cerebrosvascular disease	Discharge	1,100 (0.2%)

criterion. Many researchers omit codes 433.x0 (Occlusion and stenosis of precerebral arteries) and 435.x (Transient Ischemic Attack, TIA) because these codes are not considered stroke, but rather possible precursors of stroke. Hence, by omitting these two groups, 7,139 discharges are removed from the listing. Many other researchers will omit codes 437–438 for their low yield of patients with new stroke and thus remove another 1,568 discharges. Following these reductions, there are 9,022 discharges remaining, or 51 percent of the original, defined sample.

The accuracy of ICD-9 coding to identify patients with stroke has been investigated and reported in 1994 (17), 1997 (18), and 1998 (19). In 1994, Liebson et al. (17) studied the Rochester stroke registry patients (n=364) over three sample years (1970, 1980, and 1989) and calculated the sensitivity and the predictive value positive for individual ICD-9 codes in the first five diagnosis fields, as coded by seven hospitals. The number of observations and the percent of new (incident) strokes in each code from the 1980 and 1989 data of Liebson are displayed in **Table 2**. Liebson concluded "the use of all hospital discharge abstracts with a principal diagnosis code of 430–438.9 results in a significant overestimate of stroke incidence" (17).

In a study of the accuracy of ICD-9 coding of ischemic stroke in five academic medical centers, Benesch et al. (18) found similar inaccuracies. In a stratified sampling scheme, 649 patients with codes 433 through 436 in any primary or secondary diagnostic field were included in the analysis. Results of this study are displayed in **Table 2**.

In another study of ischemic stroke coding accuracy, Goldstein (19) investigated the use of modifier codes (fifth digit) in the primary diagnostic field in 175 patients from one VA medical center. The sample of 175 patient discharges was selected by the codes 433 (Occlusion and stenosis of precerebral arteries), 434 (Occlusion of cerebral arteries), and 436 (Acute, but ill-defined, cerebrovascular disease) in the primary diagnosis field. These results are also tabled (**Table 2**) for cross study comparison.

In combination, the above three studies consistently indicate that code 433 (Occlusion and stenosis of precerebral arteries) rarely identifies patients with a new stroke. Additionally, two of the three studies identified a very low

Table 2.
ICD-9 accuracy studies

Code		Liebson, et al. N	Liebson, et al. Percent of new strokes	Benesch, et al. N	Benesch, et al. Percent of new strokes
430.x	Subarachnoid hemorrhage	10	100%		
431.x	Intracerebral hemorrhage	23	74%		
432.x	Other intracranial hemorrhage	7	0%		
433.xx	Occlusion and stenosis or precerebral arteries	20	15%	295	6%
434.x	Occlusion of cerebral arteries	102	69%	226	85%
435.x	Transient cerebral arteries	78	12%	99	9%
436.x	Acute, but ill-defined, cerebrovascular disease	64	67%	29	83%
437.x	Other and ill-defined cerebrovascular disease	9	11%		

Code		Goldstein N	Goldstein Percent of new strokes
433.x0	Occlusion and stenosis of precerebal arteries without mention of cerebral infarction	43	2%
433.x1	Occlusion and stenosis of precerebral arteries with cerebral infarction	5	20%
434.x0	Occlusion and stenosis of cerebral arteries without mention of cerebral infarction	2	100%
434.x1	Occlusion of cerebral arteries with cerebral infarction	106	82%
436.x	Acute, but ill-defined, cerebrovascular disease	19	79%

yield of patients with stroke for code 435.xx (Transient cerebral ischemia). Finally, the more comprehensive study found codes 432 (other intracranial hemorrhage), 437 (Other and ill defined cerebrovascular disease), and 438 (Late effects of cerebrovascular disease) offer little advantage for identifying patients with stroke.

A potential source of ICD-9 misclassification of stroke could be a lack of clinical understanding of the disease and/or the coding criteria by ICD-9 coders. However, one study of patients with stroke examined the agreement rates of ICD-9 code assignment by neurologists (20). Contrary to what one would expect, the agreement rate for the neurologists, as measured by the Kappa statistic, was 0.38, indicating only "fair" agreement.

The present study is an investigation of the accuracy of ICD-9 codes for admission (reason for admission) and discharge in identifying patients with stroke from a large sample of "potential" patients with stroke from 11 Veterans Affairs Medical Centers distributed around the country. This study will provide sample sensitivity and specificity estimates for individual ICD-9 codes and propose two distinct coding algorithms to identify patients with stroke.

METHODS

Sample

The patient sample for this study was selected from two sources. The first source was a prospective study of post-stroke rehabilitation care in 11 Veterans Affairs Medical Centers (VAMC), from which potential study enrollees were identified and screened by clinician research assistants over a 14-mo enrollment period (February 1998 through March 1999). All clinical research assistants participated in a 2.5-d training program covering screening, study enrollment criteria, and patient evaluation. Patients identified and screened for study enrollment were all those that potentially had a recent stroke event (defined as "rapid onset of symptoms which has lasted more than 24 hr and is presumed to be of vascular origin causing focal disturbance of cerebral function and excluding isolated impairment of higher cognitive function" (21)). The portion of the study sample included all patients identified in the above study over the period of July 1, 1998 through January 11, 1999 (n=387).

The second source for the study sample was 321 subjects identified retrospectively using administrative data (Patient Treatment File: PTF main) and ICD-9 codes. Subjects identified using this approach were selected from the same 11 VAMC sites and over the same period of enrollment as those in the first source group. The ICD-9 algorithm used to identify potential patients with stroke in this step was developed in prior, unpublished research to maximize the identification of patients with new stroke (sensitivity) while attempting to limit the selection of non-stroke (no evidence of new stroke) patients (specificity). This high sensitivity algorithm selected patients as follows:

- If admission or discharge primary diagnosis is 430.xx, 431.xx, 432.xx, 434.xx, or 436.xx, or
- Admission or discharge primary diagnosis is V57.xx (Rehabilitation) and any secondary diagnosis is 342.xx (Hemiparesis), 430.xx, 431.xx, 432.xx, 433.xx, 434.xx, 435.xx, 436.xx, 437.xx or 438.xx, or
- Admission or discharge primary diagnosis is 433.xx or 435.xx and any secondary diagnosis code is 342.xx, 430.xx, 431.xx, 432.xx, 434.xx, or 436.xx.

In addition to this high-sensitivity ICD-9 algorithm, a second ICD-9 algorithm maximizing specificity (high-specificity algorithm) has been created using the results of this present study. This high-specificity algorithm is described below in the Results section.

The total sample for the current ICD-9 study was 708 subjects. The gold standard for validation of a diagnosis of new stroke for each of the patients in the two source groups was as follows:

1. For each patient screened for or enrolled into the prospective stroke study, the medical record was reviewed to validate the diagnosis of stroke (defined as described above) and the patient record must have been documented with a diagnosis of stroke. All patients who were screened for the prospective study that failed the recent onset stroke criteria were classified as non-stroke patients; patients who were screened and enrolled in the prospective stroke study were classified as patients with stroke.
2. Patients who were identified retrospectively using the administrative data source had their medical records reviewed at each of the study sites by the same clinician research assistants or the physician site investigators to determine the presence or absence of a documented stroke diagnosis.

For a group of 37 patients (5 percent of total sample, 36 of which were screened), the matching medical or administrative records could not be located. These subjects were excluded from the analyses. The missing data were spread over 8 of the 11 participating sites with 1 site accounting for 16 cases. The primary cause of these unmatched records is suspected to be incorrectly recorded social security numbers rather than any systematic cause. Hence, selection bias is not suspected.

Additionally, some patients with stroke may not have met the study criteria for stroke and subsequently had been assigned as non-stroke. As a source of bias, the authors believe the frequency of this occurrence would be quite low and would have little or no impact on the study results.

RESULTS

Within the validation database of 671 patient discharges, 279 patients (42 percent) had a confirmed diagnosis of new stroke and 392 (58 percent) patient discharges were non-strokes (no evidence of new stroke). **Tables 3 and 4** list the major cerebrovascular codes (430–438) used for the admission and discharge diagnosis and whether or not the patient had a new stroke.

These descriptive tables, in summary, reveal coding differences in identifying newly diagnosed patients with stroke. Individual codes had minimal differences in stroke yield for admission diagnosis (**Table 1**) *versus* discharge diagnosis (**Table 2**). However, more importantly, 87 percent of patients with stroke (240 of the 279) were identified by the cerebrovascular ICD-9 codes in the admission diagnosis field compared to 75 percent of patients with stroke (210 of 279) identified by cerebrovascular codes in the discharge diagnosis field.

Among all the cerebrovascular codes, three shared the highest proportion of patients identified: 431.xx (Intracerebral hemorrhage), 433.x1 (Occlusion and stenosis of precerebral arteries with cerebral infarction), and 434.x1 (Occlusion of cerebral arteries with cerebral infarction) and, at the same time, identified a large proportion of all patients with stroke. Two of the three codes (433.x1 and 434.x1) had higher "hit" rates in percentage of patients with stroke identified when comparing

Table 3.
Number and percent of admission* diagnosis by ICD-9-CM code and presence of stroke

ICD-9 codes		No stroke	New strokes	Predictive Value Positive (PVP) (percent stroke)
430.x	Subarachniod hemorrhage	9	5	36%
431.x	Intracerebral hemorrhage	6	21	78%
432.x	Other intracranial hemorrhage	25	5	17%
433.x0	Occlusion and stenosis of precerebral arteries without mention of cerebral infarction	14	1	7%
433.x1	Occlusion of stenosis of precerebral arteries with cerebral infarction	2	7	78%
434.x0	Occlusion of cerebral arteries without mention of cerebral infarction	4	3	43%
434.x1	Occlusion of cerebral arteries with cerebral infarction	42	123	75%
435.x	Transient cerebral arteries	68	2	3%
436.x	Acute but ill-defined, cerebrovascular disease	62	69	53%
437.x	Other and ill-defined cerebrovascular disease	3	3	50%
438.x	Late effects of cerebrovascular disease	2	1	33%
All other		155	39	20%
Total		**392**	**279**	**42%**

•Admission diagnoses are related to the reason of the admission.

Table 4.
Number and percent of discharge diagnosis by ICD-9-CM code and presence of stroke

ICD-9 codes		No stroke	New strokes	Predictive Value Positive (PVP) (percent stroke)
430.x	Subarachniod hemorrhage	8	4	33%
431.x	Intracerebral hemorrhage	5	20	80%
432.x	Other intracranial hemorrhage	23	6	21%
433.x0	Occlusion and stenosis of precerebral arteries without mention of cerebral infarction	14	2	13%
433.x1	Occlusion of stenosis of precerebral arteries with cerebral infarction	2	5	71%
434.x0	Occlusion of cerebral arteries without mention of cerebral infarction	4	2	33%
434.x1	Occlusion of cerebral arteries with cerebral infarction	42	108	72%
435.x	Transient cerebral arteries	67	2	3%
436.x	Acute but ill-defined, cerebrovascular disease	61	57	48%
437.x	Other and ill-defined cerebrovascular disease	2	2	
438.x	Late effects of cerebrovascular disease	4	2	33%
All other		160	69	30%
Total		**392**	**279**	**42%**

admission diagnoses to discharge diagnoses, and all three identified more patients with stroke in the admission diagnosis field compared to the discharge diagnosis field. These three codes as a group will be used in sensitivity and specificity analyses as a "narrow" coding algorithm or "high-specificity" algorithm to identify patients with stroke with the fewest number of false positives (maximizing specificity).

A second ICD-9-CM coding algorithm, one that maximizes stroke sensitivity estimates, was used in the following sensitivity and specificity analyses. This coding algorithm was developed in prior work and is described in the Methods section. This algorithm will be referred to as the "broad" or "high-sensitivity" algorithm for contrast to the narrow, high-specificity algorithm defined above.

The sensitivity and specificity for the broad, high sensitivity ICD-9-CM coding algorithm is shown in **Table 5**.

Using the narrow, high-specificity algorithm (only 431.x, 433.x1, and 434.x1) in the admission diagnostic field only, the sensitivity and specificity estimates are shown in **Table 6**.

Since both of the above ICD-9 algorithms were based on admission or discharge diagnostic fields, comparable analyses of all diagnostic fields (admission, discharge, and all secondary) were performed. Based on the

Table 5.

		Stroke +	Stroke –
Broad Algorithm	+	254	236
	–	25	156

Sensitivity	91%
Specificity	40%
Predictive value positive (PVP)	52%
Predictive value negative (PVN)	86%

Table 6.

		Stroke +	Stroke –
Narrow Algorithm	+	150	50
	–	129	342

Sensitivity	54%
Specificity	87%
Predictive value positive (PVP)	75%
Predictive value negative (PVN)	73%

literature and the findings of this study, the codes 430.x, 431.x, 432.x, 434.xx, and 436.x were selected and searched for in any diagnostic field as the high-sensitivity model and 431.x, 433.x1, and 434.xx were selected and searched for as the high-specificity model. The performance estimates of these analyses were as follows: high-sensitivity algorithm: 89-percent sensitivity, 57-percent specificity, 60-percent predictive value positive, and 88-percent predictive value negative; high-specificity algorithm: 59-percent sensitivity, 84-percent specificity, 72-percent predictive value positive, and 74-percent predictive value negative.

DISCUSSION

The tradeoff between sensitivity and specificity for the broad and narrow algorithms is quite apparent. By selecting the broad algorithm, sensitivity can be doubled but, at the same time, false positive patients are increased fourfold (from 50 to 236). If a "cleaner" sample of patients with stroke is desired, the narrow algorithm can be used. However, the sample will be approximately one-half of all patients (54 percent, in this sample) and may potentially contain a selection bias if this smaller group is not representative of the larger population. These choices, and their effects, must be considered prior to patient selection.

The algorithms using all diagnostic fields performed almost as well as the algorithms based only on admission diagnosis, but they did not maximize the sensitivity and specificity estimates.

Consistent with prior studies, our findings confirmed that identification of patients with stroke by using ICD-9 codes can be difficult and fraught with unknown effects. Our findings agree with the study of Liebson (17), which identified codes 431.x, 434.xx, and 436.x as high-yielding codes (to identify patients with stroke). Low yielding codes, namely, 432.x, 433.xx, 435.x, were also consistent in both Liebson's study and this investigation.

In comparison with studies of ischemic stroke identification, the results of this study agree with Benesch (18) and Goldstein (19) in identifying ICD-9 codes 434.xx and 436.x as being the most accurate stroke markers. Consistent with the finding of Benesch is the low yield of codes 433.xx and 435.xx with the exception of 433.x1. With the use of the fifth digit modifier, this study found code 433.x1 to be one of the top three identifiers (predictive value positive, PVP) of stroke, although there were few total subjects in this category.

Our study, however, differs from the prior literature on the accuracy of ICD-9-CM stroke coding in a substantial way: sampling methodology. Accuracy of any diagnostic classification system has two dimensions: how well it classifies diseased patients and how well it classifies non-diseased patients. These dimensions are commonly called sensitivity and specificity, respectively. In order to measure both dimensions fairly, you must have representative samples of both diseased and non-diseased patients. If your sample is representative, you can determine the four cells necessary for sensitivity and specificity calculations: true positives (A), false positives (B), true negatives (C), and false negatives (D), depicted in **Table 7**.

Table 7.

		Disease +	Disease –
Classification	+	A	B
	–	D	C

Prior studies on the accuracy of ICD-9 codes have used sample databases that underestimated patients with stroke and/or did not represent patients without stroke. For example, study sample selection based on ICD-9 codes 430–438 will always miss "false negative" patients (cell D, **Table 7**) because these patients fall outside of the "classification" range. When this happens, sensitivity measurements are overestimated because the denominator is reduced [sensitivity=A÷(A+D)]. Prior studies on the accuracy of ICD-9 stroke codes have also not measured specificity or sampled non-stroke patients. As a result, data from cell C (**Table 7**) are not available. Therefore, prior studies have not identified patients in cells C and D. Without these cells, three of the four most common accuracy estimates cannot be calculated. Only PVP [A÷(A+B)] can be calculated in the absence of cells C and D. Sensitivity [A÷(A+D)], specificity [C÷(B+C)], and predictive value negative [C÷(C+D)] cannot be accurately determined from these prior studies.

Ideally, in order to fill in cells A, B, C, and D and compute the most accurate estimates of sensitivity and specificity, an investigator would review the entire population of patients. For this study of 11 VA hospital sites, this task would require review of approximately 45,000 medical records. Given this unrealistic task, this study used a sampling strategy that identified both stroke cases and a subsample of non-stroke (no evidence of new stroke) cases, independent of ICD-9 codes.

Contrary to the sampling schemes found in the literature, the current study sample selection came from two sources. The first source used in creating the study sample was derived from the screening and enrollment process of a prospective stroke study. Many of the referrals and patient identification schemes (e.g., patient symptoms, radiology reports) used in this prospective stroke study were completely independent of ICD-9 codes; therefore, false negative (cell D) patients were potentially sampled. Hence, the denominator (A+D) for a sensitivity analysis has been more accurately estimated. Research assistants also identified many "non-stroke" or "stroke-like" diagnoses in their screening for prospective study enrollees. All of these patients were included in the validity study sample, providing a limited subsample of "non-stroke" patients (C), allowing specificity to be approximated.

The second source for the study sample used a highly sensitive ICD-9 algorithm to maximize the number of patients in cells A and B. This sampling addition creates a balanced study database of stroke and non-stroke (no evidence of new stroke) patients, identified by methods both dependent on and independent of ICD-9 codes.

How might the accuracy of diagnosis coding be improved? Two obvious sources of potential improvements to coding accuracy are coding criteria and clinician documentation. By enhancing and expanding the coding criteria for stroke and cerebrovascular disease, particularly regarding the temporal onset of the disease, the accuracy of specific code assignment and modifiers should improve. Similarly, clinicians must carefully document diagnoses in their discharge summaries that accurately represent new and existing diseases and conditions. For example, if a newly hospitalized patient has hemiparesis as sequela to stroke occurring 4 mo prior to admission and no new cerebrovascular disease is present, a secondary diagnosis of "Hemiparesis" (code 342) should be used rather than "Acute, but ill-defined, cerebrovascular disease" (code 436). Similarly, if late effects of stroke are present, the code identifying these conditions should be used (code 438) rather than codes appropriate for newly presented cerebrovascular events or conditions (430–437).

This study was performed within Veterans Health Administration facilities. As such, the generalizability or external validity of these study findings may be considered a limitation of this investigation. Due to the consistency of the findings with non-VA studies, the investigators do not believe this to be a significant limitation but realize that ICD-9 coding practices may vary substantially when differing financial incentives exist.

CONCLUSION

In conclusion, this study identifies significant variation in the accuracy of cerebrovascular ICD-9 codes in identifying patients with stroke. Sources for this variation may be hospital-specific coding practices and/or policies, inter-coder reliability, or lack of precision of the individual codes. Depending upon the ICD-9 codes used to identify patients with stroke, the sensitivity and specificity of coding algorithms can be altered to fit sample needs. Policy analysts and investigators are urged to be cautious when using administrative data to identify patients with stroke.

ACKNOWLEDGMENTS

The authors wish to thank the following research assistants and investigator for their effort and commitment to the research: Deborah Berg, RN; Lori Danzl, PT; Missy Ford, OT; Angelique King, OT; Jennifer McCuen, PT; Nicole Pilatos, RKT; Randal Schleenbaker, MD; Jean Stocks, RN; Karen Vogel, RN; Kelly Voisey, SLP; Patricia Woodward, RN; and Dana Tucker, BS.

REFERENCES

1. International classification of disease. Volume 1. Clinical modification. American Medical Association. 9th Edition. Baltimore, MD: HCIA; 1995. p. 307–10.
2. Monthly vital statistics report. Centers for Disease Control and Prevention, National Center for Health Statistics, U.S. Department of Health and Human Services; 1997 Dec. Report No.: 46:5.
3. Advance data. Vital and health statistics. Centers for Disease Control and Prevention, National Center for Health Statistics, U.S. Department of Health and Human Services. 1998 Aug. Report No.: 301.
4. National Center for Health Statistics. FASTATS: A to Z. http://www.cdc.gov/Nchswww/fastats/stroke.htm; 11 Jan 2001.
5. 2001 Heart and Stroke Statistical Update. American Heart Association. www.americanheart.org/index.html; 11 Jan 2001.
6. McGovern PG, Pankow JS, Burke GL, Shahar E, Sprafka JM, Fosom AR, Blackburn H. Trends in survival of hospitalized stroke patients between 1970 and 1985. The Minnesota Heart Survey. Stroke 1993;24:1640–8.
7. Williams GR, Jiang JG, Matchar DB, Samsa GP. Incidence and occurrence of total (first-ever and recurrent) stroke. Stroke 1999;30:2523–8.
8. Lee AJ, Huber JH, Stason, WB. Factors contributing to practice variation in post-stroke rehabilitation. Health Serv Res 1997;32(2):197–227.
9. Taylor TN, Davis PH, Torner JC, Holmes J, Meyer JW, Jacobson MF. Lifetime cost of stroke in the United States. Stroke 1996;27:1459–66.
10. Michell JB, Ballard DJ, Whisnant JP, Ammering CJ, Samsa GP, Matchar DB. What role do neurologists play in determining the costs and outcomes of stroke patients? Stroke 1996;27:1937–43.
11. Lee AJ, Huber J, Stason WB. Poststroke rehabilitation in older Americans: The Medicare experience. Med Care 1996;34:811–25.
12. Kane RL, Chen Q, Blewett LA, Sangl J. Do rehabilitative nursing homes improve the outcomes of care? J Am Geriatri Soc 1996;44:545–54.
13. Wolinsky FD, Wan GJ, Gurney JG, Bentley DW. The risk of hospitalization for ischemic stroke among older adults. Med Care 1998;36:449–61.
14. Simons LA, McCallum J, Friedlander Y, Simons J. Risk factors for ischemic stroke: Dubbo Study of the elderly. Stroke 1998;29:1341–6.
15. Kane RL, Chen Q, Finch M, Blewett L, Burns R, Moskowitz M. Functional outcomes of posthospital care for stroke and hip fracture patients under Medicare. 1998. J Am Geriatr Soc 1998;46:1525–33.
16. Wolinsky FD, Gurney JG, Wan GJ, Bentley DW. The sequelae of hospitalization for ischemic stroke among older adults. J Am Geriatr Soc 1998;46:577–82.
17. Liebson CL, Naessens JM, Brown RD, Whisnant JP. Accuracy of hospital discharge abstracts for identifying stroke. Stroke 1994;25:2348–55.
18. Benesch C, Witter DM, Wilder AL, Duncan PW, Samsa G, Matchar DB. Inaccuracy of the international classification of diseases (ICD-9-CM) in identifying the diagnosis of ischemic cerebrovascular disease. Neurology 1997;49:660–4.
19. Goldstein LB. Accuracy of ICD-9-CM coding for the identification of patients with acute ischemic stroke. Stroke 1998;29:1602–4.
20. Kessler C, Freybeger HJ, Dittman V, Ringelstein EB. Interrater reliability in the assessment of neurovascular disease. Cerebrovasc Dis 1991;1:43–8.
21. World Health Organization. Proposal for the multinational monitoring of trends and determinants in cardiovascular disease (MONICA Project). Geneva: World Health Organization; 1983. WHO/MNC/82.1 Rev. 1.

Submitted for publication February 16, 2000. Accepted in revised form October 3, 2000.