

Identifying veterans with acute strokes with high-specificity ICD-9 algorithm with VA automated records and Medicare claims data: A more complete picture

L. Douglas Ried, PhD;^{1–3*} Randi Cameon, MS;¹ Huangang Jia, PhD;¹ Kimberly Findley;¹ Melanie Sberna Hinojosa, PhD;¹ Xinping Wang, PhD;¹ Michael J. Tueth, MD^{3–4}

¹Rehabilitation Outcomes Research Center of Excellence, Malcom Randall Department of Veterans Affairs Medical Center (VAMC), Gainesville, FL; ²College of Pharmacy and ³Department of Psychiatry, College of Medicine, University of Florida, Gainesville, FL; ⁴Psychiatry Service, Malcom Randall VAMC, Gainesville, FL

Abstract—Medicare claims data are available to Department of Veterans Affairs (VA) researchers to identify veterans with acute stroke. Our study sought to (1) ascertain whether additional acute stroke cases are identified with Medicare data and (2) assess the use of VA and Medicare inpatient automated data for assigning the stroke date. The study population was veterans living in Veterans Integrated Service Network 8 with an acute stroke diagnosis during fiscal year 2001. High-sensitivity and high-specificity algorithms were applied to VA data sets and matched with Medicare files. We confirmed acute stroke cases and index dates using the VA Computerized Patient Record System (CPRS). VA data identified 582 veterans with acute stroke, but Medicare claims data identified 201 more such veterans. CPRS confirmed 94% of the VA and 77% of the Medicare cases. The median difference between CPRS and automated index dates was 11 days for VA and 4 days for Medicare data. Use of both VA and Medicare data provides a more complete sample of veterans with acute stroke.

Key words: acuity, cerebrovascular disease, dual users, high-sensitivity algorithm, high-specificity algorithm, ICD-9 codes, Medicare, rehabilitation, selection criteria, stroke, veterans.

INTRODUCTION

An estimated 500,000 strokes occur each year in the United States, 11,000 to 15,000 among veterans in the Department of Veterans Affairs (VA) Health System

(VHS) [1–4]. Roughly 70 percent of these patients survive the acute phase and must deal with serious physical and mental health sequelae [5]. Ongoing research is critical for improving acute treatment options and advancing new modalities for optimal poststroke recovery. Consequently, accurately identifying veterans with acute stroke is essential for determining the true incidence of stroke and assuring that research results accurately apply to the intended cohort. Yet accurately identifying veterans with acute stroke for research poses multiple challenges.

Traditionally, VA researchers have primarily used VA data sources to analyze veterans' health status because of their relative ease of access and known reliability and

Abbreviations: CPRS = Computerized Patient Record System, FY = fiscal year, ICD-9 = International Classification of Diseases-9th Edition, MedPAR = Medicare Provider Analysis and Review, PTF = Patient Treatment File, SCRSSN = scrambled Social Security number, SD = standard deviation, TIA = transient ischemic attack, VA = Department of Veterans Affairs, VHS = VA Health System, VIREC = VA Information Resource Center, VISN = Veterans Integrated Service Network.

*Address all correspondence to L. Douglas Ried, PhD; Rehabilitation Outcomes Research Center of Excellence, 151B, Malcom Randall VAMC, 1601 SW Archer Blvd, Gainesville, FL 32608; 352-376-1611, ext 4969; fax: 352-271-4270. Email: doug.ried@va.gov

DOI: 10.1682/JRRD.2007.02.0033

validity. Yet U.S. veterans are often treated in healthcare systems other than the VHS. For example, in a previous study, up to 73 percent of veteran stroke patients aged 65 and older used both VA and Medicare resources for their poststroke care [6–7]. VA databases alone likely do not tell the whole story for these dual users. Researchers studying stroke should be aware that without merging VA and non-VA databases, many veterans with strokes may be missed. Fairly recently, the VA Information Resource Center (VIREC) began acquiring Medicare claims data for VA researchers' use because of this limitation.

In addition to this limitation, accurately identifying veterans with acute stroke requires careful selection of cerebrovascular-related International Classification of Diseases-9th Edition (ICD-9) codes and subcodes. Codes 430 to 438 and their subcodes are typically used for classification of stroke for epidemiological reporting, administrative, and research purposes [8]. Unfortunately, the accuracy of ICD-9 codes assigned to individuals' stroke episodes varies widely for a number of reasons. First, strokes manifest with a wide variety of symptoms and sequelae and each is classified with a specific code [9]. Some of these codes describe the etiology of the stroke (hemorrhagic vs ischemic), others depict the anatomical location (subarachnoid vs intracerebral), while still others reflect the temporal nature of the stroke (acute but ill-defined vs late effects of cerebrovascular disease). Second, these codes are generally assigned for payment or administrative purposes and may not always accurately reflect the full clinical picture. In other cases, codes properly assigned at the time of an acute stroke may erroneously follow patients throughout long-term recovery for chronic stroke effects. Situations such as these, in which diagnosis classification is susceptible to errors inherent in the assignment of ICD-9 codes, pose challenges when correctly identifying the acute nature of the patients' strokes and are critical to the research focus.

High-sensitivity and -specificity algorithms were developed with the use of cerebrovascular-related ICD-9 codes for identifying veterans with acute stroke in automated VA data [9]. The rationale for selecting the high-sensitivity algorithm is to optimize the number of eligible veterans identified; yet with the high-sensitivity algorithm, false positive patients are also greatly increased (91% sensitivity, 40% specificity) [9]. If, on the other hand, one needs a narrow sample with fewer false positives, one would prefer the high-specificity algorithm (87% specificity) [9]. Still, the high-specificity sampling

strategy in isolation will miss some proportion of "false negative" stroke patients (54% sensitivity) [9].

While the ICD-9-based high-specificity algorithm is highly accurate for identifying acute strokes with VA inpatient data, it has not yet been tested on VIREC-based Medicare claims data. The overall goal of our study was to develop a replicable strategy for identifying a cohort of patients diagnosed with an acute stroke by applying these algorithms to both VA and Medicare data. Specifically, our objectives were to (1) ascertain whether additional acute stroke cases are identified with Medicare data and (2) assess using both VA and Medicare inpatient automated data for assigning the stroke event date for research purposes.

METHODS

Population and Setting

The study population consisted of VHS enrollees living in Veterans Integrated Service Network (VISN) 8, with one or more stroke-related medical care encounters during fiscal year (FY) 2001 (October 1, 2000, to September 30, 2001). VISN 8 is one of 21 VISNs and offers a full spectrum of medical, surgical, and mental health services. Healthcare is delivered through an integrated system of 8 medical centers, 10 multispecialty outpatient clinics, and 36 primary care community-based outpatient clinics. The service area includes 60 of 67 Florida counties, 19 rural counties of South Georgia, Puerto Rico, and the U.S. Virgin Islands. Approximately 2 million veterans reside in the VISN 8 service area and nearly 500,000 of them were served in FY2006.

Data Sources

Automated data files were obtained from the VA Austin Automation Center and the VIREC. VA databases were the inpatient Patient Treatment Files (PTFs), including the PTF Main and the Extended Care Main, and Outpatient files. Medicare data were obtained from the VIREC. We used the Medicare Denominator file to determine Medicare beneficiaries' eligibility and sociodemographic data, the Medicare Provider Analysis and Review (MedPAR) file to obtain Medicare inpatient data, the Medicare Part B Carrier file for noninstitutional care, and the Outpatient file to report the patients' non-hospital-based outpatient care in rehabilitation, urgent care, or assisted-living settings.

We confirmed eligible veterans' stroke dates using the Computerized Patient Record System (CPRS). We used CPRS to investigate the reliability and validity of assigning the automated discharge date as the index date for stroke-related epidemiological studies using these secondary data sources. We confirmed the stroke date by reviewing detailed admission and discharge notes, medical and nursing progress notes, problem lists, ICD-9 codes, and laboratory results including magnetic resonance imaging and computed tomography scans contained in CPRS. We obtained prior permission from the Information Security Officer at each of the VISN 8 facilities before accessing veterans' electronic medical records.

Identification of Acute Stroke and Stroke Date

Our first study objective was to ascertain whether additional acute stroke cases could be identified by the Medicare data housed at VIREC. Our first step (**Figure 1**) was to cast as wide a net as possible by applying the high-sensitivity algorithm to VA inpatient and outpatient data sets [8]. Afterward, we applied the high-specificity algorithm to those veterans identified by the high-sensitivity algorithm. To meet the first study objective, we abstracted veterans' scrambled Social Security numbers (SCRSSNs) from VA files by applying the high-sensitivity algorithm and then matched them to those SCRSSNs found in the Medicare data sets. We applied the high-specificity algorithm to these same SCRSSNs and identified unique stroke patients from the Medicare claims data [8]. We noted whether the veteran was identified via the VA data, Medicare data, or both.

The second objective of our study was to assess the use of both the VA and Medicare inpatient automated data for determining the stroke event date for research. We searched records obtained from the VA and Medicare claims data sources for codes indicating an acute stroke event within the FY2001 time frame. We initially assigned the discharge date abstracted from these automated data sets as the index date. We used the discharge date rather than the admission date because generally fewer misclassification errors regarding the actual diagnosis in ICD-9 codes occurred at discharge versus at the time of admission. If we found valid dates in more than one database, we assigned the discharge date based on a hierarchy predicted to be the most accurate for this specific purpose. We prioritized the data sources in the following order: (1) Inpatient PTF Main, (2) Medicare Inpatient Claims Data, and (3) Extended Care Main. Our

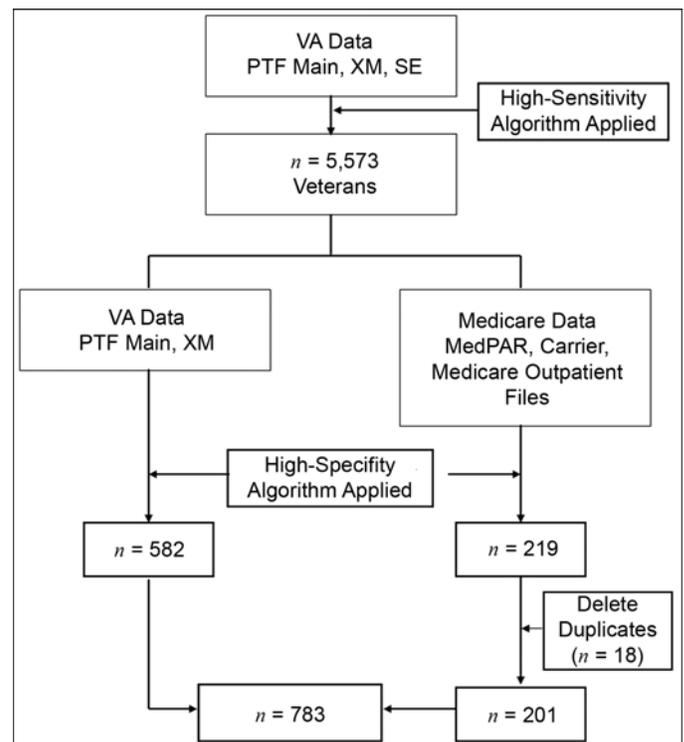


Figure 1.

Flow diagram describing patient identification process. PTF Main = Patient Treatment Files Main, XM = Patient Treatment Files Extended Care, SE = Outpatient Care Event File, MedPAR = Medicare Provider Analysis and Review, Carrier = Medicare Part B carrier, VA = Department of Veterans Affairs.

rationale was to use the two inpatient data sources to assign the primary discharge date because they are likely to be the most accurate, especially the claims data for payment purposes. Next, since no discharge dates were entered in the two inpatient files, the best estimate of discharge date in automated databases is likely the first contact with VHS after a patient's discharge. We had to obtain the discharge date from VA outpatient sources for only four (<1%) veterans. For all three data sources, we abstracted the date used as the "gold standard" from CPRS when available.

We used this same gold standard date for our second objective, confirmation of the stroke event and date within CPRS. We thoroughly reviewed the CPRS to validate the date of the acute stroke. We searched VistAWeb (Veterans Health Information Systems and Technology Architecture intranet Web application VA patient database) to include data for those who also obtained care outside of VISN 8. For all those whose

stroke date could not be confirmed within CPRS or VistAWeb, we noted the reasons. We calculated the mean and median differences in the number of days between the discharge date found in the VA automated data and the CPRS confirmed stroke date. We applied this strategy separately for veterans identified within the VA and Medicare data sets to estimate the difference and variability of the computer-assigned index date for subsequent studies.

Our study protocol was approved by the Institutional Review Board at the University of Florida and the Research and Development Subcommittee for Clinical Investigations at the Malcom Randall VA Medical Center in Gainesville, Florida. We conducted statistical analyses using SAS version 8.1 (SAS Institute, Cary, North Carolina).

RESULTS

Case Identification

When we applied the high-sensitivity algorithm to VA inpatient and outpatient data sets, we identified 5,573 veterans. We then applied the high-specificity algorithm to the automated records and identified 582 unique veterans with acute stroke. When we matched the SCRSSNs of these 5,573 veterans to records in the VIREC-based MedPAR inpatient file, we identified 219 cases using the same high-specificity algorithm. Eighteen veterans were identified in both data sources. When we eliminated the veterans already known from the VA data, we identified 201 additional, unique individuals from the VIREC-based Medicare claims data. Slightly more than 74 percent of the final sample of 783 unique veterans were identified from VA data and nearly 26 percent from the Medicare claims data.

Case Confirmation

Next, we reviewed all 783 cases identified from VA and Medicare databases using the high-specificity algorithm via CPRS for documentation of the stroke date. Overall, nearly 90 percent ($n = 700$) of the acute stroke event dates were confirmed. Sorting by data source, we confirmed nearly 94 percent (545 of 582) of the cases obtained from the VA inpatient data and nearly 77 percent (155 of 201) of the unique cases obtained from the Medicare claims data. That we confirmed nearly 77 percent of the cases using CPRS was surprising, given that the acute nature of these veterans' strokes was initially identified via codes contained in the Medicare database, not VA automated databases.

Among the 11 percent ($n = 83$) that remained unconfirmed by CPRS review, 22 (3%) did not have a stroke (e.g., the discharge ICD-9 code reflected the admission plan to "rule out stroke," but the discharge diagnosis was actually transient ischemic attack [TIA], seizure disorder, migraine). Thirty-one (4%) had their stroke confirmed, but outside the targeted time frame (e.g., the ICD-9 code of 436 was assigned, "acute, but ill-defined cerebrovascular disease" instead of 438, "late effects of cerebrovascular disease"). In addition, 30 (4%) had their stroke confirmed in CPRS, but with no identifiable date in or outside FY2001.

Ideally, one would prefer to identify all veterans with an acute stroke from one database (e.g., only VA data). However, the importance of using both data sources to optimize the yield of veterans with an acute stroke was reinforced by the small overlap among the cases ($n = 18$) contained in both files (**Figure 2**). For example, the likelihood of finding VA-based data cases also within Medicare is only about 10.4 percent and the specificity is 13.6 percent. The positive predictive value (the percentage of positive VA cases that are also Medicare cases) and negative predictive value (the percentage of VA cases absent from the VA data that are also absent from the Medicare data) are 3.3 and 34.9 percent, respectively. Conversely, the likelihood of finding a Medicare case also in the VA data is 3.3 percent, the specificity is 34.9 percent, and the positive and negative predictive values are 10.4 and 13.6 percent, respectively. The sensitivity and specificity data from these two matrixes suggest that, in either case, the use of just one source of data is inadequate compared with the use of both sources.

Confirmation of Stroke Date Using Discharge Date

Finally, we calculated the difference in the number of days between the CPRS-confirmed index date and the discharge date abstracted from the VA automated data to assess using the VA automated data as the stroke index date for epidemiological studies. For the VA data ($n = 545$), the mean \pm standard deviation difference was 19.0 ± 26.2 days and the median difference was 11 days (10th percentile = 2 days; 90th percentile = 43 days), indicating that the average actual stroke event date documented within CPRS was approximately 11 to 19 days before the discharge date abstracted from the automated data. For the Medicare data, the mean \pm SD difference was 6.5 ± 51.0 days and the median difference was 4 days (10th percentile = -19 days; 90th percentile = 27 days).

| (a) Identifying Cases in VA With Medicare Data | | | | | | |
|---|-------|----------|---|----------------|-------|---------------------------|
| | | | | VA Cases | | Actual Cases (<i>n</i>) |
| | | | | + | - | |
| Sensitivity | 3.1% | | | | | |
| Specificity | 95.9% | | | | | |
| PPV | 8.2% | Medicare | + | 18 | 201 | 219 |
| NPV | 89.5% | Cases | - | 564 | 4,790 | 5,354 |
| | | | | 582 | 4,991 | 5,573 |
| (b) Identifying Cases in Medicare With VA Data | | | | | | |
| | | | | Medicare Cases | | Actual Cases (<i>n</i>) |
| | | | | + | - | |
| Sensitivity | 8.2% | | | | | |
| Specificity | 89.5% | | | | | |
| PPV | 3.1% | VA | + | 18 | 564 | 582 |
| NPV | 95.9% | Cases | - | 201 | 4,790 | 4,991 |
| | | | | 219 | 5,354 | 5,573 |

Figure 2.

Identifying veterans' acute stroke cases with Department of Veterans Affairs (VA) and Medicare databases. NPV = negative predictive value, PPV = positive predictive value, + = case identified by database, - = case not identified by database.

DISCUSSION

Our first study objective was to ascertain whether additional unique veterans suffering an acute stroke could be identified from VIREC-based Medicare claims data versus from VA data only. Results revealed that nearly 26 percent ($n = 201$) of unique veterans from the total sample of 783 were identified from the Medicare data when we used the high-specificity algorithm first applied to VA automated data. We confirmed nearly 90 percent (700 of the 783) of the acute stroke events and dates using CPRS, which shows that use of Medicare data is warranted because over three quarters of the time, we were able to document the veteran's stroke within CPRS. These data show that the use of both data sources is important to more accurately estimate the true burden of stroke to our veterans.

In comparing the accuracy of the two data sources, we confirmed approximately 94 percent of the acute stroke events identified from VA data, supporting the accuracy of Reker et al's. high-specificity algorithm when applied specifically to VA data [9]. While one would expect ICD-9 codes indicating an acute stroke to be included in the database of the venue where the veteran received care, a major contribution of this study is that these veterans' strokes could be confirmed within CPRS. Remarkably, we confirmed more than three quarters (77%) of the veterans' strokes within CPRS, even though we identified them from Medicare data using Reker et al's. algorithm. So, keep in mind, our inability to

confirm the Medicare estimate of the stroke date using CPRS does not denote the absence of a stroke within the specified time frame; it simply means that we could not confirm a veteran's stroke using only VA data. Understandably, non-VA clinical information is needed to unequivocally confirm information from automated Medicare sources because it is based on claims for payment.

Importantly, we confirmed close to 77 percent ($n = 155$ of 201) of the stroke events and dates identified from the Medicare data within CPRS. Remarkably, information sufficient for confirming over three quarters of the acute stroke cases found within the Medicare data also could be located in CPRS, although they were not identified from VA automated data. In other words, even though these veterans had confirming clinical information within the computerized medical record, it was not coded in such a way that we could identify it from the VA automated data sets as an acute stroke using the high-sensitivity algorithm. After further investigation, we determined the reasons for missing these cases from within VA data sources were sound. We noted that the data reflected that the veterans' initial hospitalizations for the acute events were at non-VA facilities, from which subsequent transfer or outpatient visit to a VA facility took place after their initial stabilization. In most cases, when the veterans were followed up at the VA, they were assigned the ICD-9 code of 436 (acute but ill-defined cerebrovascular disease) from an outpatient visit. Thus, in the Medicare data, no ICD-9

code was generated to identify an acute stroke event within the inpatient files and so they were identified by the high-sensitivity algorithm but not the high-specificity algorithm based on inpatient VA data. Consequently, while the confirmation rate was lower in the Medicare inpatient data compared with the VA data, the Medicare data contributed substantially to the final sample. These findings lend some credence to study methods that might use Medicare data without incurring the expense of abstracting individual veterans' medical records for nationwide epidemiologic studies. Excluding Medicare data may unintentionally exclude a meaningful proportion of those with acute strokes and underestimate the rate and burden of stroke in veterans.

Given our findings, researchers should be optimistically cautious when identifying veterans for research regarding acute stroke. The selection criteria for classifying the veteran's stroke should be carefully defined, especially when acuity is an important factor. However, even if the researcher carefully selects and rigorously adheres to the criteria, the result is only as good as the coding data entered for later abstraction. Coding for acute stroke samples is not always internally consistent or reliable because of differences in ICD-9 coding across the country. Often the code initially assigned for acute stroke (e.g., 436, acute but ill-defined cerebrovascular disease) follows patients throughout their long-term rehabilitation. In this case, the code should be reassigned (e.g., 438, late effects of cerebrovascular disease) later to reflect that the veteran received care for the sequelae of a previous stroke versus an acute stroke. When strokes are classified based on acuity, a specific time frame incorporated into the codes would be helpful. For example, some specific length of time following the stroke could be categorized as "acute," after which time other codes could be used. Currently, no specific agreed-upon time frame exists.

As described in the previous section, we were not surprised to find that some of the Medicare-based events could not be confirmed within CPRS. However, we were surprised to note that the acute stroke events of 25 veterans identified by the high-specificity algorithm using VA automated inpatient data could not be confirmed in CPRS. After further investigation, we determined that 12 of these were likely to have been (1) coded incorrectly based on a stroke diagnosis on admission but discharged with another diagnosis or (2) mistakenly coded "carotid stenosis with cerebral infarction" rather than "carotid stenosis without infarction." Six had no identifiable stroke date

within CPRS but had mention of recent stroke by self-report to the admitting physician with no other formal documentation of the stroke or stroke date, and seven had the stroke confirmed but outside the specified time frame.

One important implication of these findings is that estimates of acute stroke incidence are likely underestimated and underreported if researchers use only VA data. This has direct implications for estimating the economic, medical, and humanistic costs related to stroke in the nationwide population of veterans. Because the U.S. Government is the primary payer in both VA and Medicare, using either source of data in isolation has fiscal implications for fiscal and provider estimates to care for poststroke treatment and rehabilitation. Complete knowledge of the risk pool provides the Government payer information regarding total stroke care for the population of veterans versus just those receiving care in the VA. More complete information obtained from both systems of care would allow the primary payer, the Federal Government, to develop more inclusive care/treatment strategies for the veterans that could also improve care because of better coordination and potentially reduce overall costs from realizing new economies of scale.

The second objective of our study was to assess using VA and Medicare automated data for identifying the date of veterans' acute stroke for epidemiological studies. The stroke date obtained from the automated data was compared with the date identified in CPRS to estimate case misclassification and time-to-event errors (inherent in any secondary analysis with similar data). The number of days difference in the automated and medical record dates suggests that researchers should be cautious before drawing conclusions solely from automated data sources when the study depends highly on accurate identification of the event (e.g., outcomes studies). For example, the calculated difference between the CPRS date and automated date was "negative" for a segment of the veterans; i.e., the automated discharge date for the alleged acute stroke event was actually before the event if the first documented stroke occurrence found in CPRS is used as the gold standard. This phenomenon was more frequent in the Medicare data than the VA automated data. Since Medicare data are based on claims data, that the acute admission for the stroke was to the non-VA facility and before the admission to the VA facility for follow-up care is plausible. Consequently, the first notation of the acute stroke in CPRS would likely occur when the veteran was admitted to the VA facility after the acute event was

stabilized. In this scenario, without confirming the actual timing of the individual events, the assigned stroke date appears to be clinically implausible and the discharge date assigned to the Medicare claim would rightfully precede the initial VA contact. An additional reason for caution is that we could not confirm a small portion of the veterans defined as having an acute stroke within the Medicare data from CPRS. As before, our inability to confirm the Medicare estimate of the stroke date using CPRS does not denote the absence of a stroke; it simply means that we could not confirm it using only VA data and that non-VA clinical information is needed.

However, the primary implication of this finding is that the index date abstracted from the VA automated data is a reasonable estimate of the acute stroke date in most cases. For example, a reasonable explanation for an average 7-day difference (4-day median) between the Medicare discharge date and the date entered in CPRS is that the veteran was stabilized in the non-VA facility and discharged and that the first episode of VA-based care for the veteran's stroke occurred as a component of their long-term outpatient rehabilitation plan. Under those circumstances, the date of the stroke entered in CPRS might have been based on either discharge documents from the non-VA facility brought by the veteran to his or her initial VA encounter or information gathered during the admission interview. In the case of the 19-day median difference CPRS and the VA data, this explanation is also reasonable, because the automated date was the discharge date and the CPRS stroke date was based on clinical data from the medical record. Consequently, with further investigation, these findings may provide a reasonable "adjustment" for the index date estimate.

Also, with the advent of electronic charting, the practice of cutting and pasting information within past medical history and problem lists needs to be reexamined. While having this information readily available in a busy practice setting is helpful, without due caution and timely revisions, cutting and pasting the information may lead to misclassification of acuity, with patients receiving "current" codes based on prior problems. Another complicating scenario we noted was that codes used to reflect an initial plan to "rule out stroke" were not changed to indicate the final discharge diagnosis (e.g., TIA). Future work should identify and evaluate reasons for common mistakes in assigning ICD-9 codes. The results could be incorporated into training programs for those who assign ICD-9 codes to improve the reliability and validity of

clinical and automated data sources to treat patients and detect acute stroke cases. The effects of erroneous coding can be costly, with clinical as well as administrative, fiscal, quality of care, and research implications.

For these same reasons, researchers should be optimistically cautious when accepting incidence and prevalence statistics, even from authoritative national data sources. For example, the American Heart Association and Centers for Disease Control and Prevention categorize all patients with stroke under the more generic heading "cerebrovascular disease" and group all codes between 430 and 438, excluding 435 for TIA [10–11]. These organizations obtain their estimates of annual incidence from local community-based studies (e.g., Framingham) and then extrapolate to national population estimates. The National Center for Health Statistics also combines new, recurrent, and chronic effects data when reporting national stroke statistics [12]. Thus, only those codes indicating acute stroke events should be considered when stroke incidence is estimated for epidemiological purposes and, without due caution, stroke incidence statistics will be overreported and patients' stroke status may be misclassified for research purposes [13].

The findings from this study should be used within the context of its limitations. First, a segment of U.S. veterans with information available only in Medicare were likely not identified initially by the high-sensitivity algorithm applied to VA data. Under this circumstance, these veterans did not receive any of their acute or long-term post-stroke care in the VA. So, veterans receiving all their care outside VA are likely underrepresented. Conversely, researchers using Medicare data should be aware that the same limitations apply when they include individuals who are also cared for by VA. In either case, the extent to which these veterans are "missed" is unknown and should be estimated if one is to obtain a comprehensive picture. Second, even if dual users are identified within both VA and Medicare data sources, non-VA medical records will occasionally need to be abstracted to unequivocally confirm these veterans' stroke events and dates for research purposes to reduce the likelihood of misclassification error and bias. However, the cost-benefit of doing so must be considered and managed within researchers' resource constraints. Third, even using Medicare data excludes those veterans <65 years who are dual-enrolled in another type of insurance program, such as employer-based insurance programs, although disabled veterans <65 may be represented in the Medicare data. In this case, both the previous

limitations also apply; namely, if non-VA care is these veterans' only source, they may not be included in VA data, and if they are dual users, only a portion of their utilization and care information will be included.

Our use of CPRS as the gold standard for confirming the stroke is only as good as the information contained in CPRS. In most cases, in addition to the ICD-9 code, we examined other clinical indicators of the acute nature of the stroke contained in CPRS before making the final assignment. For example, we used radiology records and nursing and physician notes to confirm the acute nature of the event. At other times, we used these same indicators to disconfirm the automated data (e.g., ICD-9 code 436) when the provider notes clearly indicated that the acute stroke event was coded incorrectly and the event occurred even a decade earlier. However, in other cases, when we were unable to confirm the acute stroke in the VA-based data but found indication of the acute stroke in the Medicare data because the original acute care was provided near the veteran's home and VHS provided follow-up care, we confirmed the stroke. For these reasons, we recommend cautious optimism in researchers' interpretations and providers' incorporation into practice and administrators' decision making based on findings from solely Medicare data or VA data. Finally, because our study focused on facilities within VISN 8, these analyses would have to be replicated with data outside VISN 8 to determine if these results apply to all veterans.

CONCLUSIONS

Research related to stroke has qualitative benefits, including improved outcomes and quality of life for veterans and their families. The research may also have significant fiscal implications for the healthcare delivery system. VA inpatient data sets are highly reliable for identifying acute strokes for research. While Medicare inpatient data require a greater degree of scrutiny for one to confirm its accuracy, excluding veterans identified in Medicare data sets may unintentionally exclude a significant proportion of acute stroke cases from consideration. Use of Medicare inpatient data as an adjunct to VA data appears to provide a more complete picture of veterans' health and utilization and should be done with optimistic caution.

ACKNOWLEDGMENTS

We thank Dean Reker, PhD, for his valuable input and editorial comments. We also thank the staff at the VIREC and the VHA for their assistance in extracting relevant Medicare data.

This material was based on work supported by Rehabilitation Research and Development Service grant D3487R to Dr. Ried and with the resources and use of facilities at the Rehabilitation Outcomes Research Center of Excellence at the Malcom Randall VA Medical Center, Gainesville, Florida.

The views and opinions expressed in this article reflect those of the authors and do not necessarily reflect those of the VA.

The authors have declared that no competing interests exist.

REFERENCES

1. Agency for Health Care Policy and Research. Post-stroke rehabilitation guideline technical report: Vol. 1. Rockville (MD): U.S. Department of Health and Human Services; 1995.
2. Choi D. Stroke. *Neurobiol Dis.* 2000;7(5):552–58. [\[PMID: 11042076\]](#)
3. Reker DM, O'Donnell JC, Hamilton BB. Stroke rehabilitation outcome variation in Veterans Affairs rehabilitation units: Accounting for case-mix. *Arch Phys Med Rehabil.* 1998;79(7):751–57. [\[PMID: 9685086\]](#)
4. Cowper DC. New stroke QUERI: Reducing risks and improving outcomes. *QUERI Quarterly.* 2004;6(2):1–4.
5. O'Sullivan SB, Schmitz JT. Physical rehabilitation: Assessment treatment. 3rd ed. Philadelphia (PA): F.A. Davis; 1994.
6. Jia H, Zheng YE, Cowper DC, Wu SS, Vogel BW, Duncan PW, Reker D. Coordinating care: How veterans use stroke services in the VA and beyond. *Fed Prac.* 2006;23(6):21–24.
7. Jia H, Zheng Y, Reker DM, Cowper DC, Wu SS, Vogel BW, Young GC, Duncan PW. Multiple system utilization and mortality for veterans with stroke. *Stroke.* 2007;38(2): 355–60. [\[PMID: 17194888\]](#)
8. American Medical Association. International classification of diseases, clinical modification. 9th ed. Baltimore (MD): U.S. Department of Health and Human Services, Health Care Financing Administration; 1995. p. 307–10.
9. Reker DM, Hamilton BB, Duncan PW, Yeh SC, Rosen A. Stroke: Who's counting what? *J Rehabil Res Dev.* 2001; 38(2):281–89. [\[PMID: 11392661\]](#)

10. 2001 Heart and Stroke Statistical Update. Dallas (TX): American Heart Association; 2002.
11. Graves EJ, Owings MF. 1996 Summary: National Hospital Discharge Survey. 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 [cited 2007 Aug 14]. Available from: <http://www.cdc.gov/nchs/data/ad/ad301.pdf>
12. National Center for Health Statistics [homepage on the Internet]. Hyattsville (MD): Centers for Disease Control and Prevention; c2006–07 [updated 2006 Dec 4; cited 2007 Aug 14]. Fast Stats A to Z; [about 1 screen]. Available from: <http://www.cdc.gov/nchs/fastats/Default.htm>
13. Williams GR, Jiang JG, Matchar DB, Samsa GP. Incidence and occurrence of total (first-ever and recurrent) stroke. Stroke. 1999;30(12):2523–28. [PMID: 10582972]

Submitted for publication February 14, 2007. Accepted in revised form June 22, 2007.