

## Race/ethnicity: Who is counting what?

**Huanguang Jia, PhD;<sup>1\*</sup> Yu E. Zheng, MS;<sup>1–2</sup> Diane C. Cowper, PhD;<sup>1</sup> James P. Stansbury, PhD;<sup>1,3</sup> Samuel S. Wu, PhD;<sup>1–2</sup> W. Bruce Vogel, PhD;<sup>1</sup> Pamela W. Duncan, PhD;<sup>1</sup> Dean M. Reker, PhD<sup>1</sup>**

*<sup>1</sup>Rehabilitation Outcomes Research Center, Gainesville Department of Veterans Affairs Medical Center, Gainesville, FL; Departments of <sup>2</sup>Statistics and <sup>3</sup>Anthropology, University of Florida, Gainesville, FL*

**Abstract**—Misclassification of race and ethnicity in administrative data may produce misleading results if it is overlooked or ignored. In this study, we examined the racial/ethnic classifications of 1,084 veterans with stroke in Florida who received inpatient and outpatient services within the Department of Veterans Affairs (VA) healthcare system and who were also eligible for Medicare between 2000 and 2001. We compared the reliability of racial/ethnic classifications between VA inpatient data, VA outpatient data, and Medicare data. Our results showed that (1) the rate of unknown racial/ethnic classification in VA outpatient and inpatient data was high, (2) minimizing the unknowns by substituting known values from other data when available would greatly enhance the overall and individual classification reliability, (3) black and white classifications in the VA data had stronger agreement with Medicare data, and (4) Medicare data may under-represent Hispanic patients.

**Key words:** administrative data, ethnicity, Department of Veterans Affairs, Medicare, race, racial/ethnic classification, reliability, Social Security Administration, stroke, validity, veterans.

### INTRODUCTION

Stroke is one of the leading causes of death and a major cause of serious and long-term disability in the United States population [1] as well as in the veteran population [2–3]. Numerous studies have shown racial/ethnic disparities in stroke. In a recent study using a continuous annual household-based survey of the United States population from 1999 to 2001, McGruder et al.

reported that racial/ethnic disparities existed in stroke prevalence, risk behaviors, and medical history [4]. In a recent selective review of literature published between 1991 and 2003, Stansbury et al. provided a broad perspective on race/ethnicity across the continuum of stroke care [5]. Their review confirmed the increased stroke incidence and mortality among black Americans and demonstrated the higher prevalence of hypertension among black and Hispanic patients and atrial fibrillation among non-Hispanic white stroke patients.

Within the Department of Veterans Affairs (VA) healthcare system, Jones et al. examined racial variation and initial stroke severity among 982 veteran patients who received acute stroke care [6]. Using the Canadian Neurological Scale, Jones et al. showed that unadjusted scores were significantly lower for black patients than for white patients (7.96 vs 9.32,  $p = 0.04$ ), which indicated

---

**Abbreviations:** CMS = Centers for Medicare and Medicaid Services, CY = calendar year, EDB = enrollment database, FY = fiscal year, HCFA = Health Care Financing Administration, MBR = master beneficiary record, PTF = Patient Treatment File, SAS = Statistical Analysis Software, SSA = Social Security Administration, VA = Department of Veterans Affairs, VAMC = VA medical center, VIREC = VA Information Resource Center.

\*Address all correspondence to Dr. Huanguang Jia, PhD; VA Rehabilitation Outcomes Research Center 151B, 1601 SW Archer Rd, Gainesville, FL 32608; 352-376-1611, ext 4926; fax: 352-271-4540. Email: [Huanguang.Jia@med.va.gov](mailto:Huanguang.Jia@med.va.gov)

DOI: 10.1682/JRRD.2005.05.0086

greater stroke severity among black patients. Reports on racial/ethnic disparities in the use of stroke-related procedures and services showed mixed results. In an earlier study among a national sample of veterans with ischemic stroke, Oddone et al. reported that the likelihood of angiography for black patients was 47 percent that of white patients and Hispanic patients were only 75 percent as likely as non-Hispanic white patients to receive the procedure [7]. On the contrary, Goldstein et al. found no significant differences between white and nonwhite patients in the use of a variety of stroke-related procedures and services among 1,073 veterans diagnosed with acute ischemic stroke [8]. In two separate studies on racial/ethnic disparities in poststroke rehabilitation outcomes, Horner et al. showed that black veteran patients had greater residual physical deficits from stroke than white patients [9] and that low-income black veteran patients experienced poorer functional recovery over a 12-month period [10]. Most of these studies were based on data extracted from the VA inpatient and/or outpatient databases. Less, however, is known about the agreement of racial/ethnic classification for stroke patients between the VA inpatient and outpatient databases and Medicare data.

Although the overall veteran population has been declining, VA healthcare enrollees have been increasing. The number of veteran patients age 65 and over is projected to increase rapidly in the next 10 years. Thus, we can expect that the number of individuals who are dually eligible for VA and Medicare programs will continue to increase. This is the case particularly for veterans with stroke, since the disease occurs more frequently in the elderly. Several VA researchers have compared the characteristics of veterans who received services under the two healthcare programs (VA and Medicare) [11–13]. No publications, however, have compared the racial/ethnic classification of veteran stroke patients between the VA and Medicare data sources.

The VA has a long history of computerizing its clinical information. At present, data for all episodes of care provided within the system are transmitted by the local VA facility to a central repository at the VA Austin Automation Center (Austin, Texas), where specific clinical data elements are maintained in various SAS (Statistical Analysis Software) data sets (SAS Institute, Cary, North Carolina). Among the commonly used data sets are inpatient and outpatient care, extended care, and inpatient short-stay observation care. In these data sets, not all

variables are of equal quality, particularly regarding the number of records with missing information. For example, for fiscal year (FY) 2000 data, patient race/ethnicity was missing from 4.5 percent of records in the main inpatient data set and from 24.9 percent of records in the outpatient visit data set [14–15]. The reliability and validity of the VA national data on race has not been definitively assessed in the published literature [16].

This study examined the racial/ethnic classifications of 1,084 veteran stroke patients who received inpatient and outpatient services within the VA healthcare system and who were also eligible for Medicare between 2000 and 2001. We calculated the racial/ethnic classification agreement for these stroke patients between VA inpatient and outpatient data as well as VA and Medicare data. Such information will be helpful in understanding the quality of racial/ethnic disparity studies. As previous studies have indicated, misclassification of race/ethnicity in research data may produce spurious conclusions if the misclassifications are overlooked or ignored [17]. Specifically, use of existing administrative data for assessing differences in rehabilitation outcomes and racial/ethnic rehabilitation outcome disparities may cause misleading results [18].

## METHODS

### Data Source

For this study, we used the VA inpatient (Patient Treatment File [PTF]) and outpatient medical SAS data sets [14–15,19] housed at the VA Austin Automation Center and the Medicare denominator file (for VA patients only) obtained from the VA Information Resource Center (VIReC) in Hines, Illinois. The PTF is the primary source of inpatient data for veterans treated in the VA healthcare system. It includes three major inpatient files: main, extended, and bed section. The main and extended care data sets contain patient's demographic information and summary information about each inpatient episode. The PTF bed section data set contains more specific information on the bed section to which patients were transferred for each hospital stay. For this study, the PTF main file was used. The outpatient file has two subsets: the outpatient visit and outpatient event files. These files contain diagnostic and procedural information on each outpatient encounter. The outpatient visit file was used in this study. The VA data are stored by Federal FY,

whereas the Medicare data are stored by calendar year (CY). To make these different time frames compatible, we consistently used the CY system in these analyses. Three years of data (FY2000–2002) from VA databases were extracted for compatibility with 2 years of data (CY2000–2001) from Medicare databases.

### Sample Selection

In a retrospective cohort study designed to investigate VA and non-VA rehabilitation services use by veterans with stroke, 1,084 patients were identified who (1) lived in the state of Florida, (2) had a stroke diagnosis as defined by Reker et al.'s high sensitivity International Classification of Disease, 9th edition, codes [20], (3) received VA inpatient and outpatient services, and (4) were also identified in the Medicare denominator file between 2000 and 2001. The study was approved by the institutional review board at the University of Florida and the Research and Development Subcommittee for Clinical Investigations at the VA medical center (VAMC) in Gainesville, Florida.

### Racial/Ethnic Classification

We used the original racial/ethnic classification in each of the three data sources (VA inpatient, VA outpatient, Medicare denominator file). The origin of race/ethnicity in the VA databases is different from that in the Medicare denominator file. In VA data, race/ethnicity is most frequently extracted from clinical documents and/or observation of the administrative staff. The classifications in the inpatient and outpatient data include white, black, Hispanic black, Hispanic white, American Indian, Asian, and unknown. Racial/ethnic classification in Medicare data ideally is self-reported. The Medicare denominator file that is released annually contains the beneficiaries' demographic data and eligibility status for public use. The primary source for identifying the beneficiaries' race/ethnicity in the denominator file is the administrative enrollment database (EDB) maintained by the Centers for Medicare and Medicaid Services (CMS) (known as the Health Care Financing Administration [HCFA] during the study period). The EBD is created based on the information from the Social Security Administration's (SSA's) master beneficiary record (MBR) file, and the original source of the MBR's race/ethnicity is the SSA's Form SS-5. This form is completed voluntarily by a person when applying for a Social Security number or for a replacement Social Security card.

Before September 1980, the race item in the SS-5 form consisted of three categories: white, black, or other. Thereafter, the race item was expanded to include white (not Hispanic); black (not Hispanic); Hispanic; Asian, Asian-American, or Pacific Islander; and North American Indian or Alaskan Native.

The HCFA took about a decade to integrate the SSA's expanded racial/ethnic information in its EDB and denominator file. Meanwhile, special efforts, such as surveys among the target Medicare beneficiaries, were taken by HCFA to improve the quality and accuracy of its racial/ethnic classification data. At present, the race variable in the denominator file consists of the following seven categories: white, black, other, Asian, Hispanic, North American Native, and unknown [17,21]. In this study, we merged the Asian classification into "other" because of the small amount of data.

Several studies have examined the reliability and validity of racial/ethnic classification in Medicare data [17,21–22]. By linking the Medicare Current Beneficiary Survey data with the EDB, Arday et al. [21], for example, assessed the accuracy of racial/ethnic classification in the Medicare EDB before and after the update in 1997. The study showed that after the 1997 update, the sensitivity of the EDB was 97 percent for white, 95 percent for black, and <60 percent for all other categories and the positive predictive value was >96 percent for white, black, and Hispanic persons, but <80 percent for all others.

### Analysis

For this study, we calculated crude concordance and kappa ( $\kappa$ ) statistics to measure the agreement rate and strength of racial/ethnic classifications between the different data sources: VA inpatient versus VA outpatient and Medicare versus VA. The statistic is commonly used to examine interrater reliability; it compares the probability of agreement to that expected if the ratings were independent [23]. It has also been used in calculating racial/ethnic classification reliability in different studies [18,21,24]. Because of a lack of consensus in interpreting the  $\kappa$  values between 0 and 1, we followed Altman's guidelines [25] in interpreting the strength of agreement in this study: poor agreement 0.20, fair = 0.21 to 0.40, moderate = 0.41 to 0.60, good = 0.61 to 0.80, and very good  $\geq 0.81$ .

## RESULTS

**Table 1** summarizes the frequencies of each racial/ethnic classification in VA inpatient and outpatient databases. Two sets of information are presented in the table: the “Index” rows show the frequencies of initial or index racial/ethnic classifications of the 1,084 patients in the two databases and the “With Replacement” rows list the frequencies of each classification after unknowns were replaced by the next most recent available racial/ethnic value from the same database in a 2-year period. To be specific, if a patient’s index or initial race/ethnicity was coded as unknown, we then checked his or her same data file for the next most recent value and replaced the unknown with the next available known racial/ethnic value. During this process, 253 unknowns (24 in inpatient and 229 in outpatient data) were identified and recoded in the data according to the next most recent available racial/ethnic value.

**Table 2** shows the agreement rate (concordance) and agreement strength ( $\kappa$ ) of the overall and individual clas-

sification both before and after replacement between VA inpatient and outpatient data. As shown, the replacement of the unknowns with the next available value greatly enhanced the reliability of overall and several individual classifications. Because of the higher number of unknown designations among white stroke patient index records, the improvement was most apparent for this group. Nonetheless, considerable improvements in reliability were noted for all minority ethnic groups as well as in the overall results.

As we discussed in the Methods section, race/ethnicity was classified differently in the VA and Medicare data. In the VA data, Hispanics are categorized as Hispanic black and Hispanic white, whereas in the Medicare data only one category exists for Hispanic. In order to make these different classifications comparable, we created a final VA classification file for the study cohort by adjusting the VA data (both inpatient and outpatient) to make it categorically equivalent to the Medicare data. The process started with merging the VA inpatient-outpatient information by

**Table 1.**

Racial/ethnic classification matrix for Department of Veterans Affairs inpatient and outpatient databases.

		Outpatient Classification						Inpatient Total
		White	Black	H. Black	H. White	Asian	Unknown	
Inpatient Classification	Index	603	1	3	0	0	195	802
	With Replacement	799	1	3	0	0	19	822
	Index	2	111	0	0	0	23	136
	With Replacement	2	135	0	0	0	2	139
	Index	1	0	26	1	0	10	38
	With Replacement	1	0	36	1	0	0	38
	Index	0	0	0	3	0	1	4
	With Replacement	1	0	0	4	0	0	5
	Index	0	0	0	0	1	0	1
	With Replacement	0	0	0	0	1	0	1
Outpatient Total	Index	11	2	1	0	0	89	103
	With Replacement	8	2	1	0	0	68	79
	Total	617	114	30	4	1	318	1,084
	With Replacement	811	138	40	5	1	89	1,084

Note: Index rows show initial racial/ethnic classification at index inpatient hospitalization of first-noted outpatient visit. With Replacement rows show classification after unknown values were replaced by next available known value in same data file.

H = Hispanic.

**Table 2.**

Percent concordance and reliability coefficient ( $\kappa$ ) of racial/ethnic classifications between Department of Veterans Affairs inpatient and outpatient data.

Inpatient	n	Concordance (%)	$\kappa$	95% CI
<b>White</b>				
Index	802	80.4	0.58	0.53 to 0.63
With Replacement	822	96.8	0.91	0.88 to 0.94
<b>Black</b>				
Index	136	97.4	0.87	0.83 to 0.92
With Replacement	139	99.4	0.97	0.95 to 0.99
<b>Hispanic-Black</b>				
Index	38	98.5	0.76	0.64 to 0.87
With Replacement	38	99.4	0.92	0.86 to 0.98
<b>Hispanic-White</b>				
Index	4	99.8	0.75	0.41 to 1.00
With Replacement	5	99.8	0.80	0.53 to 1.00
<b>Asian</b>				
Index	1	100.0	1.00	1.00 to 1.00
With Replacement	1	100.0	1.00	1.00 to 1.00
<b>Unknown</b>				
Index	103	77.6	0.33	0.27 to 0.38
With Replacement	79	97.0	0.79	0.72 to 0.87
<b>Overall</b>				
Index	1,084	76.8	0.57	0.53 to 0.61
With Replacement	1,084	96.2	0.91	0.88 to 0.94

Note: Index rows show agreement and reliability coefficient of initial racial/ethnic classification at index inpatient hospitalization or first-noted outpatient visit. With Replacement rows show agreement and reliability coefficient after unknowns were replaced by next available known value in same data file. Concordance refers to proportion of 1,084 patients in which two classifications were concordant in dichotomous classification for each individual group and in multiple categories (white, black, Hispanic-white, Hispanic-black, Asian, unknown) for overall classification (e.g., in 80.4% of cases, patients were concordantly classified according to dichotomy white/nonwhite).

CI = confidence interval.

keeping the perfect matches in the two VA data files after the replacement as shown in **Table 1**. Second, we further replaced the unknowns in one data file with the known values in the other data file. Third, we resolved the nine contradictory classifications between the two databases (**Table 1**) by further checking the individual's classification in both data files and choosing the most frequent value. As a result, from the nine controversial classifications, three were reclassified as white, two as black, and four as Hispanic in the final file. Finally, we merged the Hispanic white with the Hispanic black data to make the classification consistent with the Hispanic category in Medicare data. The frequencies of the overall and individual classifications between the final VA file and Medicare file are presented in **Table 3**. Interestingly, among the 907 whites in the Medicare data, 29 of them were identified as Hispanic in the VA data.

**Table 4** shows the agreement rate (concordance) and strength ( $\kappa$ ) of the overall and individual classifications

between the final VA file and the Medicare data. After more of the unknowns were replaced and the nine controversial classifications between the VA inpatient and outpatient data were resolved, the classification agreement rate (concordance) and strength ( $\kappa$ ) between the two data sources were 88.5 percent and 0.66 overall, 89.9 percent and 0.68 for white, 97.3 percent and 0.89 for black, 96.7 percent and 0.37 for Hispanic, and 99.6 percent and 0.28 for other. Following Altman's guidelines [25], these values indicated that the agreement was very good for black, good for the overall and white, and fair for Hispanic and other.

## DISCUSSION

This study is the first to compare racial/ethnic reliability between Medicare and VA data among veterans diagnosed with stroke. Understanding racial variation has been a primary focus for many VA researchers [6–10],

**Table 3.**

Racial/ethnic classification between Department of Veterans Affairs (VA) and Medicare data. Numbers in bold indicate classifications that were identical in VA and Medicare data.

Medicare Classification	VA Classification					Medicare Total
	White	Black	Hispanic	Other	Unknown	
White	<b>813</b>	6	29	0	59	907
Black	10	<b>134</b>	3	0	9	156
Hispanic	2	0	<b>11</b>	0	0	13
Other	2	1	2	<b>1</b>	0	6
Unknown	2	0	0	0	<b>0</b>	2
VA Total	829	141	45	1	68	1,084

**Table 4.**

Percentage of concordance and reliability coefficient ( $\kappa$ ) of racial/ethnic classification between Department of Veterans Affairs data and Medicare data.

Classification	n	Concordance (%)	$\kappa$	95% CI
White	907	89.8	0.68	0.63 to 0.74
Black	156	97.3	0.89	0.85 to 0.93
Hispanic	13	96.7	0.37	0.21 to 0.53
Other	6	99.6	0.28	-0.15 to 0.72
Unknown	2	93.5	-0.00	-0.01 to 0.00
Overall	1,084	88.5	0.66	0.61 to 0.71

Note: Concordance refers to proportion of 1,084 patients in which two classifications were concordant in dichotomous classification for each individual group and in multiple categories (white, black, Hispanic, other, unknown) for overall classification (e.g., in 89.9% of cases, patients were concordantly classified according to dichotomy white/nonwhite).

CI = confidence interval.

and several studies have examined reliability and validity of racial/ethnic classification in the Medicare data [17,21–22]. However, less is understood about the quality of the racial/ethnic classification in the commonly used VA inpatient and outpatient databases compared with Medicare data.

We found a high number of unknowns in racial/ethnic classification in both VA inpatient and outpatient databases. Further, three times as many unknowns existed in the outpatient data as in the inpatient data (29.3% vs 9.5%, respectively). These results indicate that to better understand racial/ethnic variations among veterans with stroke, use of just the initial cross-sectional classification information provided in these two databases is not enough. If the unknowns are overlooked or ignored, biased or spurious conclusions may be produced.

One recommendation for managing this issue is to check all the unknown classifications in the same database (inpatient or outpatient) and replace the unknowns with the next most recent available racial/ethnic value. This would help minimize the number of unknowns and improve the reliability of racial/ethnic classification in the

databases. Our findings among the 1,084 stroke patients showed that 23.3 percent of the unknown classifications in the inpatient data and 72.0 percent of the unknown classifications in the outpatient data could be identified and replaced by the next most recent available value in the same database. Further, these updates could help improve the reliability of the overall and several of the individual classifications by different degrees (Table 2).

Race/ethnicity in the Medicare denominator file originates in the SSA's MBR file and is generally regarded as self-reported. As such, some researchers prefer using the classification in Medicare data over the VA data [26–27]. With the recent cooperative agreement between CMS and VA, the VIREC is receiving annual Medicare data extracts directly from CMS for veterans who are dually eligible for both healthcare programs: VA and Medicare. At present, VIREC has Medicare Part A and B data for CY2000 through 2003. This provides a great opportunity for VA researchers who are engaged and interested in VA and non-VA access and services-use studies. Therefore, our findings about the racial/ethnic classification agreement

between the two sources of data (VA and Medicare) are important for researchers in the field.

We found that the overall classification agreement between the VA and Medicare data for the 1,084 veteran stroke patients was good (concordance rate = 88.5%,  $\kappa = 0.66$ , 95% confidence interval = 0.61 to 0.71). The agreement rate and strength, however, varied among the individual classifications: very good for black classification, good for white, and fair for Hispanic and other. In our analysis, we found that both other and Hispanic classifications had high agreement rates (99.6% and 96.7%, respectively) but their values were low (0.28 and 0.37, respectively) compared with the black classification in **Table 4**. The  $\kappa$  statistic is very sensitive to marginal values; if the marginal values are not well balanced, the value will be low even if the concordance rate is high [28]. We should reiterate that the VA classifications used in the comparison with the Medicare data resulted from a process of merging the inpatient and outpatient information and replacing unknown values as we discussed in the Results section.

While a broad consensus exists in the literature regarding the desirability of self-reporting of racial/ethnic information, sometimes the assumption is that Medicare data represent self-reporting and that assumption may be in error. Apparently for this study cohort, veterans living in Florida with stroke, the Medicare data may under-represent Hispanic identification (**Table 3**). This may be because of the race item in the SS-5 form that was used prior to September 1980. The three categories—white, black, and other—in the old form may cause Hispanics to be reported in the white category. Further study will be required to confirm that this is the case and to begin to understand why it may be the case.

Several limitations in this study need to be addressed. First, the generalizability of our findings is limited by our focus on one region of the VA system. It would be important to understand the racial/ethnic classifications of stroke patients, as well as VA patients in general, in different geographic regions in future studies. Our future investigations will address these issues. Second, our findings about the racial/ethnic classifications in the VA inpatient and outpatient data do not represent the classifications in these databases today. Beginning with FY2003, the VA medical SAS data show new classifications of patient race/ethnicity in response to the U.S. Office of Management and Budget and Department of Health and Human Services guidelines [29–30]. The new

data are self-reported, consist of two domains: race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or other Pacific Islander, White) and ethnicity (Spanish, Hispanic, or Latino; No, not Spanish, Hispanic, or Latino) [16,31].

## CONCLUSIONS

We found that the unknown racial/ethnic classification in VA inpatient and outpatient databases was high. Minimizing the unknowns by replacing them with the known values in the same or different databases would greatly improve the overall and individual classification reliability. The black and white classifications in the VA data had stronger agreement with that in the national Medicare denominator file. Medicare data may under-represent Hispanic identification.

## ACKNOWLEDGMENTS

Yu E. Zheng is now with the Harvard School of Public Health, Harvard University, Boston, Massachusetts. Dean M. Reker is now with the Kansas City VAMC, Kansas City, Missouri.

This material was based on work supported by the VA Health Services Research and Development (HSR&D) Service (grant IIR 03-151-1) and the VA HSR&D and Rehabilitation Research and Development Services (grant ROC 01-124). The views and opinions expressed in this manuscript 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. Gresham GE, Duncan PW, Adams HP Jr, Adelman AM, Alexander DN, Bishop DS, Diller L, Donaldson NE, Granger CV, Holland AL, Kelly-Hayes M, McDowell FH, Myers LR, Phipps MA, Roth EJ, Siebens HC, Tarvin GA, Trombly CA. Post-stroke rehabilitation. Clinical practice guideline. No. 16. Rockville (MD): Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services; 1995.
2. O'Donnell JC, Hamilton BB. Stroke rehabilitation management in the Department of Veteran's Affairs: Impact of

- patient referral source on outcomes. *Arch Phys Med Rehabil.* 1997;78(9):929–37. [\[PMID: 9305263\]](#)
3. Reker DM, O'Donnell JC, Hamilton BB. Stroke rehabilitation outcome variation in Veteran's Affairs rehabilitation units: Accounting for case-mix. *Arch Phys Med Rehabil.* 1998;79(7):751–57. [\[PMID: 9685086\]](#)
  4. McGruder HF, Malarcher AM, Antoine TL, Greenlund KJ, Croft JB. Racial and ethnic disparities in cardiovascular risk factors among stroke survivors: United States 1999 to 2001. *Stroke.* 2004;35(7):1557–61. [\[PMID: 15192252\]](#)
  5. Stansbury JP, Jia H, Williams LS, Vogel WB, Duncan PW. Ethnic disparities in stroke: Epidemiology, acute care, and postacute outcomes. *Stroke.* 2005;36(2):374–86. [\[PMID: 15637317\]](#)
  6. Jones MR, Horner RD, Edwards LJ, Hoff J, Armstrong SB, Smith-Hammond CA, Matchar DB, Oddone EZ. Racial variation in initial stroke severity. *Stroke.* 2000;31(3):563–67. [\[PMID: 10700486\]](#)
  7. Oddone EZ, Horner RD, Monger ME, Matchar DB. Racial variations in the rates of carotid angiography and endarterectomy in patients with stroke and transient ischemic attack. *Arch Intern Med.* 1993;153(24):2781–86. [\[PMID: 8257254\]](#)
  8. Goldstein LB, Matchar DB, Hoff-Lindquist J, Samsa GP, Horner RD. Veterans Administration Acute Stroke (VASt) Study: Lack of race/ethnic-based differences in utilization of stroke-related procedures or services. *Stroke.* 2003;34(4):999–1004. [\[PMID: 12649513\]](#)
  9. Horner RD, Matchar DB, Divine GW, Feussner JR. Racial variations in ischemic stroke-related physical and functional impairments. *Stroke.* 1991;22(12):1497–1501. [\[PMID: 1962323\]](#)
  10. Horner RD, Swanson JW, Bosworth HB, Matchar DB; VA Acute Stroke (VASt) Study Team. Effects of race and poverty on the process and outcome of inpatient rehabilitation services among stroke patients. *Stroke.* 2003;34(4):1027–31. [\[PMID: 1264220\]](#)
  11. Fleming C, Fisher ES, Chang CH, Bubolz TA, Malenka DJ. Studying outcomes and hospital utilization in the elderly. The advantages of a merged data base for Medicare and Veterans Affairs hospitals. *Med Care.* 1992;30(5):377–91. [\[PMID: 1583916\]](#)
  12. Wright SM, Petersen LA, Lamkin RP, Daley J. Increasing use of Medicare services by veterans with acute myocardial infarction. *Med Care.* 1999;37(6):529–37. [\[PMID: 10386565\]](#)
  13. Borowsky SJ, Cowper DC. Dual use of VA and non-VA primary care. *J Gen Intern Med.* 1999;14(5):274–80. [\[PMID: 10337036\]](#)
  14. Veterans Affairs Information Resource Center (VIREC). The Medical SAS® Inpatient Data sets—FY2000: A VIREC Resource Guide. Hines (IL): Edward J. Hines, Jr. VA Hospital; 2001.
  15. Veterans Affairs Information Resource center (VIREC). The Medical SAS® Outpatient Data sets—FY2000: A VIREC Resource Guide. Hines (IL): Edward J. Hines, Jr. VA Hospital; 2001.
  16. VIREC exploring options for researchers needing data on race. VIREC Researchers' Guide to VA Data: Data Issue Brief. March 2004.
  17. Lauderdale DS, Goldberg J. The expanded racial and ethnic codes in the Medicare data files: Their completeness of coverage and accuracy. *Am J Public Health.* 1996;86(5):712–16. [\[PMID: 8629724\]](#)
  18. Stansbury JP, Reid KJ, Reker DM, Duncan PW, Marshall CR, Rittman M. Why ethnic designation matters for stroke rehabilitation: Comparing VA administrative data and clinical records. *J Rehabil Res Dev.* 2004;41(3A):269–78. [\[PMID: 15543444\]](#)
  19. Murphy PA, Cowper DC, Seppala G, Stroupe KT, Hynes DM. Veterans Health Administration inpatient and outpatient care data: An overview. *Eff Clin Pract.* 2002;5(3 Suppl):E4. [\[PMID: 12166925\]](#)
  20. Reker DM, Hamilton BB, Duncan PW, Yeh SC, Rosen A. Stroke: Who's counting what? *J Rehabil Res Dev.* 2001;38(2):281–89. [\[PMID: 11392661\]](#)
  21. Arday SL, Arday DR, Monroe S, Zhang J. HCFA's racial and ethnic data: Current accuracy and recent improvements. *Health Care Financ Rev.* 2000;21(4):107–16. [\[PMID: 11481739\]](#)
  22. Eggers PW, Greenberg LG. Racial and ethnic differences in hospitalization rates among aged Medicare beneficiaries, 1998. *Health Care Financ Rev.* 2000;21(4):91–105. [\[PMID: 11481748\]](#)
  23. Agresti A. Categorical data analysis. 2nd ed. Hoboken (NJ): John Wiley and Sons, Inc; 2002.
  24. Blustein J. The reliability of racial classifications in hospital discharge abstract data. *Am J Public Health.* 1994;84(6):1018–21. [\[PMID: 8203669\]](#)
  25. Altman D. Practical statistics for medical research. London: Chapman and Hall; 2001.
  26. Halanych JH, Miller DD, Lin H, Wang F, Berlowitz D, Pogach LM, Frayne SM. Disparities in diabetes care in Hispanic, black, and white veterans who use VHA services [abstract]. In: HSR&D National Meeting 2005: Improving Care for Chronic Illnesses; 2005 Feb 16–18; Baltimore, MD.
  27. Tseng CL, Rajan M, Miller DR, Hawley G, Crystal S, Xie M, Tiwari A, Safford M, Pogach L. Use of administrative data to risk adjust amputation rates in a national cohort of Medicare-enrolled veterans with diabetes. *Med Care.* 2005;43(1):88–92. [\[PMID: 15323928\]](#)
  28. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol.* 1990;43:543–549. [\[PMID: 2348207\]](#)

- 
- 29. United States Department of Health and Human Services [homepage on the Internet]. Washington (DC): United States Department of Health and Human Services; c1997–2006 [updated 1997 Oct 24; cited 2005 Apr 30]. Policy for improving race and ethnicity data [1 screen]. Available from: <http://aspe.os.dhhs.gov/datacnci/inclusn.htm>
  - 30. New race and ethnicity variables added to VHA medical SAS inpatient main data set. VIREC Researchers' Guide to VA Data: Data Issue Brief. June 2003.
  - 31. Rosewell RH. VHA Directive 2003–027. Capture of race and ethnicity categories. May 29, 2003. Available from: [http://vaww1.va.gov/vhapublications/ViewPublication.asp?pub\\_ID=253](http://vaww1.va.gov/vhapublications/ViewPublication.asp?pub_ID=253)

Submitted for publication May 13, 2005. Accepted in revised form December 9, 2005.