

## Use of Medicare and DOD data for improving VA race data quality

Kevin T. Stroupe, PhD;<sup>1–3\*</sup> Elizabeth Tarlov, RN, PhD;<sup>1–2</sup> Qiuying Zhang, MS;<sup>1</sup> Thomas Haywood, MPH;<sup>2</sup> Arika Owens, MPH;<sup>2</sup> Denise M. Hynes, RN, MPH, PhD<sup>1–2,4</sup>

<sup>1</sup>Center for Management of Complex Chronic Care, Edward Hines Jr. Department of Veterans Affairs (VA) Hospital, Hines, IL; <sup>2</sup>VA Information Resource Center, Hines, IL; <sup>3</sup>Institute for Healthcare Studies, Feinberg School of Medicine, Northwestern University, Chicago, IL; <sup>4</sup>Department of Medicine, College of Medicine, University of Illinois, Chicago, IL

**Abstract**—We evaluated the improvement in Department of Veterans Affairs (VA) race data completeness that could be achieved by linking VA data with data from Medicare and the Department of Defense (DOD) and examined agreement in values across the data sources. After linking VA with Medicare and DOD records for a 10% sample of VA patients, we calculated the percentage for which race could be identified in those sources. To evaluate race agreement, we calculated sensitivities, specificities, positive predictive values (PPVs), negative predictive values, and kappa statistics. Adding Medicare (and DOD) data improved race data completeness from 48% to 76%. Among older patients ( $\geq 65$  years), adding Medicare data improved data completeness to nearly 100%. Among younger patients ( $< 65$  years), combining Medicare and DOD data improved completeness to 75%, 18 percentage points beyond that achieved with Medicare data alone. PPVs for white and African-American categories were 98.6 and 94.7, respectively, in Medicare and 97.0 and 96.5, respectively, in DOD data using VA self-reported race as the gold standard. PPVs for the non-African-American minority groups were lower, ranging from 30.5 to 48.2. Kappa statistics reflected these patterns. Supplementing VA with Medicare and DOD data improves VA race data completeness substantially. More study is needed to understand poor rates of agreement between VA and external sources in identifying non-African-American minority individuals.

**Key words:** African American, data quality, ethnicity, health science, Hispanic, Medicare, minority, race, validity, veterans.

## INTRODUCTION

Race/ethnicity-based differences in healthcare and health status in the United States are well known and continue to receive much research attention. Research has demonstrated that U.S. minorities, particularly African-American and Hispanic patients, receive lower quantity and quality of healthcare in many settings and for a wide range of conditions. Many of these differences are not explained by clinical factors, patient preferences, or ability to pay (as measured by health insurance and income) and thus represent inequities in care.

While these disparities have been well documented, their root causes and solutions remain unclear, requiring

**Abbreviations:** APIO = Asian, Pacific Islander, or Other; DEERS = Defense Enrollment Eligibility Reporting System; DMDC = Defense Manpower Data Center; DOD = Department of Defense; FY = fiscal year; MedSAS = Medical SAS (data sets); NPCD = National Patient Care Database; PPV = positive predictive value; SSA = Social Security Administration; SSN = Social Security number; VA = Department of Veterans Affairs; VADIR = VA/DOD Identity Repository; VHA = Veterans Health Administration.

\*Address all correspondence to Kevin T. Stroupe, PhD; Center for Management of Complex Chronic Care, Edward Hines Jr. VA Hospital (151H), 5000 South 5th Ave, Bldg 1B260, Hines, IL 60141-5151; 708-202-3557; fax: 708-202-2316. Email: [Kevin.stroupe@va.gov](mailto:Kevin.stroupe@va.gov)

DOI:10.1682/JRRD.2009.08.0122

further research [1–4]. Monitoring progress toward eliminating disparities in health and healthcare is a major U.S. public health goal [1–2,5]. One challenge to these research and monitoring activities in the United States is the paucity of reliable and consistent collection and reporting of race/ethnicity data [6–8].

The Department of Veterans Affairs (VA) Office of Research and Development has identified health disparities and minority health as a priority research area [9], and many VA studies with race/ethnicity as a central focus are in progress or have been completed [10–22]. As the largest integrated healthcare system in the United States and a pioneer in electronic health information, the VA has vast data stores that provide rich opportunities for health services research. In addition to clinical information, VA databases frequently used in research contain patient demographic information including race/ethnicity. However, the quality and completeness of this race/ethnicity information has been identified as a potential limitation to research [23–26].

Obtaining veteran race/ethnicity information from external sources, including Medicare and Department of Defense (DOD) databases, has the potential to improve data completeness in VA research studies. However, little information is available to inform researchers about the utility of this approach. In this study, we evaluated the improvement in VA race data completeness that could be achieved by linking VA data with data from Medicare and DOD. Further, we examined the agreement in race values between the Veterans Health Administration (VHA) and these external data sources.

## BACKGROUND

The VA has a national network of facilities that provides a comprehensive set of healthcare services, including inpatient and outpatient care, medications, and medical equipment, to more than 5 million U.S. veterans annually, approximately 20 percent of whom are racial/ethnic minorities [27]. All veterans eligible for VA care are offered the same set of services and pay no premiums, although some veterans are subject to copayments for medications for conditions not related to military service and some veterans with financial means surpassing a specified threshold also pay copayments for other services [28–29]. Given the large portion of patients from racial/ethnic minority groups who receive care at VA

facilities, the availability of national data on healthcare use and outcomes, and the limited financial barriers to care in the VA, the VA is a valuable setting for studying racial/ethnic disparities [25,27]. Although most financial barriers to healthcare found in the private sector have been removed for VA users, racial/ethnic disparities in healthcare utilization and outcomes have been found in the VA [13,15,20,27,30–36]. Other studies, however, have found no disparities in care or outcomes or have found that the disparities that do exist in the VA population are reduced in size compared with the disparities found in other non-VA populations [12,14–16,21,37–40]. VA's contribution to reducing health disparities through improved understanding of factors responsible for their absence or attenuation as well as the continued existence of racial/ethnic disparities in some areas of VA care highlight the need for continued research and monitoring. Accurate and complete patient race/ethnicity information is critical to these endeavors.

Veterans' racial/ethnic affiliation in VA data is entered into the local healthcare facility electronic medical record known as the Computerized Patient Record System by healthcare facility personnel and then transmitted with patient healthcare encounter data to the VA's centralized data repository at the Austin Information Technology Center, where it is stored in the National Patient Care Database (NPCD). Data extracts from the NPCD, known as the Medical SAS (MedSAS) data sets, are frequently used by researchers and contain race/ethnicity data as well as clinical and other demographic information [26,41].

In accordance with a 1997 revision of Office of Management and Budget Directive 15, which established standards for the classification of Federal data on race/ethnicity, and VHA Directive 2003-027, VA healthcare encounter records from 2004 forward contain race/ethnicity information that is self-reported or reported by a representative who is authorized to speak for the patient (i.e., patient proxy-reported) [42–43].

Although the VA did not record the method of collection prior to 2003 when VA implemented the new data collection standards, it is widely assumed to have been predominantly observer-reported by clinic personnel. The transition to self-report (or patient proxy-report) as the preferred method of collection was an outgrowth of the evolution in understanding race to be a social rather than biological construct; self-identity is the most accurate

and useful, and perhaps the only, valid measure of race/ethnicity [6,8,44–45].

Unfortunately, patient race/ethnicity information is frequently missing in VA healthcare encounter records [24–26]. A review of 114 studies focusing on racial/ethnic disparities in VA found that these studies reported missing race/ethnicity data rates as high as 48 percent [25,46]. Approaches for addressing these missing data have included creating a “missing” or “unknown” category, using patient race/ethnicity information obtained from other sources, or excluding patients with missing race/ethnicity values from the study [25]. Moreover, in more than 40 percent of the studies reviewed, the authors did not discuss the issue of or methods for addressing missing race/ethnicity, even when the information was from data sources in which race/ethnicity values were known to be missing [25]. Thus, previous examinations of racial/ethnic disparities in VA have failed to completely and consistently address the issue of missing data with unknown consequences for study results [25].

In this study, we examined the feasibility and utility of using non-VA data sources (Medicare and DOD) to address missing data problems in VA healthcare data. That is, we addressed two questions: (1) To what extent can missing patient race information be reduced using these sources? and (2) How likely is it that the information obtained from these sources will mirror the information that would have been available in VA data had it been obtained from the patient? Determining the agreement between self-reported VA race/ethnicity information and the information from external sources provides insight into the utility of using external sources to supplement incomplete race/ethnicity information in VA data. Because the vast majority of the 42 percent of elderly VA users and nearly 20 percent of younger users are enrolled in Medicare, results from this study could provide a method to address missing race/ethnicity values for a substantial portion of VA users. The utility of DOD data (more recently made available in the VA) for addressing missing data problems in VA has not previously been explored.

## METHODS

### Study Design

This was a retrospective cohort study of a representative 10 percent sample of individuals who received VA healthcare between October 1, 2003, and September 30,

2005 (fiscal years [FYs] 2004 and 2005). We identified patients in this cohort whose race in VA data was either missing or unknown (i.e., contained no “usable” value), and we determined the proportion for whom that information could be obtained from either Medicare or DOD data sources. For veterans whose VA records did contain a usable value, we determined the agreement between the VA values and those in Medicare and DOD data.

### Study Sample

Our representative 10 percent sample consisted of 574,971 individuals who received VA healthcare in FY 2004 and 2005. We excluded 1,590 (0.3%) individuals whose age, calculated from the date of birth in the VA record, was implausible—younger than 18 and older than 110 years. We also excluded 3,363 (0.6%) individuals who had two or more different race values in the FY 2004 and 2005 data (approximately half reported multiple racial identities and the other half reported a single but different race over time).

To examine agreement between the data sources, we conducted a record match between VA and Medicare data and VA and DOD data. To ensure that the records from the three data sources contained information on the same individuals, we used conservative matching criteria: Social Security number (SSN) plus date of birth or SSN plus sex plus two of the three parts of the date of birth (month, day, year).

### Data Sources

We obtained information on race from the VA MedSAS data sets for FY 2004 and 2005. These data sets are national workload data for VA-provided and VA-funded healthcare [42–43,47–48]. For this study, we used the outpatient “Visit” and inpatient “Acute Main” MedSAS data sets. Each record in the outpatient Visit file reflects the services provided to an individual at a VA facility on a single day. Information in these records, therefore, may be generated from multiple provider encounters (e.g., clinic visits) or services (e.g., radiology examination), all provided on the same day and at the same facility. The Acute Main file includes one record for each discharge from a VA acute care hospital stay in the respective FY. Race categories in VA data are white, black or African American, American Indian or Alaska Native, Asian, and Native Hawaiian or Other Pacific Islander (all referred to in this study as usable values). The data also contain “Declined to Answer,” “Unknown,” and missing values.

Guidance during the time of our study period [43] instructed personnel to enter “Unknown” if the veteran either returned the Application for Health Benefits (1010-EZ) form to enroll in VA healthcare services without completing the race/ethnicity sections or refused to answer the question when he or she checked in for an outpatient visit or inpatient admission. Therefore, the frequency of “Declined to Answer” values in the data probably does not accurately reflect the number who declined and the “Unknown” category has unclear meaning. The vast majority (about 93%) of unusable race values in this study were due to null values. VA collects and reports Hispanic ethnicity separately from race.

We obtained Medicare race/ethnicity information from the Medicare Vital Status file. This data set contains information about each beneficiary ever entitled to Medicare, including demographic information populated from the Medicare program’s enrollment database [49]. It is updated annually to reflect changes in enrollment and vital status. Medicare race/ethnicity information comes primarily from the Social Security Administration (SSA) and has known data quality problems of its own, including the absence of a separate indicator of Hispanic ethnicity and a substantial proportion of “Unknown” values [50–51]. Medicare race/ethnicity categories are white, black, Asian, Hispanic, North American Native, and Other.

We obtained DOD race information from the VA/DOD Identity Repository (VADIR) database, which is owned by the Office of Enterprise Development of the VA Office of Information. VADIR contains DOD data elements for all veterans whose military separation date was 1980 or later and for some veterans discharged before 1980 [52]. The VA obtains these data from the Defense Manpower Data Center’s (DMDC’s) Defense Enrollment Eligibility Reporting System (DEERS) database through an established VA/DOD data-sharing agree-

ment. Included in the DEERS database is self-reported race/ethnicity obtained from servicemembers when they first join the military as part of their military entrance processing. Active Duty servicemembers can change this information at any time at the service personnel office or online. Current rules allow servicemembers to decline to provide their race/ethnicity. In this article, we refer to race/ethnicity data obtained from VADIR as DOD data. Race values in the DOD data are white, Asian or Pacific Islander, black, American Indian or Alaskan Native, Other, and Unknown. DOD collects and reports Hispanic ethnicity separately; that information was not used in this study.

In **Table 1**, we have presented the race categories in the VA, DOD, and Medicare databases. Differences were present in categories of non-African-American minority groups across the three data sources. To facilitate comparison across databases, we combined some categories to create four mutually exclusive categories: white; black or African American; North American Native; and Asian, Pacific Islander, or Other (APIO). For the 1 percent of individuals in the Medicare data whose race/ethnicity was classified as Hispanic, we compared this with Hispanic or Latino ethnicity in VA MedSAS data sets.

### Analysis

We identified two groups of patients: those with and without a usable race value in the VA MedSAS data sets. For the patients without a usable value (the “missing” subsample), we linked VA with Medicare records and identified those whose race information was present in those records. We then calculated the proportion of the missing subsample whose race information was present after the record linkage. We also calculated the improvement in data completeness in our full VA sample of 570,018 individuals that resulted from the record linkage.

**Table 1.**

Race classification mapping across Department of Veterans Affairs (VA), Medicare, and Department of Defense (DOD) data.

VA	DOD	Medicare	Classification Constructed for Consistency Analysis
White	White	White	White
Black or African American	Black	Black	Black or African American
American Indian or Alaska Native	American Indian or Alaska Native	North American Native	North American Native
Asian	Asian or Pacific Islander	Asian	Asian, Pacific Islander, or Other
Native Hawaiian or Other Pacific Islander	Other	Other	Asian, Pacific Islander, or Other

We computed these proportions among patients in two age groups: elderly ( $\geq 65$ ) and nonelderly ( $< 65$ ).

We conducted a similar analysis using data originating from DOD but focused on individuals  $< 65$  years. Since the VADIR database includes DOD data for only a limited number of veterans who separated from the military before 1980, very few individuals in our elderly sample would have DOD data in their VADIR record. In addition, because only approximately 20 percent of nonelderly veterans are enrolled in the Medicare program, Medicare data will have less value in improving race data completeness among individuals  $< 65$  years than among elderly individuals. Therefore, DOD data have the greatest potential to add value for the younger population.

In the final step of our missing race analysis, we linked records from all three sources—VA, Medicare, and DOD—and calculated the proportion of the full sample with a usable race value after all data sources were combined.

For the individuals in our sample who had a usable race value in VA data, we compared Medicare and DOD race values to VA values in a linked record data set (the “race consistency” subsample). We calculated sensitivities, specificities, positive predictive values (PPVs), negative predictive values, and kappa statistics to evaluate race category agreement between VA and Medicare and VA and DOD data. Again, we limited the DOD analysis to individuals  $< 65$  years.

To explore whether Medicare data might provide a useful source to supplement missing ethnicity information in the VA MedSAS data sets, we compared concordance of the patients reporting Hispanic or Latino ethnicity in VA data with the Hispanic ethnicity category in Medicare data.

## RESULTS

Our full study sample comprised 570,018 individuals, 295,010 (52%) of whom had no usable race information in the MedSAS data sets and therefore were the focus of our completeness analysis (**Figure 1**, **Table 2**). The remaining 48 percent of the full sample (275,008 individuals) was the focus of the consistency analysis.

**Table 2** presents sample characteristics for those with and without a usable race value in VA data. Due to the large sample size, these groups differed statistically on all sample characteristics, but very few of the differ-

ences could be considered meaningful in any practical sense. The largest differences were found for sex, geographic region, and period of military service. Individuals lacking a usable race value in VA data were less likely to be male, reside in the South, or have served during the Vietnam era and were more likely to reside in the West than those with a usable value.

### Race Completeness Analysis

**Figure 2** shows the results of linking Medicare data to fill in missing values among the subgroup whose race was missing in VA data. Results for the elderly and nonelderly age groups are broken out in **Figures 2(b)** and **2(c)**. Of the 295,010 individuals in the missing race subsample, 157,189 (53%) had a Medicare record. As expected due to Medicare eligibility criteria, the Medicare record match rate was much higher among elderly than nonelderly individuals (97% vs 18%). The Medicare record contained a usable race value 99 percent of the time overall, in 99 percent of individuals  $\geq 65$  years and in 98 percent of individuals  $< 65$  years.

In **Figure 3**, we show the influence of combining VA and Medicare data on our full study sample of VA patients. Adding Medicare data improved race data completeness in the full sample from 48 to 76 percent. In the older age group, adding Medicare data improved completeness from 47 to 98 percent while completeness in the younger age group improved from 49 to 58 percent.

**Figure 4** shows the results of linking DOD data with VA data to fill in missing race among nonelderly veterans in VA data. Of the 162,882 individuals  $< 65$  years in our missing race subsample, 134,892 (83%) had a VADIR record (**Figure 4**). The VADIR record contained a usable DOD race value in 45 percent of those cases.

In **Figure 5**, we show the influence of combining VA and DOD data on the subgroup of nonelderly individuals without a usable race value. Adding DOD data improved data completeness in this group from 49 to 68 percent.

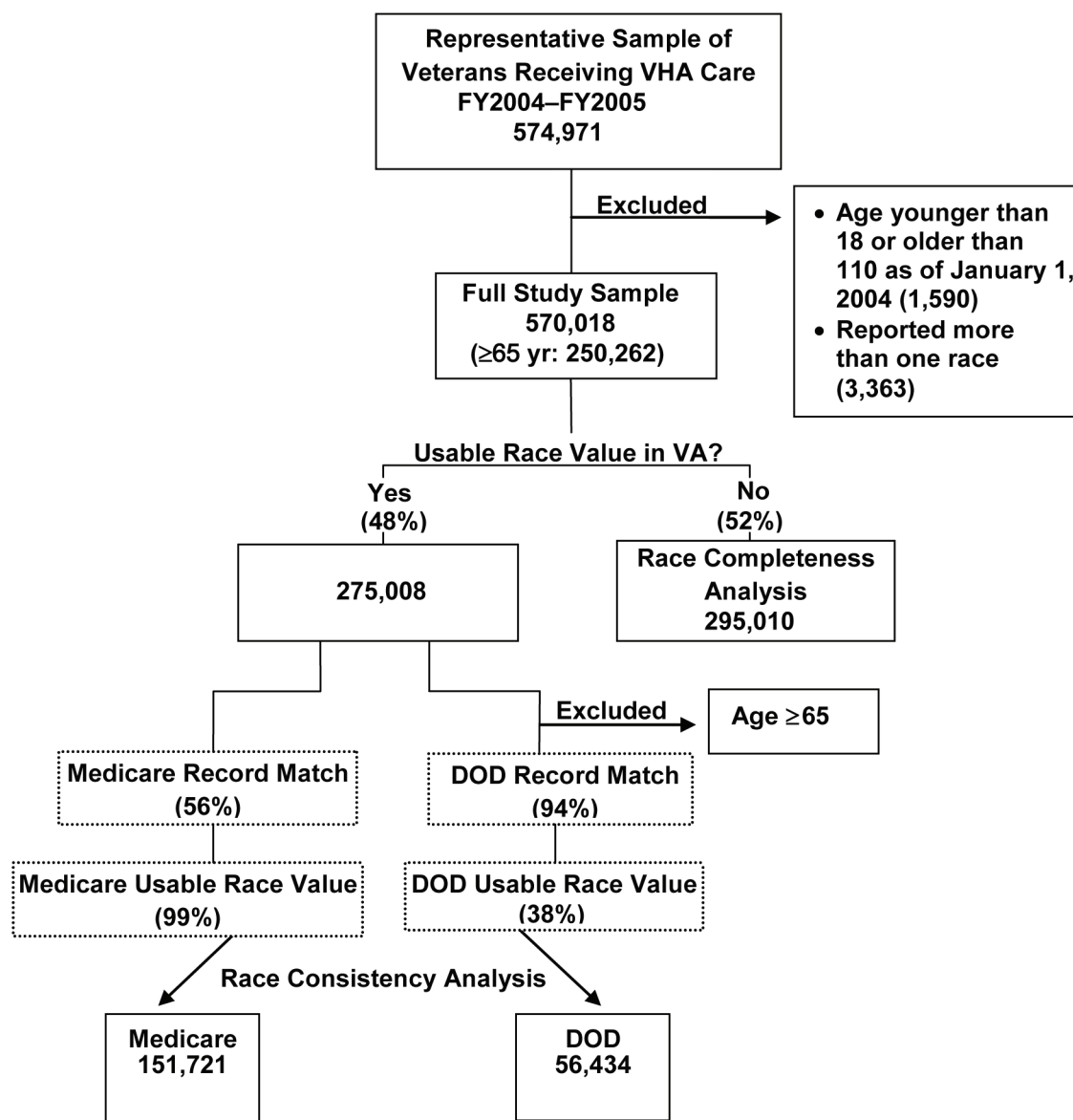
Finally, we combined both Medicare and DOD data to examine the benefit gained from using all three data sources together to fill in missing race among individuals  $< 65$  years. Race data completeness improved from 49 to 76 percent in that group (**Figure 6**). Among the nonelderly, combining Medicare and DOD data improved completeness 8 percentage points over that achieved by adding DOD data alone and 18 percentage points over that achieved by adding Medicare data alone.

### Race Consistency Analysis

**Figure 7** shows the concordance between VA and Medicare (**Figure 7(a)**) and VA and DOD data (**Figure 7(b)**). A high degree of concordance was found between VA and Medicare data for individuals identified as white (99%) or African American (96%) in VA data. Among individuals who were North American Native in VA data, only 36 percent were recorded as such in Medicare data, and among those who were APIO in VA, just 47 percent

were APIO in Medicare data. The majority who had discordant race information in the VA North American Native and APIO groups were recorded as white in Medicare data (55% and 47% of those groups, respectively).

We also found a high degree of concordance between VA and DOD data for individuals identified as white (93%) or African American (95%) in VA data. Among individuals who were North American Native in VA data, only 39 percent were recorded as such in DOD data,



**Figure 1.**

Sample derivation. DOD = Department of Defense, FY = fiscal year, VA = Department of Veterans Affairs, VHA = Veterans Health Administration.

**Table 2.**  
Sample characteristics.\*

Characteristic	Had Usable Race in VA Data <sup>†</sup>			
	Yes ( <i>n</i> = 275,008)		No ( <i>n</i> = 295,010)	
	<i>n</i>	%	<i>n</i>	%
Age <sup>‡</sup> : ≥65 years	118,134	43.0	132,128	44.8
Sex: Male	258,702	94.1	261,278	88.6
Marital Status: Married	153,211	55.7	165,687	56.2
Geographic Region <sup>§</sup>				
Northeast	45,088	16.4	48,562	16.5
South	121,735	44.3	109,929	37.3
Midwest	61,957	22.5	62,698	21.2
West	46,228	16.8	73,821	25.0
Period of Military Service				
Post-Vietnam and Desert Storm	52,070	18.9	47,444	16.1
Vietnam	100,863	36.7	85,008	28.8
Korea (including pre and post)	62,895	22.9	63,549	21.5
World War II	50,414	18.3	57,500	19.5
Other <sup>¶</sup>	8,766	3.2	41,509	14.1

\*Unique combination of Social Security number, date of birth, and sex defines individual.

<sup>†</sup>Chi-square tests show statistically significant differences across groups for all characteristics at  $p < 0.001$ .

<sup>‡</sup>Age on January 1, 2004.

<sup>§</sup>Geographic region is based on VA network in which patient lives: Northeast—VISNs 1 to 4; South—VISNs 5 to 9, 16, and 17; Midwest—VISNs 10 to 12, 15, 23; West—VISNs 18 to 22.

<sup>¶</sup>Other includes World War I, Spanish-American War, and other.

VA = Department of Veterans Affairs, VISN = Veterans Integrated Service Network.

while among those who were APIO in VA, 65 percent were APIO in DOD data. The majority who had discordant race information in the VA North American Native and APIO groups were recorded as white in DOD data (46% and 27% of those groups, respectively).

Compared with Medicare data, DOD data had poorer concordance for the VA white group (93% vs 99%) and similar concordance for the African-American group (95% vs 96%). In contrast, concordance between DOD and VA data was better than the concordance between Medicare and VA data for the North American Native group (39% vs 36%) and markedly better for the APIO group (65% vs 47%).

Measures of agreement between VA data and each of the external data sources are shown in **Table 3** (Medicare data) and **Table 4** (DOD data), which assume the VA self-reported or proxy values to be the gold standard. In both data sources, sensitivities and PPVs for the white and African-American categories were high. In Medicare, the PPVs were 98.6 for the white and 94.7 for the African-American categories. In DOD, PPVs were 97.0 for the white and 96.5 for the African-American categories. Kappa statistics ranging from 0.86 (for the white

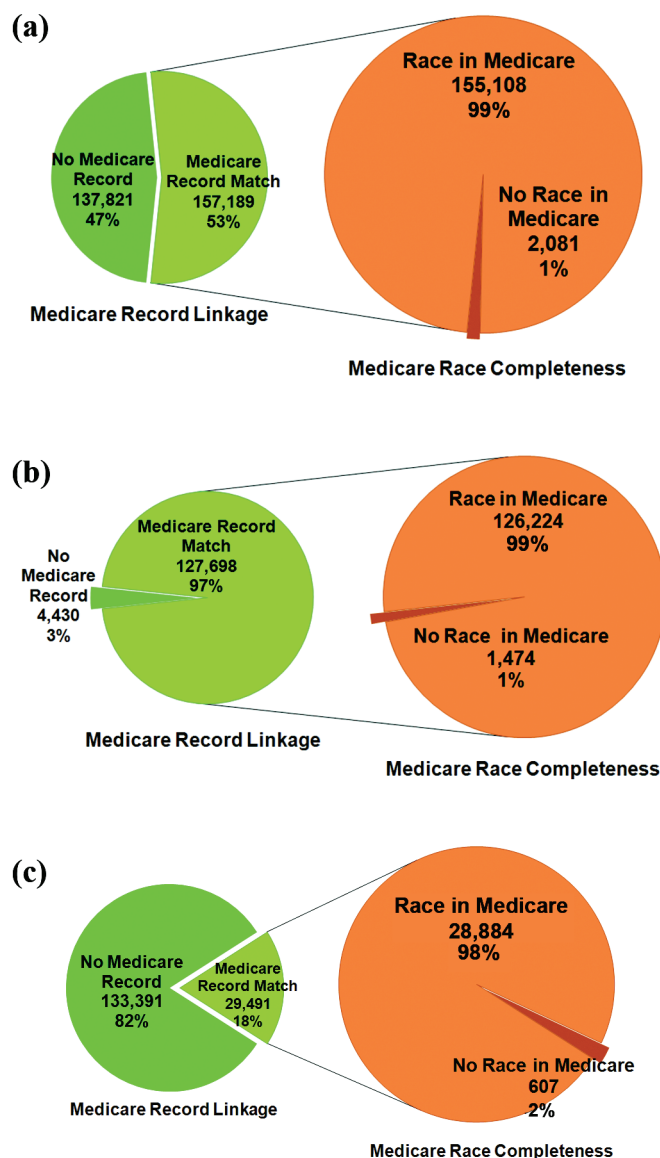
category in DOD data) to 0.95 (for the African-American category in Medicare data) indicate high levels of agreement and reflect the high specificities and sensitivities shown.

In both Medicare and DOD data, sensitivities and PPVs for the North American Native and APIO groups were much lower. In Medicare, the PPVs were 38.0 for the North American Native and 48.2 for the APIO categories. In DOD, PPVs were 35.3 for the North American Native and 30.5 for the APIO categories. Kappa statistics ranging from 0.37 (for the North American Native category in both data sources) to 0.47 (for the APIO category in DOD data) indicate only fair agreement.

### Hispanic Ethnicity Analysis

Of the 5,606 (3.7%) veterans in our sample who reported Hispanic or Latino ethnicity in the VA MedSAS data sets and were also in the Medicare data, only 25 percent were recorded as Hispanic in the Medicare data (**Figure 8**). The majority of patients reporting Hispanic ethnicity were recorded as white (64%) in the Medicare data.

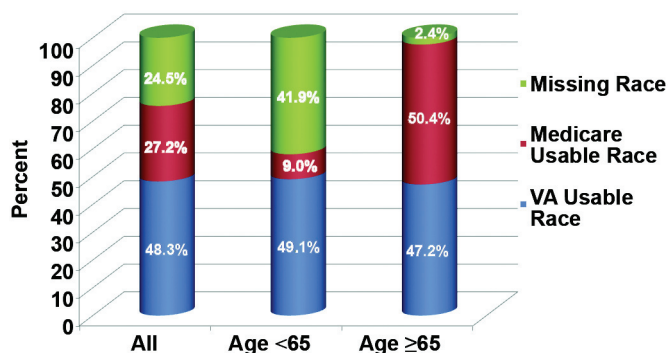


**Figure 2.**

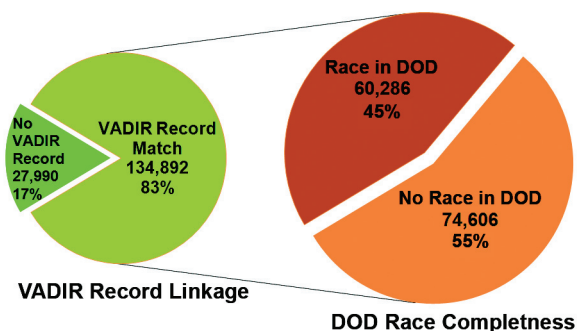
Improving Department of Veterans Affairs (VA) race data completeness using external sources: Medicare. (a) All individuals without VA usable race. (b) Elderly ( $\geq 65$  years) without VA usable race. (c) Nonelderly ( $<65$  years) without VA usable race.

## DISCUSSION

In this report, we evaluated the improvement in VA race data completeness that could be achieved by linking VA data with data from Medicare and the DOD. Medicare merged with VA data substantially improved race data completeness; the proportion of the full sample with

**Figure 3.**

Adding Medicare data improves race data completeness. Sample sizes: All = 570,018; Age  $<65$  = 319,756; Age  $\geq 65$  = 250,262. VA = Department of Veterans Affairs.

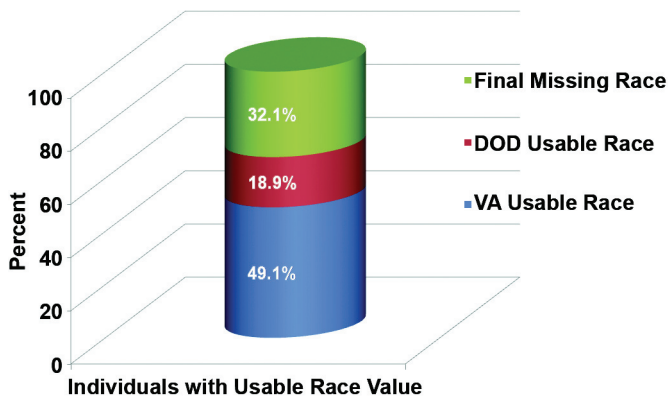
**Figure 4.**

Improving Department of Veterans Affairs (VA) race data completeness using external sources: Department of Defense (DOD). VADIR = VA/DOD Identity Repository.

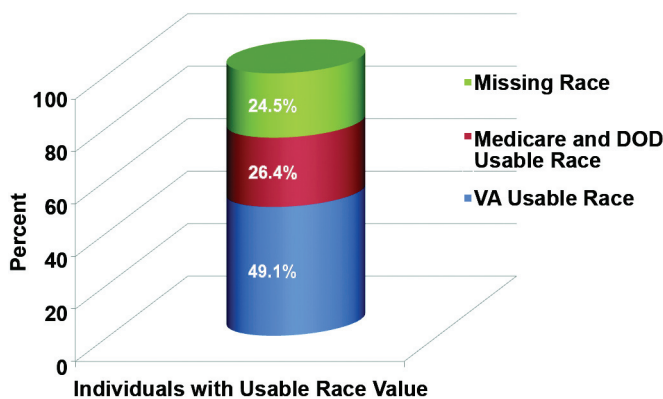
a usable value increased by 56 percent, resulting in 98 percent completeness among individuals  $\geq 65$  years. Medicare data also improved completeness among the 18 percent of the younger age group who were Medicare enrollees. More modest improvements were realized with DOD data; race completeness increased by 38 percent (from 49% to 68%) among those  $<65$  years. The greatest improvement in data completeness in the younger age group was achieved when both Medicare and DOD data were used to supplement VA data. In a merged data set that included VA and the two external data sources together, more than 75 percent of individuals  $<65$  years had a usable race value.

We also examined the agreement in race values between VA and the two external data sources and found



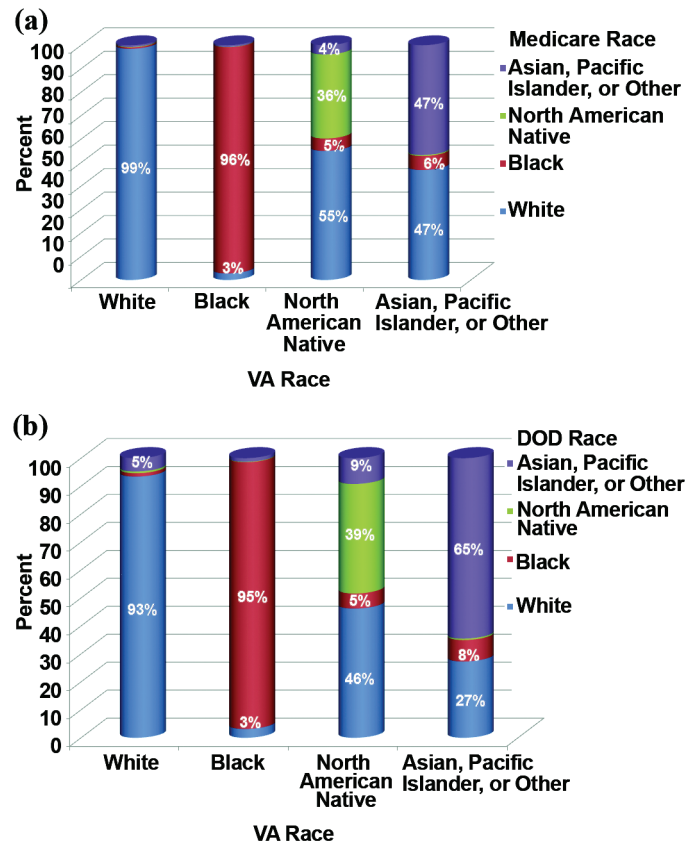


**Figure 5.** Adding Department of Defense (DOD) data improves race data completeness among nonelderly (<65 years) ( $n = 319,756$ ). VA = Department of Veterans Affairs.



**Figure 6.** Adding Medicare and Department of Defense (DOD) data improves race data completeness among nonelderly (<65 years) ( $n = 319,756$ ). VA = Department of Veterans Affairs.

high levels of agreement between VA and each source for self-reported white and African-American individuals. Agreement for self-reported North American Native and APIO individuals was fair; a large portion of these individuals was recorded as white in Medicare and DOD data. These results suggest that researchers who use Medicare or DOD race data to supplement VA data will under-identify the non-African-American minority groups and that a substantial proportion of those groups will be misclassified as white. Since together those groups represent less than 2 percent of the sample in the VA-Medicare merged data and just 3 percent of the sam-



**Figure 7.** Concordance between Department of Veterans Affairs (VA) race values and race/ethnicity values from external sources. **(a)** Medicare ( $n = 151,721$ ). **(b)** Department of Defense (DOD) ( $n = 56,434$ ). Data values not shown for percentages less than 2.5.

ple in the VA-DOD merged data, the likely effect on research study results is small, except in cases in which the study sample is small.

Our finding of much lower sensitivity and PPV of Medicare race data for the North American Native and APIO classifications than for the white and African-American classifications is consistent with results of other studies examining the validity of Medicare data on race (we are unaware of other studies examining DOD race data quality) [51,53–55]. However, these other studies found much higher PPVs for the North American Native classification than we found. For example, Waldo compared Medicare data to self-reported race in the Medicare Current Beneficiary Data and found a PPV of 69.5 for the Medicare North American Native classification [55], while we found a PPV of 38.0 in our study. This low PPV is owing to the large proportion of individuals who

**Table 3.**

Accuracy of race data from Medicare compared with VA.

Race	VA	Medicare		Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Kappa
		Yes	No					
White	Yes	129,305	2,007	98.5	91.3	98.6	90.3	0.89
	No	1,772	18,637	(98.4–98.5)	(90.0–91.7)	(98.6–98.7)	(89.9–90.7)	
Black	Yes	17,142	638	96.4	99.3	94.7	99.5	0.95
	No	966	132,975	(96.1–96.7)	(99.2–99.3)	(94.3–95.0)	(99.5–99.6)	
North American Native	Yes	183	328	35.8	99.8	38.0	99.8	0.37
	No	299	150,911	(31.6–40.1)	(99.8–99.8)	(33.6–42.5)	(99.8–99.8)	
Asian, Pacific Islander, or Other	Yes	991	1,127	46.8	99.3	48.2	99.2	0.47
	No	1,063	148,541	(44.6–48.9)	(99.2–99.3)	(46.1–50.4)	(99.2–99.3)	

CI = confidence interval, NPV = negative predictive value, PPV = positive predictive value, VA = Department of Veterans Affairs.

**Table 4.**

Accuracy of race data from DOD compared with VA.

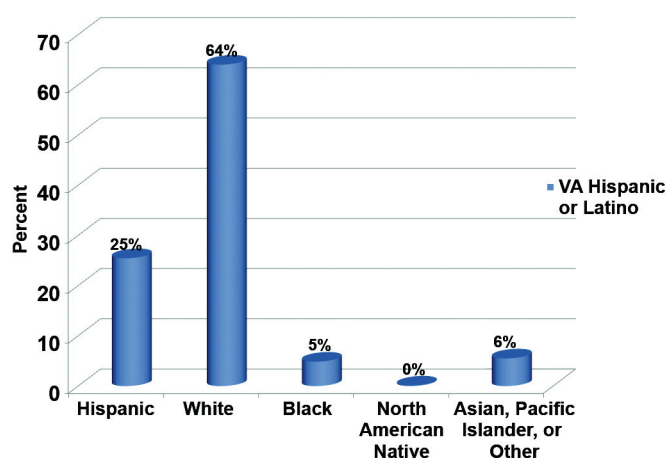
Race	VA	DOD		Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Kappa
		Yes	No					
White	Yes	35,187	2,472	93.4	94.2	97.0	87.7	0.86
	No	1,085	17,690	(98.4–98.5)	(93.2–93.7)	(96.8–97.2)	(87.3–88.2)	
Black	Yes	16,247	777	95.4	98.5	96.5	98.0	0.95
	No	583	38,827	(95.1–95.7)	(98.4–98.6)	(96.2–96.8)	(97.9–98.2)	
North American Native	Yes	145	225	39.2	99.5	35.3	99.6	0.37
	No	266	55,798	(34.2–44.4)	(99.5–99.6)	(30.7–40.1)	(99.6–99.6)	
Asian, Pacific Islander, or Other	Yes	891	490	64.5	96.3	30.5	99.1	0.39
	No	2,030	53,023	(61.9–67.0)	(96.2–96.5)	(28.8–32.2)	(99.0–99.2)	

CI = confidence interval, DOD = Department of Defense, NPV = negative predictive value, PPV = positive predictive value, VA = Department of Veterans Affairs.

were identified as North American Native in Medicare data but some other race in VA data. The large majority of these “false positives” were classified as white in VA data. In fact, 58 percent of the 482 individuals identified as North American Native in Medicare were recorded as white in VA. The DOD data had very similar proportions and distributions of false positives for North American Natives. Additionally, a VA study comparing self-reported race from a survey in the VA’s electronic health record found more than 85 percent concordance for whites and African Americans but only 20 percent concordance for North American Natives [56].

Reasons for this pattern of discordance in the North American Native and APIO groups are unclear. In this analysis, we have treated VA data as the gold standard against which Medicare and DOD data were compared. VA policy dictates that race/ethnicity be obtained from the patient or proxy. However, the data entry system does not prevent the entry of values that are not self-reported

(for example, values based on clinic staff observation), and we have no way of verifying the true source of the information. In a recent study comparing VA self-reported race to observer-reported race in earlier VA data (prior to the 2003 mandated switch to the self-reported data collection standard), the investigators found that 58 percent of self-reported North American Natives were identified as white in observer-reported data [26]. Medicare has made special efforts to improve the quality of its race data and since 1999 has been using data provided by the Indian Health Service to identify enrollees who are North American Natives. These efforts have increased the identification of North American Natives by an estimated 68 percent [57]. We cannot rule out, then, that the misclassification is occurring in VA rather than Medicare data. Furthermore, since race information in VA, Medicare, and DOD were collected at different points in an individual’s lifetime and race is a social rather than a biological construct, we also cannot rule out that individuals’



**Figure 8.** Medicare race/ethnicity among Department of Veterans Affairs (VA) self-reported Hispanics.

racial identifications may have changed over time. While more than 93 percent of whites and blacks marry within their own racial group, 70 percent of Asians and 33 percent of American Indians do so [58]. Consequently, a substantial portion of individuals in the APIO and North American Native groups are likely to be multiracial. Among multiracial individuals, self-identity has been found to change over time, with North American Natives having the most instability in their racial identity [59]. So, it is also possible that a majority of individuals of North American Native heritage are reporting white as their race in VA.

We also found some differences between the two external data sources in their concordance with VA data. For example, sensitivity of Medicare data for the VA white group was 98 percent but only 93 percent in DOD data. In contrast, sensitivity of DOD data for the VA Other group (comprising Asian, Native Hawaiian, Pacific Islander, and Other) was 65 percent but only 47 percent in Medicare data. While race in both external data sources is principally self- or proxy-reported (in the case of SSA-originated Medicare data, parents may have applied for a SSN on behalf of a child), there could be several reasons for these differences in concordance. In our study and consistent with others' findings, the greatest discordance between VA and each of the external data sources was observed for the non-African-American minority groups and most of the misclassification involved identification as white rather than the minority group [3,26]. The proportion of non-African-American

minorities in our VA/DOD analysis sample (3.1%) was nearly twice that in the VA/Medicare analysis sample (1.7%). Therefore, we would expect poorer concordance overall in the VA/DOD than in the VA/Medicare analysis. Additionally, some of the discordance could be related to shifts over time in likelihood of self-identifying as belonging to a minority group (on average, individuals in the VA/DOD analysis are 28 years younger than those in the VA/Medicare analysis) and/or to different preferences for revealing racial affiliation in the various settings (VA/Medicare, DOD) [44,60]. Finally, while DOD race information is purportedly self-reported and can be updated at any time, we were unable to find an organizational directive, a manual, or instructions that operationalize this.

As the largest integrated healthcare system in the United States, with a large minority population who face minimal financial barriers to access to care relative to the private sector, the VA has served as the setting for a substantial number of investigations into racial/ethnic disparities in healthcare utilization and quality of care. Researchers have often used the VA's electronic health information system as the source of race/ethnicity information but have faced the perennial problem of incomplete values in these databases, the prevalence of which has been reported to be as high as 48 percent in previous research [25]. We have shown that the use of Medicare data to supplement VA data will reduce the missing data quite substantially, to approximately 25 percent in a representative sample and to close to zero for the 53 percent of patients (97% of those  $\geq 65$  years; 18% of those  $< 65$  years) who were enrolled in Medicare. We have also shown a high level of agreement between VA and Medicare data for the white and African-American categories, which suggests that the information obtained from Medicare to supplement VA data will mirror the information that would have been available in VA data had it been obtained from the patient. This information is particularly important for researchers because Medicare race/ethnicity data is now available in the VA Vital Status file. Our results show that researchers can use this information to fill in missing white or African-American race with confidence; however, our results also show that researchers should use caution with race information for non-African-American minorities.

For veterans not enrolled in Medicare, this study is also the first to show the utility of supplementing race information with DOD data for individuals  $< 65$  years. Unfortunately, DOD DEERS data are not available for

veterans discharged before the 1980s. Therefore, DOD data will not be a useful source of race/ethnicity information for older veterans at this time. The improvement in race data completeness was more modest with DOD data than with Medicare data. However, if race is a particular focus of research in this population, DOD data is a source that researchers could consider. Moreover, a high level of agreement was found between white and African-American categories in VA and DOD data for individuals with a usable race value in both data sources, which supports the usefulness of these data to fill in incomplete VA data for these categories.

For researchers needing race information, we would recommend that they supplement incomplete information with Medicare data from the VA Vital Status file. Because such a high proportion of VA non-African-American minorities are recorded as white in Medicare data, the most reliable classification when supplementing VA with Medicare data is a dichotomous grouping of African American versus not African American. For researchers needing race information for a younger cohort, supplementing VA and Medicare data with DOD data and again combining the non-African-American individuals into a single category may be a consideration. Researchers focusing on non-African-American minorities might consider other sources such as Indian Health Service data or conducting a survey.

Our study has some limitations. In order to have comparability in categories across the three data sets, we combined the Asian, Native Hawaiian, Other Pacific Islander, and "Other" classifications into one category. Other published literature examining Medicare race data has shown much higher sensitivity and PPV for the Asian than the "Other" category [53–55]. By combining these, we likely underestimated the agreement between Medicare—and possibly DOD—and VA data for individuals identified as Asian in VA data. We were unable to assess the utility of DOD race data for supplementing VA race information for elderly VA patients because of VADIR data limitations. We were unable to match VA with VADIR records in 17 percent of the subgroup we tried to match, those <65 years. This 17 percent (9,032) matched on SSN but not other match criteria (date of birth, sex). None of those individuals had a race value in their VADIR record. Therefore, inclusion of these individuals would not have affected either our completeness or consistency analysis.

Questions remain about best approaches to addressing problems presented by missing race information in VA data. Future studies should explore the potential contribution of Indian Health Service data as well as the additional benefit derived from DOD data obtained directly from the DMDC. Further exploration of VADIR data completeness would be highly desirable if it is to be used in future VA research.

## CONCLUSIONS

Using Medicare data to fill in missing race information in VA records improves data completeness substantially. Among veterans <65 years, the benefit derived from supplementation with DOD data was substantial and use of the two data sources together improves completeness by 18 percentage points beyond that achieved with Medicare data alone. Medicare and DOD had similar rates of agreement with VA data. Use of either of these two external data sources will result in high rates of accurate classification of patients who are either African American or white. More study is needed to understand poor rates of agreement between VA and external sources in identifying race for individuals who are neither white nor African American. The best approach to managing the problem of missing race/ethnicity information may vary from study to study. This study has demonstrated that a potentially useful approach is to supplement VA data with Medicare and DOD data.

## ACKNOWLEDGMENTS

### Author Contributions:

*Study concept and design:* K. T. Stroupe, E. Tarlov, D. M. Hynes.

*Data acquisition, management, and programming:* T. Haywood, Q. Zhang.

*Analysis and interpretation of data:* T. Haywood, Q. Zhang, K. T. Stroupe, E. Tarlov.

*Drafting of manuscript:* K. T. Stroupe, E. Tarlov, A. Owens.

*Critical revision of manuscript for important intellectual content:* T. Haywood, Q. Zhang, K. T. Stroupe, E. Tarlov, D. M. Hynes, A. Owens.

*Obtained funding:* D. M. Hynes.

*Administrative, technical, or material support:* A. Owens.

*Study supervision:* K. T. Stroupe, E. Tarlov, D. M. Hynes.

**Financial Disclosures:** The authors have declared that no competing interests exist.

**Funding/Support:** This material was based on work supported by the VA Information Resource Center, VA Health Services Research &

Development Service (grant SD4 98-004) and the VA/CMS Data for Research Project (SDR 02-237). Dr. Hynes was also supported by a VA Research Career Scientist Award and the VA Health Services Research and Development Service (grant IIR 03-196).

**Additional Contributions:** The views expressed in this article are those of the authors and do not necessarily represent the position of the VA. Dr. Stroupe is now with the Program in Health Services Research, Stritch School of Medicine, Loyola University Chicago, Maywood, Illinois.

## REFERENCES

1. Smedley BD, Stith AY, Nelson AR; Institute of Medicine (U.S.). Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care. Unequal treatment: Confronting racial and ethnic disparities in health care. Washington (DC): National Academies Press; 2003.
2. National healthcare disparities report, 2005 [Internet]. AHRQ Publication No. 06-0017. Rockville (MD): U.S. Department of Health and Human Services, Agency for Healthcare Research and Quality; 2005. Report No.: 06-0017. Available from: <http://www.ahrq.gov/qual/nhdr05/fullreport>.
3. Kressin NR, Petersen LA. Racial differences in the use of invasive cardiovascular procedures: Review of the literature and prescription for future research. *Ann Intern Med*. 2001;135(5):352–66. [PMID: 11529699]
4. Ayanian JZ. Determinants of racial and ethnic disparities in surgical care. *World J Surg*. 2008;32(4):509–15. [PMID: 18196327]  
DOI:10.1007/s00268-007-9344-4
5. Healthy People 2010. Understanding and improving health, 2nd edition [Internet]. Washington (DC): U.S. Department of Health and Human Services; 2000. Available from: <http://www.healthypeople.gov/Document/tableofcontents.htm#under>.
6. Ver Ploeg M, Perrin E; National Research Council (U.S.). Panel on DHHS Collection and Race and Ethnicity Data. Eliminating health disparities: Measurement and data needs. Washington (DC): National Academies Press; 2004.
7. Ulmer C, McFadden B, Nerenz DR; Institute of Medicine (U.S.). Subcommittee on Standardized Collection of Race/Ethnicity Data for Healthcare Quality Improvement. Race, ethnicity, and language data: Standardization for health care quality improvement. Washington (DC): National Academies Press; 2009.
8. Ford ME, Kelly PA. Conceptualizing and categorizing race and ethnicity in health services research. *Health Serv Res*. 2005;40(5 Pt 2):1658–75. [PMID: 16179001]  
DOI:10.1111/j.1475-6773.2005.00449.x
9. Research on health disparities and minority health [Internet]. Washington (DC): Veterans Health Administration Research & Development; 2009 [updated 2009 Nov 3; cited 2009 Aug 10]. Available from: <http://www.research.va.gov/about/research-health-disparities.cfm>.
10. Agoston I, Cameron CS, Yao D, Dela Rosa A, Mann DL, Deswal A. Comparison of outcomes of white versus black patients hospitalized with heart failure and preserved ejection fraction. *Am J Cardiol*. 2004;94(8):1003–7. [PMID: 15476612]  
DOI:10.1016/j.amjcard.2004.06.054
11. Alexander D, Chatla C, Funkhouser E, Meleth S, Grizzle WE, Manne U. Postsurgical disparity in survival between African Americans and Caucasians with colonic adenocarcinoma. *Cancer*. 2004;101(1):66–76. [PMID: 15221990]  
DOI:10.1002/cncr.20337
12. Aujesky D, Long JA, Fine MJ, Ibrahim SA. African American race was associated with an increased risk of complications following venous thromboembolism. *J Clin Epidemiol*. 2007;60(4):410–16. [PMID: 17346616]  
DOI:10.1016/j.jclinepi.2006.06.023
13. Chakkeri HA, O'Hare AM, Johansen KL, Hynes D, Stroupe K, Colin PM, Chertow GM. Influence of race on kidney transplant outcomes within and outside the Department of Veterans Affairs. *J Am Soc Nephrol*. 2005;16(1):269–77. [PMID: 15563568]  
DOI:10.1681/ASN.2004040333
14. Deswal A, Petersen NJ, Soucek J, Ashton CM, Wray NP. Impact of race on health care utilization and outcomes in veterans with congestive heart failure. *J Am Coll Cardiol*. 2004;43(5):778–84. [PMID: 14998616]  
DOI:10.1016/j.jacc.2003.10.033
15. Dominitz JA, Maynard C, Billingsley KG, Boyko EJ. Race, treatment, and survival of veterans with cancer of the distal esophagus and gastric cardia. *Med Care*. 2002;40(1 Suppl):114–26. [PMID: 11789626]  
DOI:10.1097/00005650-200201001-00003
16. 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]  
DOI:10.1161/01.STR.0000063364.88309.27
17. Gordon HS, Paterniti DA, Wray NP. Race and patient refusal of invasive cardiac procedures. *J Gen Intern Med*. 2004;19(9):962–66. [PMID: 15333061]  
DOI:10.1111/j.1525-1497.2004.30131.x
18. Kamalesh M, Shen J, Tierney WM. Stroke mortality and race: Does access to care influence outcomes? *Am J Med Sci*. 2007;333(6):327–32. [PMID: 17570984]  
DOI:10.1097/MAJ.0b013e318065c101
19. Oddone EZ, Horner RD, Johnston DC, Stechuchak K, McIntyre L, Ward A, Alley LG, Whittle J, Kroupa L, Taylor J. Carotid endarterectomy and race: Do clinical indications and patient preferences account for differences?

- Stroke. 2002;33(12):2936–43. [PMID: 12468794]  
DOI:10.1161/01.STR.0000043672.42831.EB
20. Petersen LA, Wright SM, Peterson ED, Daley J. Impact of race on cardiac care and outcomes in veterans with acute myocardial infarction. *Med Care*. 2002;40(1 Suppl):186–96. [PMID: 11789635]  
DOI:10.1097/00005650-200201001-00010
  21. Rosenheck R, Fontana A. Black and Hispanic veterans in intensive VA treatment programs for posttraumatic stress disorder. *Med Care*. 2002;40(1 Suppl):I52–61. [PMID: 11789632]  
DOI:10.1097/00005650-200201001-00007
  22. Volpp KG, Stone R, Lave JR, Jha AK, Pauly M, Klusaritz H, Chen H, Cen L, Brucker N, Polsky D. Is thirty-day hospital mortality really lower for black veterans compared with white veterans? *Health Serv Res*. 2007;42(4):1613–31. [PMID: 17610440]  
DOI:10.1111/j.1475-6773.2006.00688.x
  23. FY2003 & FY2004 Medical SAS datasets missing race & ethnicity data. VIREC Data Issues Brief. Washington (DC): Department of Veterans Affairs; 2004.
  24. Jia H, Zheng YE, Cowper DC, Stansbury JP, Wu SS, Vogel WB, Duncan PW, Reker DM. Race/ethnicity: Who is counting what? *J Rehabil Res Dev*. 2006;43(4):475–84. [PMID: 17123187]  
DOI:10.1682/JRRD.2005.05.0086
  25. Long JA, Bamba MI, Ling B, Shea JA. Missing race/ethnicity data in Veterans Health Administration based disparities research: A systematic review. *J Health Care Poor Underserved*. 2006;17(1):128–40. [PMID: 16520522]  
DOI:10.1353/hpu.2006.0029
  26. Sohn MW, Zhang H, Arnold N, Stroupe K, Taylor BC, Wilt TJ, Hynes DM. Transition to the new race/ethnicity data collection standards in the Department of Veterans Affairs. *Popul Health Metr*. 2006;4:7. [PMID: 16824220]  
DOI:10.1186/1478-7954-4-7
  27. Saha S, Freeman M, Toure J, Tippens K, Weeks C, Ibrahim S. Racial and ethnic disparities in the VA health care system: A systematic review. *J Gen Intern Med*. 2008;23(5):654–71. [PMID: 18301951]  
DOI:10.1007/s11606-008-0521-4
  28. VHA Directive 2007-012, Eligibility verification process for VA health care benefits. Washington (DC): Department of Veterans Affairs; 2007.
  29. VHA Directive 1909, Income verification (IV) program. Washington (DC): Department of Veterans Affairs; 2008.
  30. Ambriz EH, Woodard LD, Kressin NR, Petersen LA. Use of smoking cessation interventions and aspirin for secondary prevention: Are there racial disparities? *Am J Med Qual*. 2004;19(4):166–71. [PMID: 15368781]  
DOI:10.1177/106286060401900405
  31. Cheng EM, Siderowf AD, Swartztrauber K, Lee M, Vassar S, Jacob E, Eisa MS, Vickrey BG. Disparities of care in veterans with Parkinson's disease. *Parkinsonism Relat Disord*. 2008;14(1):8–14. [PMID: 17702625]  
DOI:10.1016/j.parkreldis.2007.05.001
  32. Collins TC, Clark JA, Petersen LA, Kressin NR. Racial differences in how patients perceive physician communication regarding cardiac testing. *Med Care*. 2002;40(1 Suppl):I27–34. [PMID: 11789628]  
DOI:10.1097/00005650-200201001-00004
  33. Copeland LA, Zeber JE, Valenstein M, Blow FC. Racial disparity in the use of atypical antipsychotic medications among veterans. *Am J Psychiatry*. 2003;160(10):1817–22. [PMID: 14514496]  
DOI:10.1176/appi.ajp.160.10.1817
  34. Groeneveld PW, Kruse GB, Chen Z, Asch DA. Variation in cardiac procedure use and racial disparity among Veterans Affairs Hospitals. *Am Heart J*. 2007;153(2):320–27. [PMID: 17239696]  
DOI:10.1016/j.ahj.2006.10.032
  35. McGinnis KA, Fine MJ, Sharma RK, Skanderson M, Wagner JH, Rodriguez-Barradas MC, Rabeneck L, Justice AC; Veterans Aging Cohort 3-Site Study (VACS 3). Understanding racial disparities in HIV using data from the veterans aging cohort 3-site study and VA administrative data. *Am J Public Health*. 2003;93(10):1728–33. [PMID: 14534229]  
DOI:10.2105/AJPH.93.10.1728
  36. Safford M, Eaton L, Hawley G, Brimacombe M, Rajan M, Li H, Pogach L. Disparities in use of lipid-lowering medications among people with type 2 diabetes mellitus. *Arch Intern Med*. 2003;163(8):922–28. [PMID: 12719201]  
DOI:10.1001/archinte.163.8.922
  37. Alexander DD, Waterbor J, Hughes T, Funkhouser E, Grizzle W, Manne U. African-American and Caucasian disparities in colorectal cancer mortality and survival by data source: An epidemiologic review. *Cancer Biomark*. 2007;3(6):301–13. [PMID: 18048968]
  38. Dobscha SK, Dickinson KC, Lasarev MR, Lee ES. Associations between race and ethnicity and receipt of advice about alcohol use in the Department of Veterans Affairs. *Psychiatr Serv*. 2009;60(5):663–70. [PMID: 19411355]  
DOI:10.1176/appi.ps.60.5.663
  39. Dominitz JA, Samsa GP, Landsman P, Provenzale D. Race, treatment, and survival among colorectal carcinoma patients in an equal-access medical system. *Cancer*. 1998;82(12):2312–20. [PMID: 9635522]  
DOI:10.1002/(SICI)1097-0142(19980615)82:12<2312::AID-CNCR3>3.0.CO;2-U
  40. Giordano TP, Morgan RO, Kramer JR, Hartman C, Richardson P, White CA Jr, Suarez-Almazor ME, El-Serag HB. Is there a race-based disparity in the survival of veterans



- with HIV? *J Gen Intern Med.* 2006;21(6):613–17. [PMID: 16808745]  
DOI:10.1111/j.1525-1497.2006.00452.x
41. 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]
  42. Revisions to the standards for the classification of federal data on race and ethnicity [Internet]. Washington (DC): Office of Management and Budget; 1997. Available from: <http://www.whitehouse.gov/omb/fedreg/1997standards.html>.
  43. VHA Directive 2003-027, Capture of race and ethnicity categories. Washington (DC): Department of Veterans Affairs; 2003.
  44. Citro CF, Cork DL, Norwood JL, editors. Panel to review the 2000 census, National Research Council. The 2000 census: Counting under adversity. Washington, DC: National Academies Press, 2004.
  45. Hirschman C, Alba R, Farley R. The meaning and measurement of race in the U.S. census: Glimpses into the future. *Demography.* 2000;37(3):381–93. [PMID: 10953811]  
DOI:10.2307/2648049
  46. Saboeiro AP, Porkorny JJ, Shehadi SI, Virgo KS, Johnson FE. Racial distribution of Dupuytren's disease in Department of Veterans Affairs patients. *Plast Reconstr Surg.* 2000;106(1):71–75. [PMID: 10883614]  
DOI:10.1097/00006534-200007000-00013
  47. VIREC research user guide: FY2004 VHA Medical SAS Inpatient dataset. Hines (IL): Veterans Affairs Information Resource Center; 2005.
  48. VIREC research user guide: FY2004 VHA Medical SAS Outpatient datasets. Hines (IL): Veterans Affairs Information Resource Center; 2005.
  49. VA Information Resource Center [Internet]. Washington (DC): Department of Veterans Affairs; 2010 [updated 2010 Jun 18; cited 2009 Aug 1]. Available from: <http://www.virec.research.va.gov>.
  50. 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]  
DOI:10.2105/AJPH.86.5.712
  51. McBean M. Medicare race and ethnicity data. Washington (DC): National Academy of Social Insurance; 2004.
  52. Department of Veterans Affairs. Privacy act; Systems of records. *Federal Register.* 2009;74(142):37093–96.
  53. 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]
  54. Eicheldinger C, Bonito A. More accurate racial and ethnic codes for Medicare administrative data. *Health Care Financ Rev.* 2008;29(3):27–42. [PMID: 18567241]
  55. Waldo DR. Accuracy and bias of race/ethnicity codes in the Medicare enrollment database. *Health Care Financ Rev.* 2004;26(2):61–72.
  56. Hamilton NS, Edelman D, Weinberger M, Jackson GL. Concordance between self-reported race/ethnicity and that recorded in a Veteran Affairs electronic medical record. *N C Med J.* 2009;70(4):296–300. [PMID: 19835243]
  57. McBean AM. Improving Medicare's data on race and ethnicity. *Medicare Brief.* 2006;15:1–7. [PMID: 17036427]
  58. Waters M. Immigration, intermarriage, and the challenges of measuring racial/ethnic identities. *Am J Public Health.* 2000;90(11):1735–37. [PMID: 11076242]  
DOI:10.2105/AJPH.90.11.1735
  59. Doyle JM, Kao G. Are racial identities of multiracials stable? Changing self-identification among single and multiple race individuals. *Soc Psychol Q.* 2007;70(4):405–23. [PMID: 19823596]  
DOI:10.1177/019027250707000409
  60. Del Pinal J, Schmidley D. Matched race and Hispanic origin responses from Census 2000 and current population survey February to May 2000. Population Division Working Papers. Washington (DC): U.S. Census Bureau; 2005. Report No.: 79.

Submitted for publication August 12, 2009. Accepted in revised form March 30, 2010.

This article and any supplementary material should be cited as follows:

Stroupe KT, Tarlov E, Zhang Q, Haywood T, Owens A, Hynes DM. Use of Medicare and DOD data for improving VA race data quality. *J Rehabil Res Dev.* 2010; 47(8):781–96.

DOI:10.1682/JRRD.2009.08.0122





