

VIREC Technical Report

VA Race Data Quality

September 2011



Contact Information

VA Information Resource Center (VIREC)
Health Services Research and Development Service
Edward Hines Jr. VA Hospital
Routing mail code: 151V
5000 S. 5th Avenue
Hines, IL 60141-5000

VIREC Help Desk: virec@va.gov

VIREC website: <http://www.virec.research.va.gov/>

The VA Information Resource Center (VIREC) is supported by Health Services Research and Development Service grant SDR 98-004.

Suggested Citation

VA Information Resource Center. *VIREC Technical Report: VA Race Data Quality*. Hines, IL: U.S. Dept. of Veterans Affairs, Health Services Research and Development Service, VA Information Resource Center, Sept 2011.

This product is available from VIREC's Internet site: <http://www.virec.research.va.gov/>.

Contents

List of Tables & Figures 4

Acronyms List..... 6

Acknowledgements..... 7

Executive Summary 8

Introduction and Objectives 11

Background: VHA Race and Ethnicity Data 12

Methods..... 15

Analyses and Results 22

Discussion..... 42

Reference List 47

Appendix A..... 51

Appendix B..... 51

Appendix C..... 55

Appendix D..... 56

Appendix E 57

Appendix F..... 58

List of Tables & Figures

Tables

Table 1.	Race and Ethnicity Information Acquisition Process in VHA Facilities	12
Table 2.	Defined Race and Ethnicity Response Categories in Department of Veterans Affairs, Medicare, and Department of Defense Data	17
Table 3.	Usable Ethnicity Values	19
Table 4.	Sample Descriptors: Data Sources and Operational Definitions, VHA Medical SAS [®] Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Datasets, FY1997–FY2005.. ..	20
Table 5.	Sample Characteristics, Aims 1–3.....	23
Table 6.	Usable Race Values, VHA Medical SAS [®] Datasets, FY1997–FY2005, by Year	24
Table 7a.	Usable Ethnicity Values, Combined Veterans Health Administration Medical SAS [®] Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Datasets, FY1997–FY2003, by Year	25
Table 7b.	Usable Ethnicity Values, Combined Veterans Health Administration Medical SAS [®] Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Datasets, FY1997–FY2005, by Year	26
Table 8.	Characteristics of Individuals With and Without a Usable Race Value, Veterans Health Administration Medical SAS [®] Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Datasets, FY1997–FY2005	27
Table 9a.	Increases in Single-Year Race Data Completeness Using Additional Years of Veterans Health Administration Medical SAS [®] Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Data, FY1997–FY2005	29
Table 9b.	Increases in Single-Year Race Data Completeness Using Additional Years of Veterans Health Administration Medical SAS [®] Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Data, Individuals Under Age 65, FY1997–FY2005	29
Table 10.	Increases in Single-Year Ethnicity Data Completeness Using Additional Years of VHA Medical SAS [®] Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Data, FY1997–FY2005	30
Table 11.	Race Classification Mapping of Department of Veterans Affairs, Medicare, and Department of Defense Data	37
Table 12.	Sample Characteristics, Aim 4	38
Table 13a.	Accuracy of Medicare Race Data Compared to Self-Reported Veterans Health Administration Medical SAS [®] Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Race Data, FY2003–2005	40
Table 13b.	Accuracy of Department of Defense Race Data Compared to Veterans Health Administration Medical SAS [®] Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Race Data, FY2003–2005	40

Table B1.	RACE Variable Values	52
Table B2.	RACE1-RACE7 Variable Values	53
Table B3.	ETHNIC Variable Values	54
Table C1.	ETHNICITY Variable Values (DoD Data).....	55
Table D1.	Comparison of Study Sample to All VA Users.....	56
Table E1.	Characteristics of Individuals With and Without a Usable Race Value in VA Data, 1997-2005.....	57
Table F1.	Results of Match Between Medical SAS [®] and Medicare Records Shown Separately for the Full Study Sample and Individuals Age 65 and Older	58
Table F2.	Results of Match Between Medical SAS [®] and DoD Records Shown Separately for the Full Study Sample and Individuals Age 64 and Younger	58

Figures

Figure 1.	Trends in Usable Race Values, VHA Medical SAS [®] Datasets, FY1997–FY2005, by Fiscal Year.....	25
Figure 2.	Use of Medicare and Department of Defense Identity Repository Data to Supplement Race Information in Veterans Health Administration Medical SAS Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Datasets, FY1997–FY2005.....	31
Figure 3a.	Changes in Race Data Completeness After Adding Medicare Data, FY1997–FY2005	32
Figure 3b.	Changes in Race Data Completeness After Adding DoD Data, FY1997–FY2005 ..	32
Figure 3c.	Changes in Race Data Completeness After Adding Medicare and DoD Data, FY1997–FY2005	33
Figure 4.	Use of Medicare and Department of Defense Identity Repository Data to Supplement Ethnicity Information in Veterans Health Administration Medical SAS Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Datasets Using the Expanded Definition, FY1997–FY2005	34
Figure 5a.	Changes in Ethnicity Data Completeness After Adding Medicare Data, FY1997–FY2005	35
Figure 5b.	Changes in Ethnicity Data Completeness After Adding DoD Data to VA Data	35
Figure 5c.	Changes in Ethnicity Data Completeness After Adding Medicare and DoD Data to VA Data.....	36
Figure 6a.	Concordance Between VA and Medicare Race Values	38
Figure 6b.	Concordance Between VA and DoD Race Values.....	39
Figure 7.	Medicare Race Among VA Self-Reported Hispanics	41
Figure 8.	Agreement of Hispanic Ethnicity in Department of Defense Data with Self-Reported Ethnicity in Veterans Health Administration Medical SAS [®] Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Race Data, FY2003-FY2005	41

Acronyms List

Acronym	Definition
AITC	Austin Information Technology Center
APIO	Asian, Pacific Islander, or Other
DEERS	Defense Enrollment Eligibility Reporting System
DoD	Department of Defense
FY	fiscal year
MedSAS [®]	Medical SAS [®]
NPCD	National Patient Care Database
PDX	patient data exchange
PPV	positive predictive value
SD	standard deviation
SSN	Social Security number
VA	Department of Veterans Affairs
VADIR	Department of Veterans Affairs Department of Defense Identity Repository
VHA	Veterans Health Administration
VIReC	Department of Veterans Affairs Information Resource Center
VISN	Veterans Integrated Service Network
VistA	Veterans Health Information Systems and Technology Architecture
VSF	Vital Status File

Acknowledgements

This research was supported by the Department of Veterans Affairs, Veterans Health Administration, Health Services Research & Development Service, VA Information Resource Center, Project Number SDR-98-004. The contributing authors are Kevin Stroupe, Ph.D., Elizabeth Tarlov, Ph.D., R.N., Qiuying Zhang, M.S., Arika Owens, M.P.H.

We gratefully acknowledge input and review from the following individuals: James F. Burgess, Ph.D., M.A., B.A., Senior Investigator, VA Boston Healthcare System; Todd H. Wagner, Ph.D., from VA Health Economics Resource Center (HERC); John Quinn, B.S., Dorothea Garrett, Program Analyst, Katrina Tuisamatatele, and Larry Hughes from National Data Systems; Laureen Kennedy from VHA OHI HDI Data Quality. Deborah Berlyne, Ph.D. and Patricia Socha, M.A. provided editorial assistance.

This report presents the findings and conclusions of the authors; it does not necessarily represent the Department of Veterans Affairs or Health Services Research & Development Service.

Executive Summary

To inform investigators using Veterans' race and ethnicity information from Veterans Health Administration (VHA) administrative databases, the Department of Veterans Affairs (VA) Information Resource Center (VIREC) studied the completeness of race and ethnicity information in the VHA Medical SAS[®] Datasets. The study also assessed the feasibility and utility of several potential approaches to reducing missingness in research datasets. This report describes the results of this study.

Aims

This study had four specific aims:

1. Describe trends over time in the proportion of patients whose race or ethnicity can be ascertained using VHA administrative (i.e., workload) data.
2. Elucidate patient characteristics associated with unknown race or ethnicity.
3. Evaluate the extent to which missing race and ethnicity information in VHA workload data can be filled in using:
 - a. Other VHA data files.
 - b. Information from Medicare.
 - c. Information from the Department of Defense (DoD).
4. Assess the likelihood that the data obtained from external sources (Medicare and DoD) will mirror the race and ethnicity information that would have been available in the VHA data if patients had provided this information.

Sample and Data Sources

We identified our sample from records in the VHA Vital Status (Master) File (VSF), which contains identifying and vital status information for every patient who has received care through the VHA since 1992, has enrolled for care in the VA, or has received compensation or pension benefits from the VA since 2002.

We obtained Medicare data on race, ethnicity, date of birth, and gender from the Medicare Vital Status File, which contains demographic information about each beneficiary who has ever been eligible for Medicare services. Our DoD race information came from the VA-DoD Identity Repository (VADIR) database, which contains data on all Veterans who separated from the military in 1980 or later and for some Veterans who left the military prior to 1980.

We randomly selected 10% of records in VSF. After excluding records lacking date-of-birth or gender information and performing a deterministic match of the sample with a list of all unique combinations of Social Security number (SSN), date of birth, and gender in the VHA inpatient and outpatient Medical SAS[®] Datasets, we obtained 939,021 matches with 936,469 unique SSNs.

Results

Aim 1

For all fiscal years combined, we found a “usable” race value (i.e., a value that was not missing or categorized as ‘unknown’) for 56% of individuals. For every year, inpatient records had substantially more complete race data than outpatient records.

The VA began collecting and reporting (Hispanic) ethnicity separately from race in FY2003. The proportion of individuals who had a usable ethnicity value increased markedly in the first three years after the transition, from 2% in FY2003 to 52% in FY2005.

Aim 2

Compared to individuals whose race we ascertained in VA data, those without a usable race value were younger (mean age 59 vs. 65) and less likely to be male (82% vs. 94%) or have non-VA healthcare coverage (34% vs. 51%). When we compared patient characteristics in single-year samples, we found that differences in the demographic characteristics of the two groups were quite substantial in FY1997 and decreased over time.

Aim 3

Filling in missing data in single fiscal year samples with usable race values from other years of VA data decreased the proportions of missing data by up to 29%. The largest decreases occurred in records from the earlier years, reflecting the greater utility of searching more recent rather than earlier years of data to fill in missing information. Adding information from additional years of VHA data also increased VHA ethnicity data completeness substantially, to 67%.

Adding Medicare data improved race data completeness to 97% among those aged 65 and older, 67% among those younger than 65, and 75% overall (over and above the 56% completeness achieved by combining all years of VA data). Adding DoD data increased race data completeness to 61% among those aged 65 and older, 71% among those younger than 65, and 67% overall. Combining Medicare and DoD with VHA data increased race data completeness to 85% in all Veterans, 76% in Veterans under age 65, and 97% in those aged 65 and older.

Medicare data provided little additional ethnicity information for our cohort. The addition of DoD data, however, improved ethnicity data completeness to 60% in both older and younger age groups. Combined Medicare and DoD data increased ethnicity data completeness to 60% in both age groups.

Aim 4

Our results showed a high level of concordance between VA and Medicare data for Whites (99%) and African Americans (96%). However, only 36% of individuals categorized as American Indian or Alaska Native in VA data were listed as North American Native in Medicare data. Furthermore, only 47% of those identified as Asian, Pacific Islander, or Other in VA data were similarly identified in Medicare data.

We also found a high level of concordance between VA and DoD data for individuals identified as White (93%) or African American (95%) in VA data. However, just 39% of individuals identified as American Indian or Alaska Native in VHA data were listed as the same in DoD data and only 65% of individuals identified as Asian, Pacific Islander, or Other in VHA data were similarly listed in DoD data.

Recommendations

- Avoid excluding data on patients whose race is not known from sample analyses because this changes important sample characteristics and could bias the study results. Consider instead creating an “unknown” category for the study’s race and ethnicity variable and analyze data separately on patients whose race or ethnicity is not known.
- To supplement missing race and ethnicity data in VHA datasets, consider searching multiple years of VA data, particularly if more recent data are available, and supplementing VA data with Medicare data, DoD data, or both.
- Add DoD data to yield substantial additional race information on younger cohorts.
- When supplementing VA data with Medicare or DoD data, use a dichotomous grouping of African Americans and non-African Americans to produce the highest rates of accurate classification.
- For research on non-African American minorities, consider using other data sources, such as Indian Health Service or survey data, to supplement VHA data on race.
- When using the VA’s VSF to supplement race and ethnicity data, identify patient matches using a combination of date of birth, gender, and SSN to maximize the likelihood of identifying the right individuals.
- Do not use the method of collection information in the VA race and ethnicity variable (e.g., Black – self-identification) since there is known systematic misreporting of the data collection method.

Introduction and Objectives

The central role of race and ethnicity in explaining differences in health and healthcare in the United States is well known and continues to receive much research attention. The Office of Research and Development, U.S. Department of Veterans Affairs (VA), has identified research on health disparities and minority health as a priority funding area¹ and many VA studies focusing on race are in progress or have been completed [1-13]. In many of these studies, investigators have relied on information obtained from Veterans Health Administration (VHA) administrative databases (i.e., workload data) to ascertain patient race and ethnicity. However, the limited quality and completeness of this information has been identified as a potential limitation to this research [14-17].

To inform investigators using Veterans' race and ethnicity information from VHA administrative databases, the VA Information Resource Center (VIREC) studied the completeness of race and ethnicity information in the VHA Medical SAS[®] datasets and assessed the feasibility and utility of several potential approaches to reducing missingness in research datasets.

The study had four specific aims:

1. Describe trends over time in the proportion of patients whose race or ethnicity can be ascertained using VHA workload data.
2. Elucidate patient characteristics associated with unknown race or ethnicity.
3. Evaluate the extent to which missing race and ethnicity information in VHA workload data can be filled in using:
 - a. Other VHA data files.
 - b. Information from Medicare.
 - c. Information from the Department of Defense (DoD).
4. Assess the likelihood that the data obtained from external sources (Medicare and DoD) will mirror the race and ethnicity information that would have been available in the VHA data if patients had provided this information.

In this report, we describe the results of this study.

Background: VHA Race and Ethnicity Data

Race and Ethnicity Data Collection and Entry

The VA collects information on VA patient race primarily from VA Form 10-10EZ (*Application for Health Benefits*) which patients complete at enrollment (Appendix A). Veterans can complete and submit VA Form 10-10EZ online, download the form from the Internet and deliver it to the facility by mail or in person, or complete the form through a telephone or face-to-face interview with personnel at the facility where the Veteran intends to receive care [18]. This information is entered into the VA Information Systems Technology Architecture (VistA) system. Following the initial enrollment process, personnel responsible for updating the patient record may fill in missing information at the time of a patient encounter (e.g., hospital admission or clinic visit) [19].

During training, staff are instructed to collect this information from patients by asking them to identify their own race and ethnicity when possible. Otherwise, staff should collect the data by proxy from a friend or relative accompanying the patient. If the patient or proxy declines to provide the information, the VHA staff are supposed to enter “Declined to Answer” into the patient’s record [19]. If the patient has not already provided the information and the staff member does not ask about the patient’s race or ethnicity (at enrollment or during a future healthcare encounter), the staff member is supposed to leave the race and ethnicity fields blank [20]. Patients are not required to provide race and ethnicity information to the VA.

The processes through which race and ethnicity information is collected and entered into a VistA system are illustrated in Table 1.

Table 1. Race and Ethnicity Information Acquisition Process in VHA Facilities

<i>Who</i> *	<i>What</i>	<i>When</i>	<i>Where</i>	<i>How</i>
<ul style="list-style-type: none"> • Patient (self-report) • Proxy 	<ul style="list-style-type: none"> • VA Form 10-10EZ, <i>Application for Health Benefits</i> • Two interview questions 	<ul style="list-style-type: none"> • Enrollment in VHA facility • Hospital admission • Outpatient visit 	<ul style="list-style-type: none"> • Online • By mail • Telephone call with local VHA facility • In-person visit to local VHA facility 	VHA facility enrollment coordinator or designee (e.g., admission interview clerk or enrollment specialist) or outpatient clinic staff member collects the information.

*Information Source

VistA systems operate independently of one another. Prior to FY2003 for inpatients and FY2004 for outpatients, when a patient enrolled in one VHA facility and then visited another VHA facility that used a different VistA system, facility staff populated the new record by collecting and entering the race and ethnicity information manually or by requesting an electronic patient data exchange (PDX) file from another facility. Staff could import patient data from these PDX

files into the local VistA system. Despite this process, some Medical SAS[®] (MedSAS) records lacked a usable race or ethnicity value even when a record from an earlier visit to a different facility contained usable data. Today, the second VistA system automatically populates the new VistA record with the patient's race and ethnicity (and other) information based on the response to an automated query issued when the patient's identification number is entered into VistA.

Once entered into VistA, race and ethnicity and other demographic information remains part of the VistA record for that patient, unless the information is altered manually. Race and ethnicity information may be altered at the patient's request. In the past, information resource management staff sometimes altered records by assigning a value of "Unknown" to blank race and ethnicity fields. This practice has been prohibited since January 2006 by VHA Directive 2006-002.

Changes in VA Race and Ethnicity Data Collection

Changes in the way the VA collects and records race information in response to a revision of the Office of Management and Budget's Statistical Policy Directive No. 15, *Race and Ethnic Standards for Federal Statistics and Administrative Reporting*, present challenges to researchers who use these data [21]. The directive revised the standards for collecting, maintaining, and presenting federal data on race and ethnicity. The new standards redefined minimum categories for reporting race and ethnicity (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White), indicated a preference for collecting self-reported or self-identified race and ethnicity using two separate questions, and established the requirement to offer respondents the option of selecting more than one racial designation. The VHA, through Directive 2003-027, implemented the mandated revisions beginning in 2003 [22].

Implementation of the revised standards has probably improved the validity of recorded race and ethnicity values in VA data because self-reported race and ethnicity information is considered the "gold standard" for collecting this information [23-26]. The change in collection method and in race and ethnicity categories in 2003 can create problems for studies whose cohorts were accrued before and after 2003. Furthermore, the new standards were not implemented simultaneously in different VHA facilities or for inpatient and outpatient data. For example, population of the new race and ethnicity fields began in FY2003 for inpatient files and in FY2004 for outpatient files.

Race and Ethnicity Data Storage

Veterans' race and ethnicity data are stored in the VistA Race Information and Patient Information Sub-files [27]. VA facilities transmit these files with patient healthcare encounter data to the Austin Information Technology Center (AITC), where this information is stored in the National Patient Care Database (NPCD). National Data Systems, the VHA division that serves as the data steward for some AITC data, directs the construction of extracts of these data, known as the Medical SAS[®] Datasets. The Medical SAS[®] Datasets contain nationwide electronic health record data (known as "workload data") for VHA-provided healthcare encounters.

Race and ethnicity information on patients who received inpatient care is stored in the Medical SAS[®] Acute Inpatient, Extended, Observation, Non-VA Main, and Inpatient Encounters Files [28]. Race and ethnicity information collected from patients receiving outpatient care is contained in the Medical SAS[®] Visit and Event Files [27].

Prior to 2003, the software that managed the transmission process was configured to block transmission of VistA inpatient records to AITC that did not have a valid race/ethnicity value or in which the race/ethnicity field was blank. For this reason, the extent of missing race information in Medical SAS[®] inpatient records prior to FY2003 is low. VistA software changes implemented in response to the new standards in FY2003 allowed inpatient records with blank race and ethnicity values to be transmitted to AITC for the first time.

The VA had no policies governing the collection of race information in the outpatient setting prior to FY2003. Until 2003, outpatient records were populated with race information obtained from inpatient records. As a result, most of the outpatient records from before 2003 with race and ethnicity information are for patients who had an inpatient stay in a VHA facility.

Methods

We selected a representative sample of patients who received treatment in the VA's outpatient or acute inpatient settings between FY1997 and FY2005 (see next section, Study Sample and Data Linkage, for details on how we selected our patient sample). We evaluated whether their records contained usable race and ethnicity values or, in other words, whether these values could be used to identify the patient's race or ethnicity. We then examined information obtained from Medicare and DoD data for the same individuals to estimate the extent to which we could reduce missingness in VA study samples by using race and ethnicity information from those external sources.

Data Sources

VHA Vital Status File

We identified our sample from records contained in the VHA Vital Status (Master) File (VSF), which contains identifying information (Social Security number [SSN], birth date, and gender) and vital status information for every patient who has enrolled for care in the VA, has received care through the VHA since 1992 (regardless of enrollment status), or has received compensation or pension benefits from the VA since 2002 [28, 29]. Because the identifying information from the data sources that feed the VSF do not always agree, the Master VSF contains one record for each combination of SSN, birth date, and gender.

We obtained our sample on January 23, 2007 from the VSF created in September, 2006. At that time, the VSF contained 19,231,724 records. We obtained the identifying data elements of SSN, birth date, and gender from the VSF.

VHA Medical SAS[®] Datasets

The VHA Medical SAS[®] Datasets are provided in SAS[®] format by fiscal year and reside at AITC [27,28]. For this study, we used the Outpatient Visit and Inpatient Acute Main files for FY1997–FY2005. Each record in the Outpatient Visit dataset contains data on the services provided to a patient at a VA facility on a single day. Information in these records can be generated from multiple provider encounters (e.g., clinic visits) or services (e.g., radiology exams), all provided on the same day and at the same facility. The Inpatient Acute Main file includes one record for each inpatient stay in a VA acute care hospital. Each fiscal year file contains information on the stays that resulted in a discharge during that year.

We obtained information on race and ethnicity, other demographic characteristics, and numbers of inpatient stays and outpatient visits from these files.

Medicare Vital Status File

We obtained Medicare data on race/ethnicity, date of birth, and gender from the Medicare Vital Status File. This file contains demographic information about each beneficiary who has ever been eligible for Medicare services. VIREC maintains the Medicare Vital Status File for Veterans in the VHA cohort (i.e., all Veterans known to the VA) [30]. VIREC updates the file annually to reflect changes in enrollment and vital status. The race/ethnicity information in Medicare records comes primarily from the Social Security Administration and has known data quality problems, including some “unknown” values [31,32].

VA-DoD Identity Repository Database

Our DoD race and ethnicity information came from the VA-DoD Identity Repository (VADIR) database. The Office of Enterprise Development in the VA’s Office of Information maintains VADIR, which contains data on all Veterans who separated from the military in 1980 or later and on some Veterans who left the military prior to 1980 [33]. A data-sharing agreement between the VA and DoD enables the VA to obtain these data from the Defense Manpower Data Center’s Defense Enrollment Eligibility Reporting System (DEERS) database.

The DEERS database has data on self-reported race and ethnicity that service members provide when they initially join the military during the military entrance process. Active-duty military members may alter their race and ethnicity information online or by contacting the service personnel office. At present, service members have the option of declining to specify their race.

Race and Ethnicity Variables

The categories of race and ethnicity in VA, Medicare, and DoD data are shown in Table 2.

Race and Ethnicity Variables in VHA Medical SAS[®] Datasets

Current VA race and ethnicity data elements include RACE1–RACE7 (allowing reporting of up to seven races) and ETHNIC (indicating Hispanic or Latino ethnicity). The formatted values for RACE1–RACE7 and ETHNIC in the Medical SAS[®] Datasets incorporate method of collection (see Appendix B). For example, formatted value 9S is “Black or African American – Self Identification,” whereas 9P indicates “Black or African American – Proxy.” However, although formatted values are available to indicate methods of data collection other than self-identification, such as proxy, the software is configured to allow only self-identification as the reported data-collection method [22]. For example, if the patient’s wife reports that the patient is White, the value in the data will be “White – Self-Identification,” even though the information was reported by a proxy. Other values appear in the data occasionally because local health information management staff may modify the software to accommodate the entry of other values for data-collection method.

The Medical SAS[®] Datasets also maintain the pre-transition race and ethnicity field, RACE, to keep the file structure consistent over time. That field continues to be populated in the outpatient Medical SAS[®] Datasets for FY2004 and later (i.e., it is extracted from the NPCD), even though VHA facilities no longer enter the information for this field into their local VistA system. In the inpatient datasets from FY2004 and later, the RACE field is empty.

Table 2. Defined Race and Ethnicity Response Categories in Department of Veterans Affairs, Medicare, and Department of Defense Data

Data Source	Time Period	Variable Name	Label
VA	FY2003 and earlier	Ethnicity and Race	Hispanic, White Hispanic, Black American Indian Asian Black White, Unknown
	FY2003 /FY2004 to present*	Ethnicity	Hispanic or Latino (yes or no)
Medicare		Race	American Indian or Alaska Native Asian Black or African-American Native Hawaiian or Pacific Islander White
		Ethnicity and Race	Hispanic White Black Asian Native American Other
DoD		Ethnicity	Multiple†
		Race	White Black Asian or Pacific Islander American Indian or Alaskan Native Other Unknown

*The transition in response categories occurred in FY2003 for inpatient data and in FY2004 for outpatient data.

†DoD uses 21 ethnicity categories (see Appendix C).

Study Variable Definitions

Usable Race Values

We defined a usable race value as one that had research utility because it assigned a race to an individual. In VA data, values of Unknown or Declined to answer (post-transition records) and missing values were not usable. In pre-transition records, usable race values were Hispanic, White; Hispanic, Black; American Indian; Black; Asian; and White. In post-transition records, usable race values were White; Black or African American; Asian; American Indian or Alaska Native and Native Hawaiian or Other Pacific Islander, without regard to the method of collection. Usable race values in VA data are shown in Appendix B. In Medicare data, only missing values were unusable. In DoD data, Unknown and missing values were unusable.

Usable Ethnicity Values

Like a usable race value, a usable ethnicity value was theoretically one that assigned an ethnicity (Hispanic or Latino, or not Hispanic or Latino) to an individual. However, the operational definition of “usable ethnicity value” was less straightforward because Medicare’s datasets contain just one combined race/ethnicity variable and the VA’s pre-transition categories defined a valid ethnicity value somewhat ambiguously. In the VA data, the pre-transition categories of Hispanic White; Hispanic Black; White; Black; Asian, and American Indian had an implied assumption that a value of White or Black indicated that the respondent had non-Hispanic ethnicity and did not allow the specification of Hispanic ethnicity for Asians and American Indians.

We defined an ethnicity value in pre-transition VA data (Table 3) as usable if the RACE field contained a value of Hispanic, White or Hispanic, Black (in which case we categorized the patient as Hispanic) *or* a value of Black or White (in which case we categorized the patient as non-Hispanic). In all other cases (i.e., values of American Indian; Asian; Unknown, or Missing), we treated the patient’s ethnicity status as unknown and the record had no usable ethnicity value. The introduction of the ETHNIC data field into the VA’s Medical SAS[®] Datasets in FY2003 for inpatient files and in FY2004 for outpatient files removed the ambiguity in the ethnicity categorization of VA data.

For Medicare data, we operationally defined a usable ethnicity value as a RACE_CD variable value of 5 (Hispanic). Because respondents had to choose between Hispanic and the race options (White; Black; Asian; Native American; Other), we could not assume that the selection of one of the RACE_CD categories other than Hispanic indicated that the respondent was not Hispanic. This Medicare categorization of race is known to under-report Hispanic ethnicity [32].

See Table 3 for list of defined VA and Medicare ethnicity values.

Table 3. Usable Ethnicity Values*

Data Source	Time Period	Variable Name	Coded Value	Label
VA	FY1997–2003	RACE	1	Hispanic, White
			2	Hispanic, Black
			4	Black
			6	White
	FY2004–2005	ETHNIC	H	Hispanic or Latino – Missing
			HO	Hispanic or Latino – Observer
			HP	Hispanic or Latino – Proxy
			HS	Hispanic or Latino – Self-identification
			HU	Hispanic or Latino – Unknown by patient
			N	Not Hispanic or Latino – Missing
			NO	Not Hispanic or Latino – Observer
			NP	Not Hispanic or Latino – Proxy
			NS	Not Hispanic or Latino – Self-identification
			NU	Not Hispanic or Latino – Unknown to patient
Medicare	FY1997–2005	RACE_CD	5	Hispanic
DoD	FY1997–2005	ETHNICITY	Any non-missing value	

*RACE variable values American Indian and Asian, Unknown, ETHNIC variable values, Declined to Answer and Unknown, RACE_CD values of White, Black, Asian, Native America, Other, and all missing values were defined as not usable.

The ethnicity variable in the VADIR database, ETHNIC, contains 21 response categories (see Appendix C). We defined a usable ethnicity value as any non-missing value. For Aim 4, in which we compared ethnicity agreement across datasets, we collapsed the ethnicity categories to two, Hispanic or Latino or not Hispanic or Latino. Values we considered Hispanic or Latino were Hispanic, Puerto Rican, Mexican, Cuban, and Latin American with Hispanic descent.

Patient Characteristics

We developed operational definitions (Table 4) of patient age, gender, marital status, and geographic region for our study based on Medical SAS[®] variables (VIREC Research User Guide: VHA Medical SAS[®] Outpatient Datasets, 2004). We obtained the value we used from the most recent record with a usable value for that variable.

See Table 4 for a complete list of the operational definitions of patient characteristics we developed for our analysis.

Table 4. Sample Descriptors: Data Sources and Operational Definitions, VHA Medical SAS[®] Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Datasets, FY1997–FY2005

Characteristic	Medical SAS [®] Dataset		Variable Name	Operational Definition
	Inpatient	Outpatient		
Age	X		BORNDAY	Age on January 1, 2004
		X	DOB	Age on January 1, 2004
Gender	X	X	SEX	M: male F: female
Marital Status	X		MS	M: married N: not married D: divorced S: separated W: widowed U or blank: unknown or missing data
		X	Marital	M: married N: never married S: single D: divorced A: separated W: widowed U or blank: unknown or missing data
Geographic Region	X	X	VISN	Northeast: VISNs 1–4 South: VISNs 5–9, 16, 17 Midwest: VISNs 10–12, 15, 23 West: VISNs 18–22

Note 1: VISN: Veterans Integrated Service Network

Note 2: The combination of Social Security number, date of birth, and gender identified each individual. For other descriptors, the value used came from the most recent record with a usable value for that variable.

Study Sample and Data Linkage

We randomly selected 10% of records in the VHA VSF. This procedure yielded 1,923,171 records with 1,854,384 unique SSNs. We excluded records lacking date-of-birth or gender information (236,564 records). We then performed a deterministic match of the sample with a list of all unique combinations of SSN, date of birth, and gender in the Medical SAS[®] Outpatient Visit and Inpatient Acute Main Datasets from the same time period. This procedure yielded 939,021 matches with 936,469 unique SSNs. The bulk of those for whom we found no match were deceased (an estimated 110,572 individuals) or did not use VA healthcare services between FY1997 and FY2005. Because we could not determine whether records containing the same SSN were for the same or different people, we operationally defined an “individual” as a person identified by a unique combination of SSN, date of birth, and gender.

To verify that our sample selection method yielded a representative sample of users of VHA outpatient and acute inpatient care, we compared the age and gender of individuals in our sample with the characteristics of all individuals (9,405,890 users, defined by unique combinations of SSN, date of birth, and gender) in the FY1997–FY2005 Visit and Acute Main datasets. We found that the age and gender distributions in the 10% and 100% samples were nearly identical (see Appendix D).

Analyses and Results

Aim 1

Aim 1: Describe trends over time in the proportion of patients whose race or ethnicity can be ascertained using VHA workload data.

Analysis

We described and compared rates of race and ethnicity data missingness overall and by year. We calculated the proportions of the study sample for which we found one or more usable race and ethnicity values in the Medical SAS[®] data during the study period. We evaluated outpatient and inpatient data separately and then together.

For the by-year comparisons, we examined all records for each individual during a single fiscal year. We calculated the proportions for which we found usable race and ethnicity values among those who had one or more records (i.e., had received outpatient care in a VA facility, inpatient care, or both) during that fiscal year.

Results

The characteristics of the study sample are shown in Table 5. The age distribution reflects that of the population of VA service users; the mean age was 62, patients aged 65 years and older comprised 47% of the sample and 88% were male. The distribution by geographic region was proportionate to patient volume in the Veterans Integrated Service Networks (VISNs) within each region. Approximately 20% of the sample had had one or more inpatient stays and 99.8% had used outpatient services during the study period. On average, individuals contributed data to the study during 4 of the 9 years.

Table 5. Sample Characteristics, Aims 1–3 (N = 939,021)

Characteristics	N	%
Age		
65 years or older	439,344	46.8
	Mean (SD)	62.1 (17.4)
Gender		
Male	829,291	88.3
Marital status		
Married	481,425	51.3
Unmarried, widowed, divorced, or separated	335,252	35.7
Missing, unknown, or invalid value	122,344	13
Geographic region		
Northeast	163,963	17.5
South	323,959	34.5
Midwest	245,957	26.2
West	205,142	21.8
Inpatient utilization		
Number of inpatient stays	Mean (SD)	2.9 (3.3)
	Median	2
Outpatient utilization		
Number of outpatient visits	Mean (SD)	39.7 (78.3)
	Median	14
Years of data contributed to study sample		
	Mean (SD)	3.9 (2.8)

Note 1: SD: standard deviation

Note 2: The combination of Social Security number, date of birth and gender identified each individual.

Table 6 shows the proportion of the study sample for which we obtained one or more usable race values in VA inpatient or outpatient data in each year of the study period. For all fiscal years combined, we found a usable race value for 56% of individuals.

Table 6. Usable Race Values, VHA Medical SAS® Datasets, FY1997–FY2005, by Year*

Fiscal Year	Sample Size [†]	Percentage with a Usable Race Value
	n	%
All years combined	939,021	56
1997	300,272	57
1998	328,050	57
1999	342,525	58
2000	367,004	58
2001	407,993	56
2002	446,416	55
2003	471,794	49
2004	488,947	62
2005	502,205	67

*A usable race value for records prior to 2003 is defined as a value that is not missing or coded as “unknown.” For records after FY 2003, a usable race value comes from the legacy RACE field or the newly introduced RACE1 field.

[†]Each single-year sample consists of the subgroup of the full study sample that used VA healthcare during that year.

The proportion of individuals with a usable race value was much higher among the subset who had received inpatient services, 88% for all years combined, than those who had received outpatient services only. This rate was highest (98%) at the beginning of the study period, in FY1997, and declined thereafter (Figure 1). Rates of usable race data in both inpatient and outpatient files were lowest in FY2003, the year of the transition to the new data collection and recording method, when a marked decline occurred. Although the proportion of usable outpatient and inpatient data subsequently increased, the inpatient rate remained substantially lower in FY2005 than in FY2002.

For ethnicity data, we present results separately for the two different data collection methods and race and ethnicity categorizations. Table 7a provides data on the proportions of the sample with usable ethnicity values between FY1997 and FY2003 obtained from the RACE field prior to use of self-report and separate reporting of race and ethnicity.

The proportions of the sample with a usable ethnicity value declined over this period, from 56.9% in 1997 to 48.0% in 2003.

Figure 1. Trends in Usable Race Values, VHA Medical SAS® Datasets, FY1997–FY2005, by Fiscal Year*

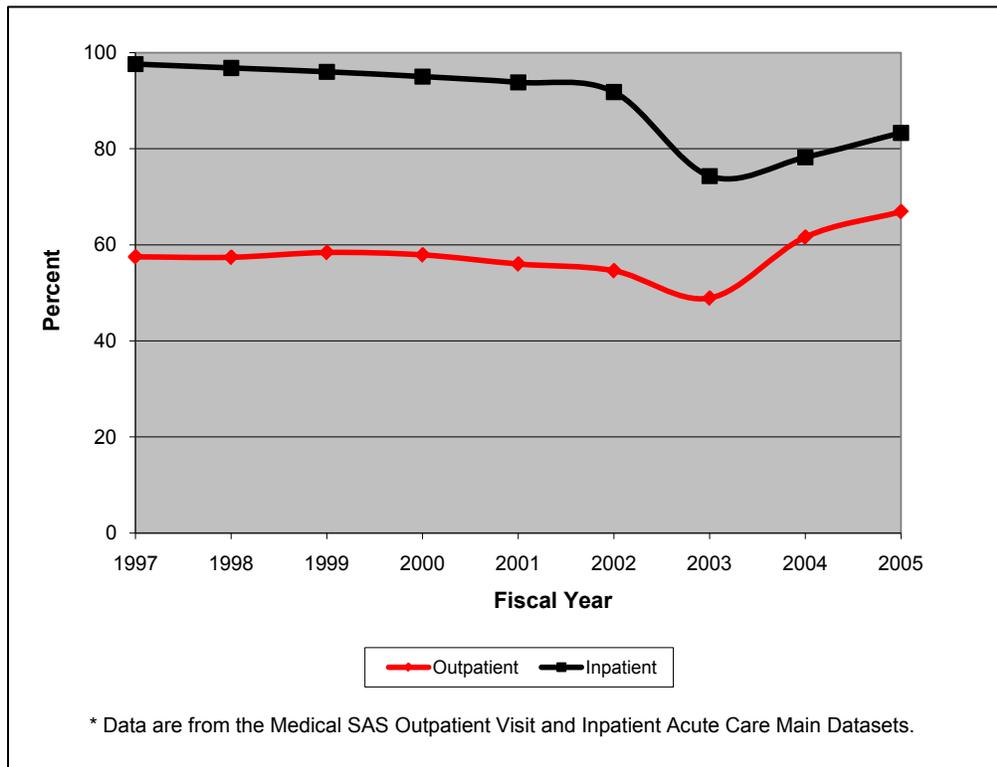


Table 7a. Usable Ethnicity Values, Combined Veterans Health Administration Medical SAS® Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Datasets, FY1997–FY2003, by Year*

Fiscal Year	Sample	Usable Ethnicity
	n	%
1997	300,272	56.9
1998	328,050	56.8
1999	342,525	57.7
2000	367,004	57.2
2001	407,993	55.4
2002	446,416	54.1
2003	471,794	48.0
Total	939,021	55.3

* A usable ethnicity value is one that definitively identifies the individual as Hispanic/Latino or not. See Table 3 for more detailed information.

In Table 7b, we provide the percentage completeness of ethnicity data for FY2003–FY2005, when ethnicity and race information were collected and recorded separately.

Table 7b. Usable Ethnicity Values, Combined Veterans Health Administration Medical SAS[®] Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Datasets, FY1997–FY2005, by Year*

Fiscal Year	Inpatient		Outpatient		Inpatient and Outpatient	
	Sample	Usable Ethnicity Value	Sample	Usable Ethnicity Value	Sample	Usable Ethnicity Value
	n	%	n	%	n	%
2003	36,106	30	471,213	(blank)	471,794	2
2004	36,472	47	488,342	39	488,947	39
2005	37,422	52	501,701	51	502,205	51

*A usable ethnicity value is one that definitively identifies the individual as Hispanic/Latino or not.

VA facilities began using the new data-collection method beginning in FY2003 but did not start reporting the new race and ethnicity fields in outpatient data until FY2004. Therefore, we present the inpatient and outpatient data separately. Among individuals who had one or more inpatient records, the proportion with a usable ethnicity value increased markedly in the first three years after the transition, from 30% to 52%. We observed a similar increase in outpatient data between FY2004 and FY2005, from 39% to 51%.

Aim 2

Aim 2: Elucidate patient characteristics associated with unknown race or ethnicity.

Analysis

For Aim 2, we divided the sample into two groups, those for whom we could and those for whom we could not find a usable race value for any year during the study period. We compared the frequency distributions for age, gender, marital status, insurance status, geography, and healthcare use in the two groups.

Results

Table 8 presents patient characteristics for the two groups studied in Aim 2. Compared to individuals whose race we ascertained in VA data, those without a usable race value were younger (mean age 59 vs. 65) and less likely to be male (82% vs. 94%) or have non-VA healthcare coverage (34% vs. 51%). Interestingly, the absence of a usable race value was associated with unknown marital status (27% vs. 2%). Those without a usable race value also had fewer inpatient stays and outpatient visits.

Due to the large sample size, all tests of differences between groups were statistically significant and, thus, non-informative. When we compared patient characteristics in single-year samples, we found that differences in the demographic characteristics of the two groups were quite substantial

in FY1997 and decreased over time (Appendix E). For example, in FY2005, we found no differences in average age or percent married among those who did and did not have a usable race value in VA data. Differences in gender and percentage with non-VA insurance were small.

Table 8. Characteristics of Individuals With and Without a Usable Race Value, Veterans Health Administration Medical SAS® Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Datasets, FY1997–FY2005

Characteristics	Usable Race Value *		
	Yes	No	
	(n=522,410)	(n=416,611)	
	%	%	
Ages			
65 years or older	51	42	
	Mean (SD)	64.5 (15.6)	59.1 (18.9)
Gender			
Male	94	82	
Marital status			
Married	54	47	
Unmarried, widowed, divorced, or separated	43	26	
Missing, unknown, or invalid value	2	27	
Insurance coverage			
Insured	51	34	
Uninsured	48	66	
Unknown	1	< 1	
Geographic region			
Northeast	17	18	
South	42	36	
Midwest	21	19	
West	19	25	
Unknown	<1	2	
Inpatient care use			
Inpatient stay (1 or more)	33.6	2	
Number of inpatient stays	Mean (SD)	1.0 (2.4)	0.03 (0.3)
Outpatient care use			
Number of outpatient visits	Mean (SD)	61.6 (97.3)	12.2 (24.3)
	Median	33	4
Years contributed to study sample	Mean (SD)	4.9 (2.8)	2.6 (2.1)

Note: SD - standard deviation

* A usable race value for records prior to 2003 is defined as a value that is not missing or coded as “unknown.” For records after FY 2003, a usable race value comes from the legacy RACE field or the newly introduced RACE1 field. All differences were statistically significant.

Aim 3

Aim 3: Evaluate the extent to which missing race and ethnicity data in VHA workload data can be filled in using:

- *Additional years of data from the Medical SAS[®] Datasets.*
- *Information from Medicare.*
- *Information from the DoD.*

Analysis

Additional Years of VHA Data

In this analysis, we examined single-fiscal year data files, focusing on the subgroup of individuals who used VA inpatient or outpatient services in that year. For each single-fiscal year subsample, we identified individuals whose records contained no usable race value. We then searched for a usable value in subsequent years of data for those individuals. We calculated the proportion of each single-year subsample for which we could convert a missing or unknown value to a usable value by this method.

We then calculated the same proportions among individuals younger than 65 as of December 31, 2004. Most members of this group were not eligible for Medicare coverage, so using additional years of VA data could be a useful method of increasing the completeness of these race data. We used the same method to calculate the proportions of single-fiscal-year subsamples for which we could find a usable ethnicity value in later years of VA data.

Supplementation of VHA Data with Information from Medicare and DoD Files

To obtain race information from non-VA sources for our VA sample, we tried to identify Medicare and DoD records for individuals who had no usable race or ethnicity information in their VA records (n = 416,611). We matched VA, Medicare, and DoD records using a deterministic procedure based on SSN, date of birth, and gender.

We calculated the proportions of the individuals with no usable race value in VHA data for whom we obtained a usable race value from Medicare or DoD data among the full sample, those younger than 65, and those aged 65 or older. We calculated the same proportions for ethnicity values.

Results

Additional Years of VHA Data

Table 9a shows the increases in the proportions of single-year samples with a usable race value when we added race information from other years of VA data.

Table 9a. Increases in Single-Year Race Data Completeness Using Additional Years of Veterans Health Administration Medical SAS® Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Data, FY1997–FY2005*

Fiscal Year	N*	Usable Race Value Found	
		Single-year Data	All Data†
		%	%
1997	300,272	57	74
1998	328,050	57	73
1999	342,525	58	74
2000	367,004	58	73
2001	407,993	56	71
2002	446,416	55	69
2003	471,794	49	67
2004	488,947	62	68
2005	502,205	67	69

*Each single-year sample is the subgroup of the study sample that used VA healthcare in that year.

†Using all available years of VA data in study period (FY1997–FY2005).

Table 9b shows the increase in race data completeness for individuals who were younger than 65 years. Using additional VA fiscal year files increased data completeness for the under-65 population to levels that were similar to the levels for the entire population, particularly for the most recent years.

Table 9b. Increases in Single-Year Race Data Completeness Using Additional Years of Veterans Health Administration Medical SAS® Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Data, Individuals Under Age 65, FY1997–FY2005*

Fiscal Year	N	Usable Race Value Found	
		Single-year Data	All Data†
		%	%
1997	147,279	50	69
1998	163,299	51	68
1999	170,635	53	70
2000	180,286	54	70
2001	194,464	55	70
2002	214,094	54	69
2003	234,019	48	67
2004	254,391	61	68
2005	271,692	67	69

*Each single-year sample is the subgroup of the study sample that used VA healthcare in that year.

†Using all available years of VA data in study period (FY1997–FY2005).

Filling in missing data in single fiscal year samples with usable race values from other years of VA data decreased the proportions of missing data by 3% (in FY2005) to 29% (in FY1997). The largest decreases occurred in data from the earlier years, reflecting the greater utility of searching more recent rather than earlier years of data to fill in missing information.

Table 10 shows the increase in the proportion of single-fiscal-year samples with a usable ethnicity value achieved with this approach. In the FY2003–FY2005 data files, the ETHNIC data element is also available and values for this variable contributed additional information. The resulting rates of completeness remained stable throughout the study period, at around 70%. Among individuals receiving care in the most recent year, FY2005, the ethnicity completeness rate was 67%, although records from additional years did not contribute meaningfully to this number.

Table 10. Increases in Single-Year Ethnicity Data Completeness Using Additional Years of VHA Medical SAS® Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Data, FY1997–FY2005

Fiscal Year	N	Single-Year Data	All Data*	Percent Increase
		%	%	
1997	300,272	56.9	73.9	30
1998	328,050	56.8	72.6	28
1999	342,525	57.7	73.3	27
2000	367,004	57.2	72.7	27
2001	407,993	55.4	71.1	28
2002	446,416	54.1	69.3	28
2003	471,794	48	66	38
2004	488,947	61.2	66.8	9
2005	502,205	66.9	67.3	0.6

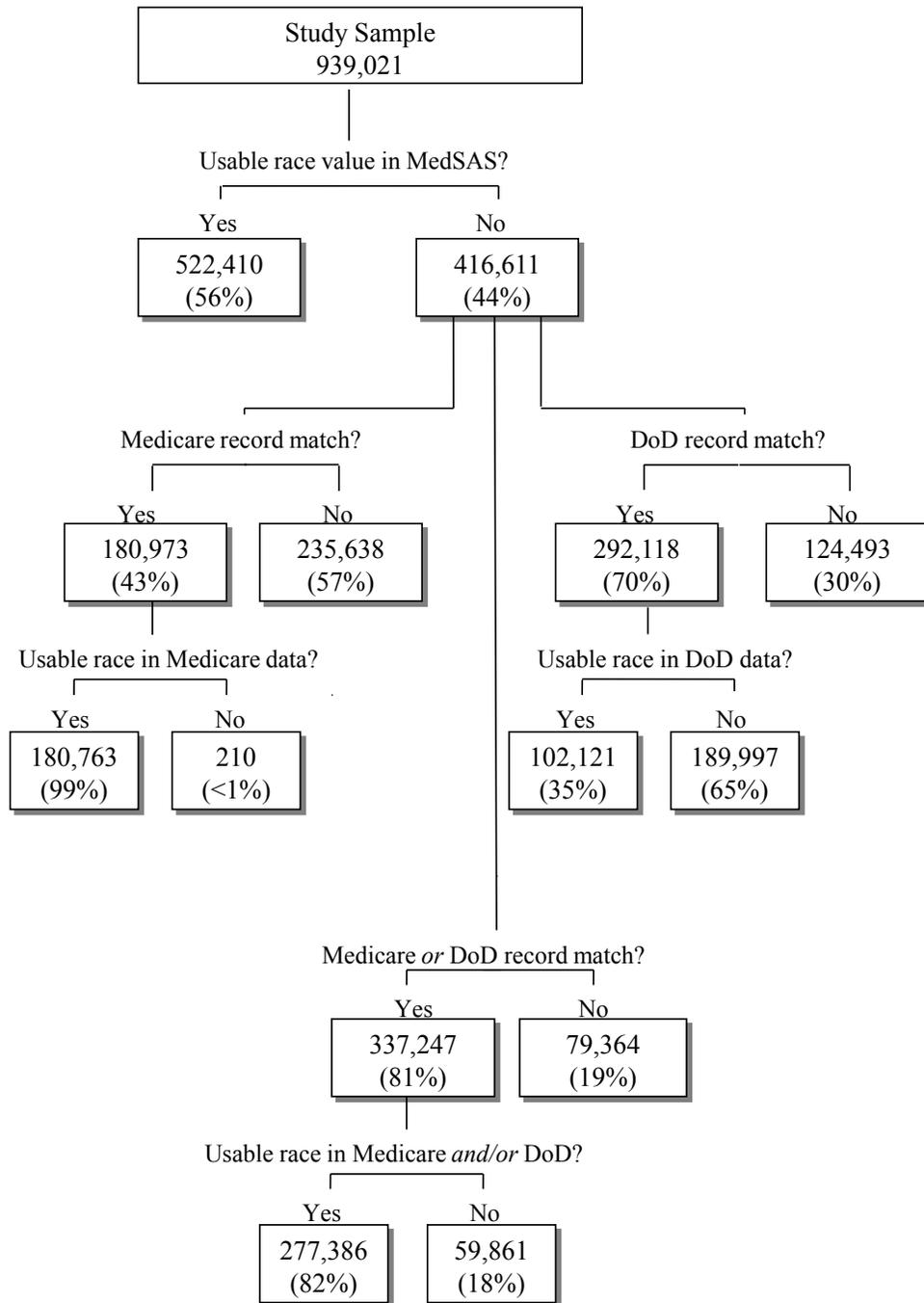
*Using all available years of VA data in the study period.

Adding information from FY2003 and later also contributed validity to the ethnicity data because these data came from patient self-reports and were collected separately from race data.

Supplementation of VHA Race Data with Information from Medicare and DoD Files

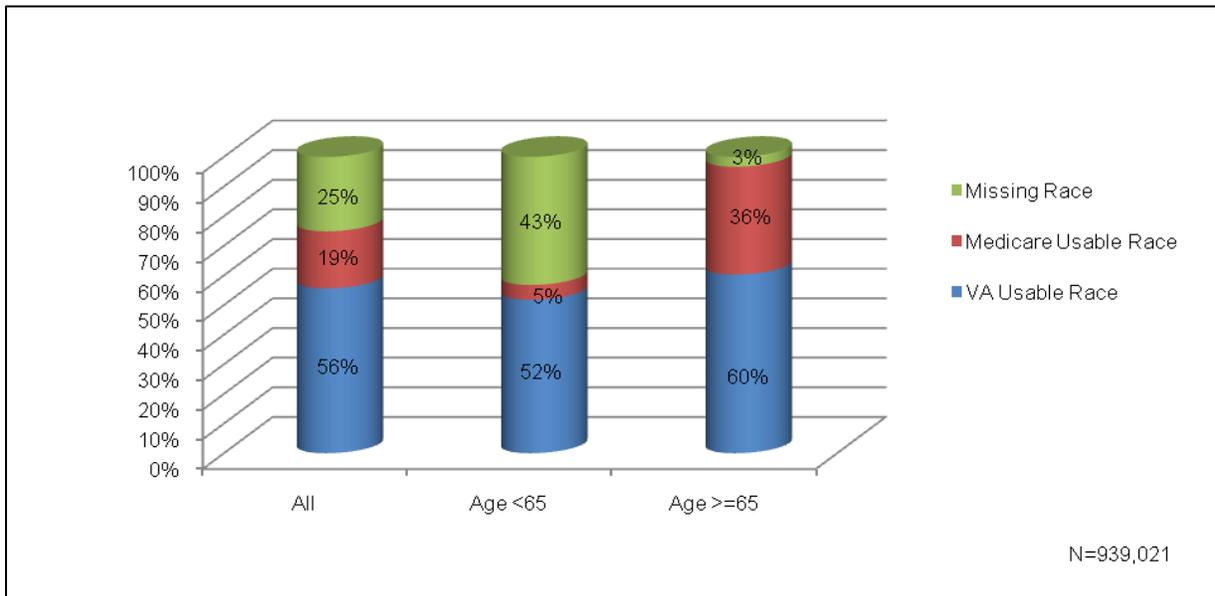
[For the 416,611 individuals (44% of the sample) for whom we found no usable race value in the FY1997–FY2005 VA data (Figure 2)]. We identified 180,973 Medicare records, for an overall match rate of 43%. Match rates were 92% among those aged 65 years and older and 10% among those younger than 65. We identified 292,118 records in DoD data that matched the VHA records, a match rate of 70% in both the older and younger age groups. Details on the match results are shown in Appendix F. We obtained race information from the Medicare record for 180,763 (99.9%) of the matches.

Figure 2. Use of Medicare and Department of Defense Identity Repository Data to Supplement Race Information in Veterans Health Administration Medical SAS Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Datasets, FY1997–FY2005



The addition of Medicare data increased race data completeness to 75% in the two age groups combined, and to 97% among those aged 65 and older and 57% among those younger than 65 (Figure 3a).

Figure 3a. Changes in Race Data Completeness After Adding Medicare Data, FY1997–FY2005



We obtained race information from the DoD records for 102,121 (35%) of the matches (Figure 2). DoD data contained a usable race value for 58% of individuals in the younger age group but for less than 2% of those aged 65 and older (not shown). The addition of DoD data increased race data completeness in our sample to 67% overall, 61% among those aged 65 and older, and 71% among those aged 64 and younger (Figure 3b).

Combining Medicare and DoD data increased race data completeness to 85% in all Veterans in our sample, 76% in Veterans under age 65, and 97% in those aged 65 and older (Figure 3c).

Figure 3b. Changes in Race Data Completeness After Adding DoD Data, FY1997–FY2005

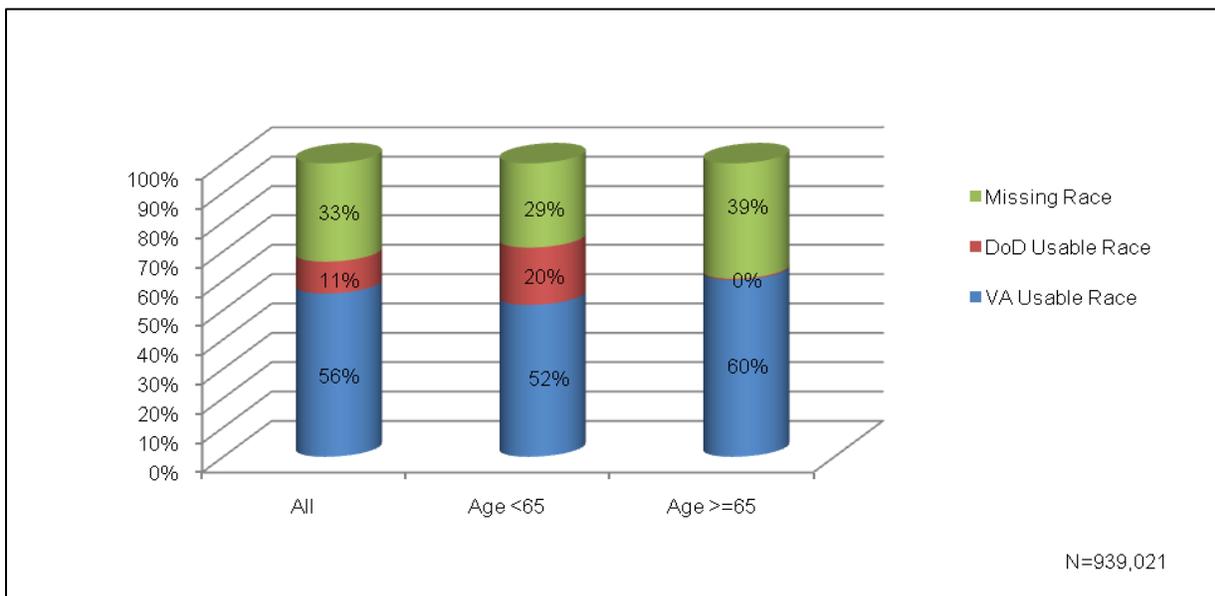
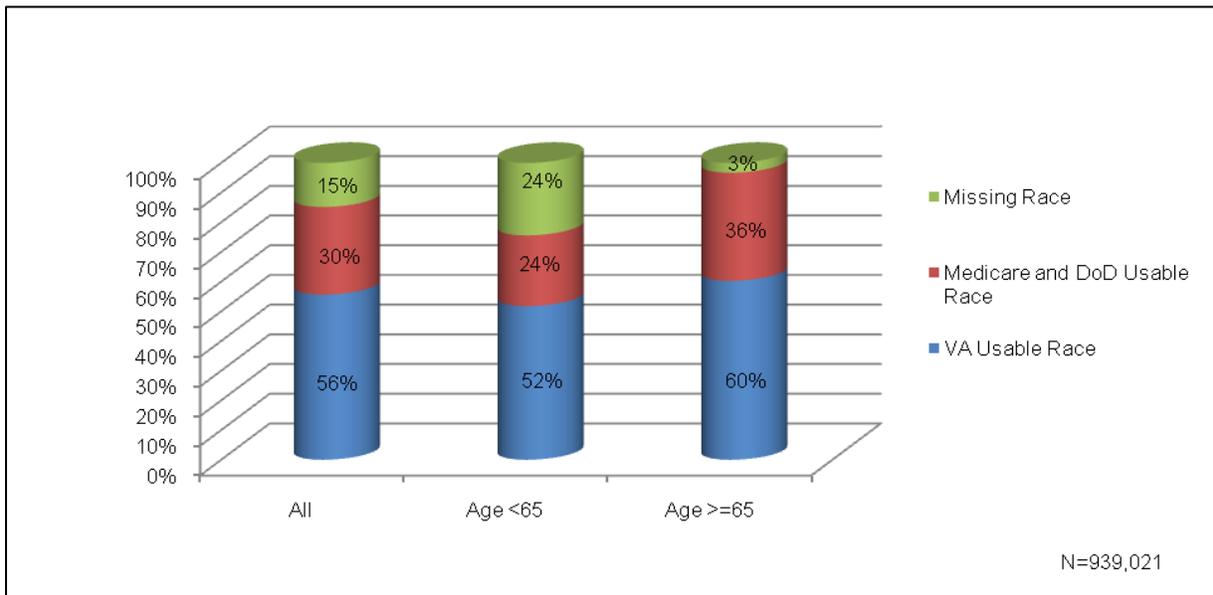


Figure 3c. Changes in Race Data Completeness After Adding Medicare and DoD Data, FY1997–FY2005



Supplementation of VHA Ethnicity Data with Information from Medicare and DoD Files

Among the 421,097 individuals for whom we found no usable ethnicity value in the FY1997–FY2005 VA data (Figure 4), we found a Medicare record match for 182,864 (43%). The match rate was 91% among those aged 65 years and older and 10% among those younger than 65 (not shown). We found a DoD record for 294,414 (70%) of all individuals in both age groups.

When we examined Medicare data for additional ethnicity information for our cohort, we found no appreciable increases in usable value percentages (Figures 4 and 5a). Medicare does not distinguish between ethnicity and race, forcing Latino beneficiaries to choose between Hispanic ethnicity and a race category. As a result, increasing the completeness of VA ethnicity data using Medicare data does not necessarily increase the validity of Latino identification in the VA data.

Figure 4. Use of Medicare and Department of Defense Identity Repository Data to Supplement Ethnicity Information in Veterans Health Administration Medical SAS Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Datasets Using the Expanded Definition, FY1997–FY2005

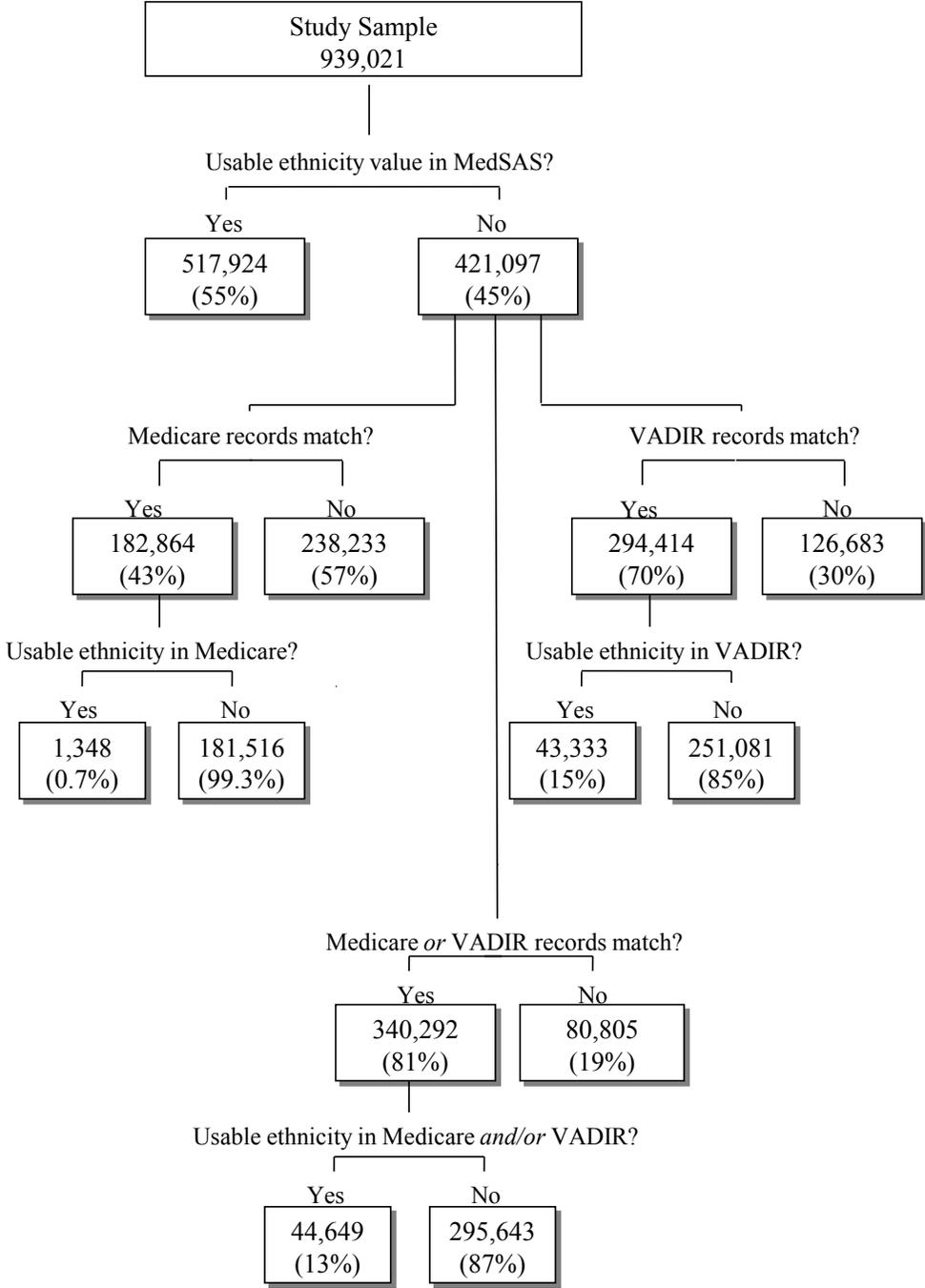
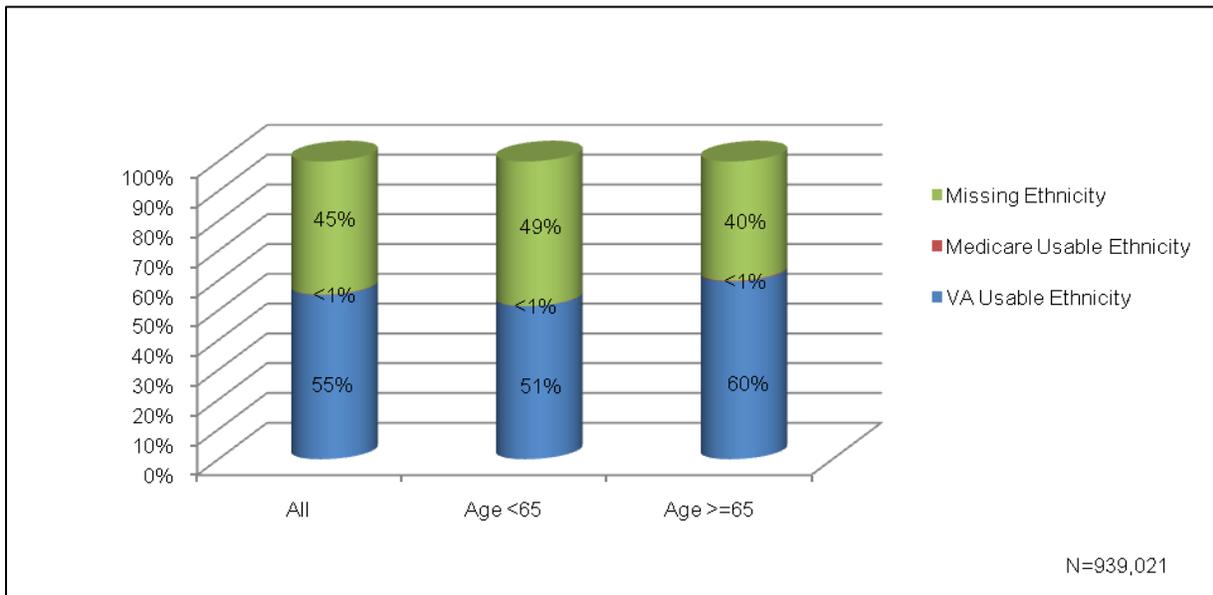
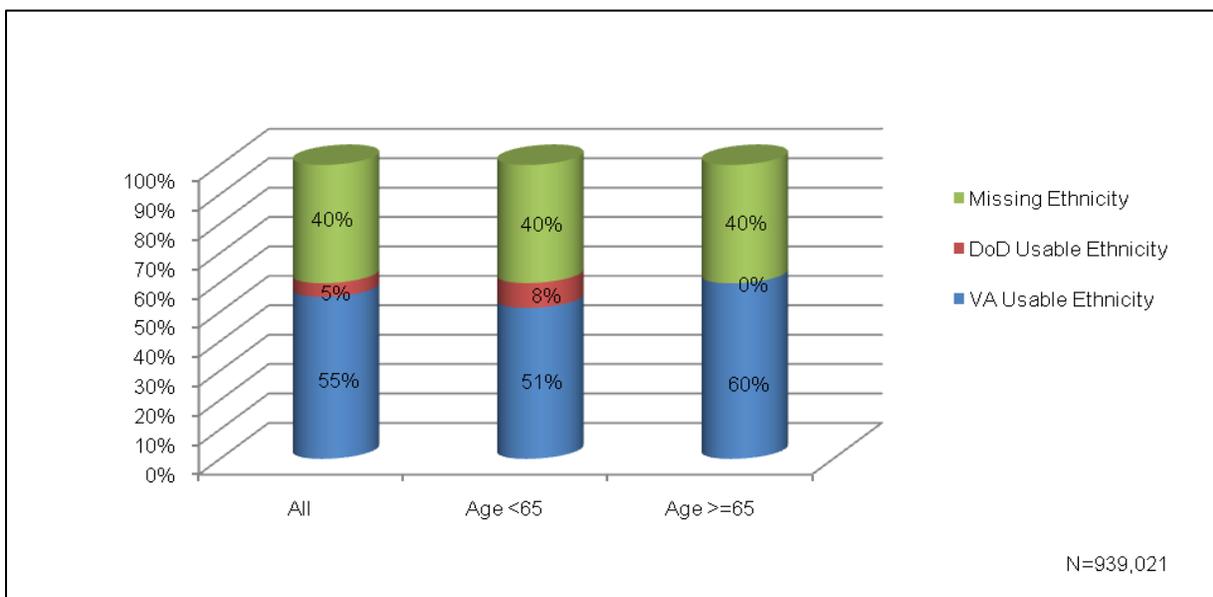


Figure 5a. Changes in Ethnicity Data Completeness After Adding Medicare Data, FY1997–FY2005



DoD data provided a usable ethnicity value for 15% of the matches (Figure 4). The match rate was 25% in the younger age group but less than 1% in the older age group (not shown). Adding DoD data improved the completeness of ethnicity data by more than 8 percentage points among Veterans aged 64 and younger (to 60%) but had no appreciable impact on the completeness of ethnicity data among those age 65 and older (Figure 5b).

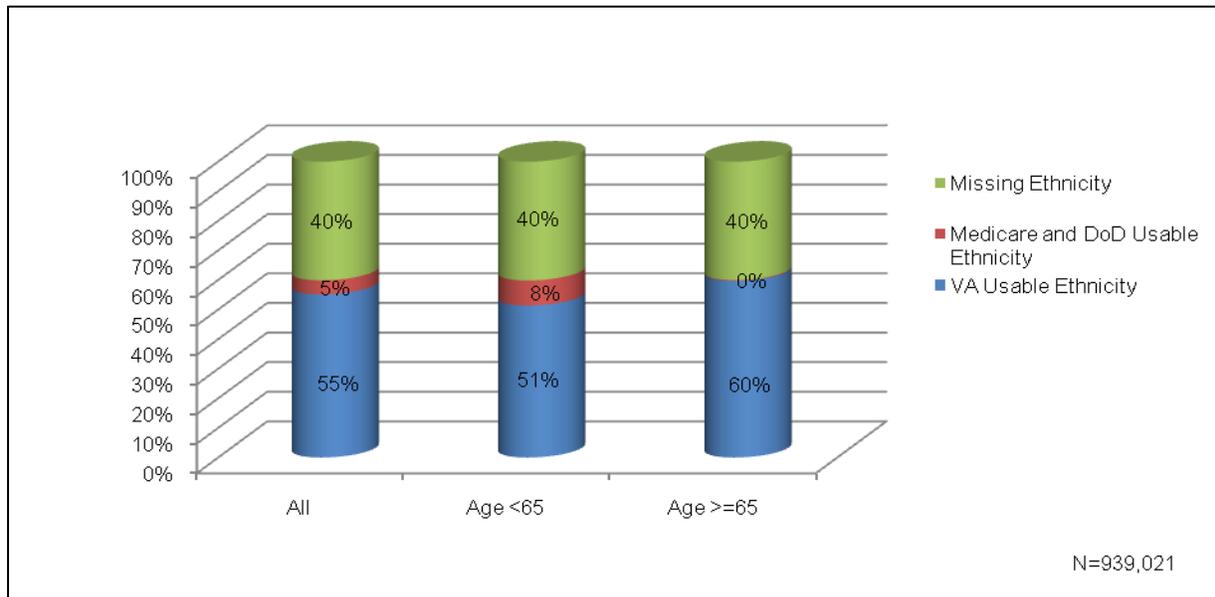
Figure 5b. Changes in Ethnicity Data Completeness After Adding DoD Data to VA Data*



*VA data are from the Medical SAS® Outpatient Visit and Inpatient Acute Care Main Datasets, FY1997-FY2005.

Combining Medicare and DoD data increased ethnicity data completeness to 60% among Veterans of all ages (Figure 5c).

Figure 5c. Changes in Ethnicity Data Completeness After Adding Medicare and DoD Data to VA Data^[*]



*VA data are from the Medical SAS[®] Outpatient Visit and Inpatient Acute Care Main Datasets, FY1997-FY2005.

Aim 4

Aim 4: Assess the likelihood that the data obtained from external sources (Medicare and DoD) will mirror the race and ethnicity information that would have been available in the VHA data if patients had provided this information.

Sample

For Aim 4, we used data from the individuals in our study sample who received VA healthcare between October 1, 2003, and September 30, 2005 (i.e., in FY2004 or FY2005) and who had a usable race value or a usable ethnicity value in VA data (n=275,008). Only values whose method of collection was entered as proxy or self-identification were accepted “self-reported” race and ethnicity. We excluded individuals with two or more different races in the FY2004–FY2005 data.

Race Categories for Cross-Dataset Comparisons.

Because race categories used by VA, Medicare, and DoD differ somewhat, we constructed a classification system for this analysis, shown in Table 11. To create comparable categories, we had to collapse VA’s Asian and Native Hawaiian or Other Pacific Islander categories into one group that we call Asian, Pacific Islander, or Other (APIO).

Table 11. Race Classification Mapping of Department of Veterans Affairs, Medicare, and Department of Defense Data

<i>Department of Veterans Affairs</i>	<i>Department of Defense</i>	<i>Medicare</i>	<i>Classification Constructed for Consistency Analysis</i>
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	

Analysis

We compared the race values in Medicare and DoD records to race values in VA records for Veterans in our sample with a usable VA race record in a linked record dataset. We evaluated race category agreement between VA and Medicare data and between VA and DoD data by calculating sensitivities, specificities, positive predictive values (PPVs), negative predictive values, and kappa statistics. We limited the DoD analysis to Veterans who were younger than age 65.

Results

The characteristics of the sample for the Aim 4 analysis are provided in Table 12.

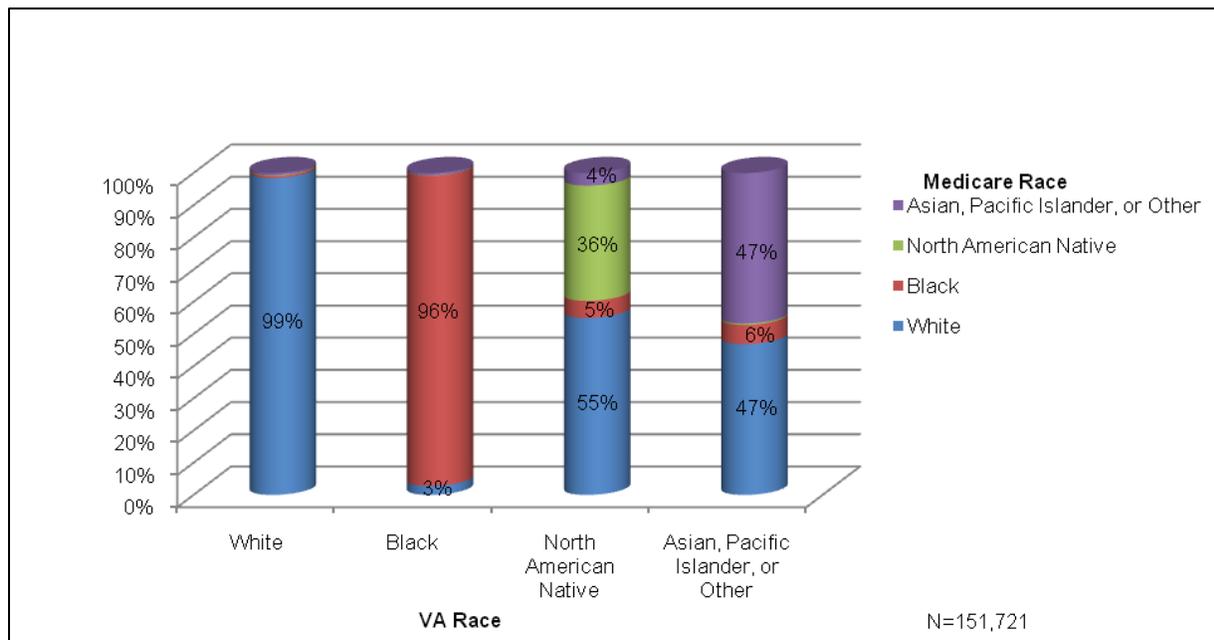
Concordance of Medicare and DoD Data with VA Race Values

Our results showed a high level of concordance between VA and Medicare data for Whites (99%) and African Americans (96%) (Figure 6a). However, only 36% of individuals categorized as North American Native in VA data were listed as North American Native in Medicare data. Only 47% of those listed as Asian, Pacific Islander, or Other (APIO) in VA data were listed as APIO in Medicare data. Most individuals with discordant race information who were listed as North American Native or APIO in the VA data were listed as White in Medicare data.

Table 12. Sample Characteristics, Aim 4 (N=275,008)

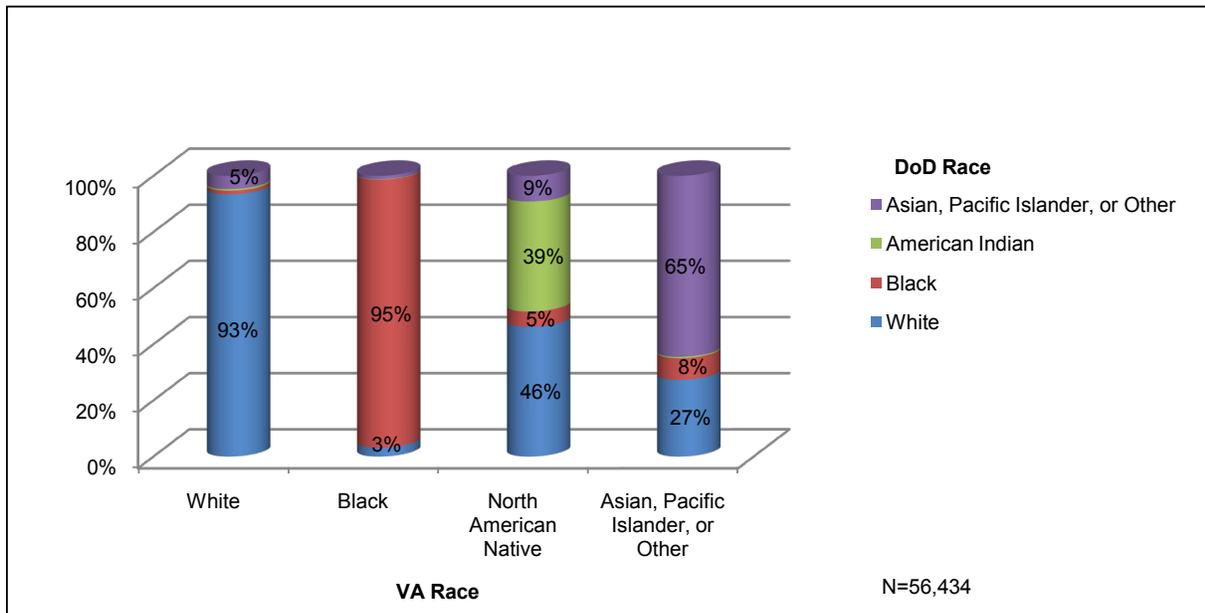
Characteristic	N	(%)
Age on January 1, 2004		
65 years or older	118,134	(43.0)
Gender		
Male	258,702	(94.1)
Marital status		
Married	153,211	(55.7)
Geographic region (based on VISN for region in which patient lives)		
Northeast (VISNs 1–4)	45,088	(16.4)
South (VISNs 5–9, 16, 17)	121,735	(44.3)
Midwest (VISNs 10–12, 15, 23)	61,957	(22.5)
West (VISNs 18–22)	46,228	(16.8)
Period of military service		
Post–Vietnam War and Desert Storm	100,863	(36.7)
Vietnam War	50,414	(18.3)
Korean War (includes service before and after the Korean War)	62,895	(22.9)
World War II	52,070	(18.9)
Other (including World War I and Spanish-American War)	8,766	(3.2)

Figure 6a. Concordance Between VA and Medicare Race Values



We also found a high level of concordance between VA and DoD data for individuals identified as White (93%) or African American (95%) in VA data (Figure 6b). However, just 39% of individuals identified as North American Native in VHA data were listed as North American Native in DoD data and only 65% of APIO individuals according to VHA data were listed as APIO in DoD data. Most North American Natives and APIOs with discordant race information were listed as White in DoD data.

Figure 6b. Concordance Between VA and DoD Race Values



DoD data had less concordance (93%) than Medicare data (99%) with VA data for White individuals, but the two datasets had similar concordance levels for African Americans (95% for Medicare and 96% for DoD data). In contrast, DoD data had better concordance (39%) than Medicare data (36%) with VA data for North American Natives and much better concordance for APIO individuals (65% for DoD, 47% for Medicare data).

Table 13a shows agreement levels between VA and Medicare data, and Table 13b shows agreement levels between VA and DoD data. Sensitivities and PPVs in Medicare and DoD data were high for White and African American individuals. The kappa statistics for these groups indicate high levels of agreement and reflect high specificities and sensitivities. Sensitivities and PPVs for North American Natives and APIOs in Medicare and DoD data were much lower than for Whites and African Americans. The kappa statistics for these two groups indicate only fair agreement.

Table 13a. Accuracy of Medicare Race Data Compared to Self-Reported Veterans Health Administration Medical SAS® Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Race Data, FY2003–2005

				<i>Accuracy and Agreement Measure</i>				
<i>Race</i>	<i>VA</i>	<i>Medicare</i>		<i>Sensitivity</i>	<i>Specificity</i>	<i>Positive Predictive Value</i>	<i>Negative Predictive Value</i>	<i>Kappa</i>
		<i>Yes</i>	<i>No</i>					
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)	(90.9 - 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)	
American Indian	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	1,062	148,541	(44.6 - 48.9)	(99.2 - 99.3)	(46.1 - 50.4)	(99.2 - 99.3)	

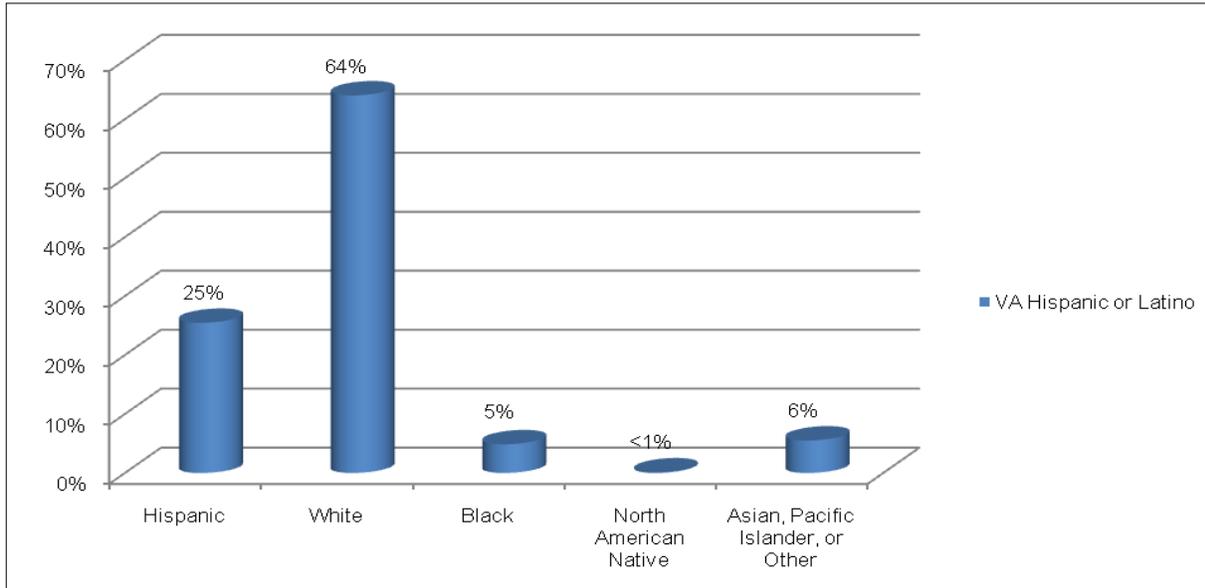
Table 13b. Accuracy of Department of Defense Race Data Compared to Veterans Health Administration Medical SAS® Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Race Data, FY2003–2005

				<i>Accuracy and Agreement Measure</i>				
<i>Race</i>	<i>VAs</i>	<i>DoD</i>		<i>Sensitivity</i>	<i>Specificity</i>	<i>Positive Predictive Value</i>	<i>Negative Predictive Value</i>	<i>Kappa</i>
		<i>Yes</i>	<i>No</i>					
White	Yes	35,187	2,472	93.4	94.2	97.0	87.7	0.86
	No	1,085	17,690	(93.2 - 93.7)	(93.9 - 94.6)	(96.8 - 97.2)	(87.3 - 88.2)	
Black	Yes	16,247	777	95.4	98.5	96.5	98.0	0.94
	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)	(99.5 - 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)	

Concordance of Medicare and DoD Data with VA Ethnicity Values

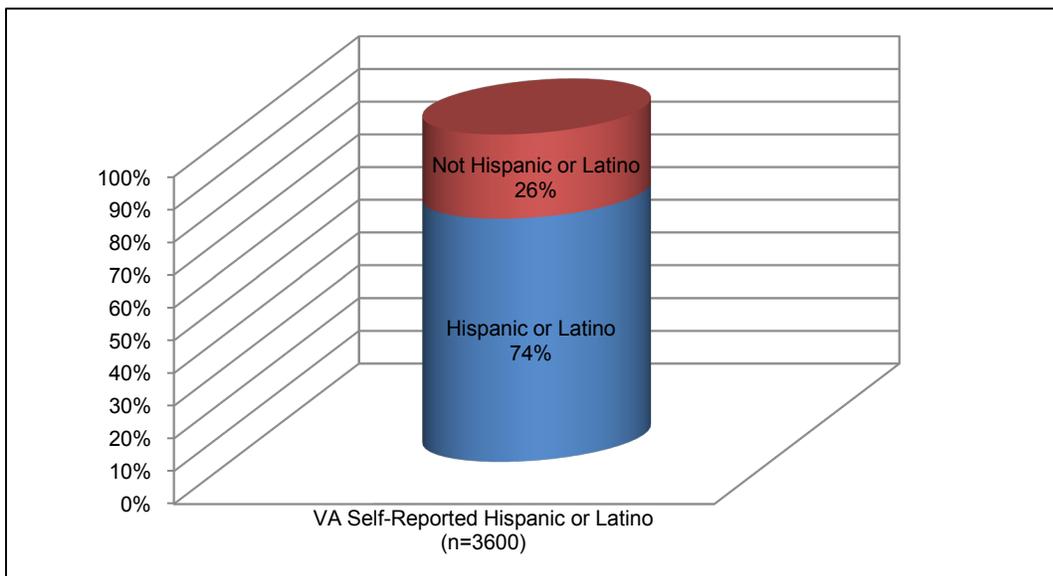
Only 25% of the 5,606 Veterans listed as having Hispanic or Latino ethnicity in the VA datasets who had Medicare records were listed as Hispanic in the Medicare dataset (Figure 7). Most patients listed as Hispanic in VA data were listed as White (64%) in the Medicare data.

Figure 7. Medicare Race Among VA Self-Reported Hispanics



Of the 3,600 Veterans with Hispanic ethnicity according to VA data who had a DoD record, 74.3% were listed as Hispanic in the DoD records (Figure 8).

Figure 8. Agreement of Hispanic Ethnicity in Department of Defense Data with Self-Reported Ethnicity in Veterans Health Administration Medical SAS[®] Outpatient (“Visit”) and Inpatient (Acute Care “Main”) Race Data, FY2003-FY2005



Discussion

In this report, we describe the scope of missing race and ethnicity information in VHA workload data. We compare the characteristics of VHA enrollees whose race is not known to those for whom a usable race value is available. We examine the feasibility and utility of searching additional years of VHA data to fill in missing race values and of supplementing VHA data with race and ethnicity information from Medicare and DoD records. Finally, we examine the likelihood that the data obtained from Medicare and the DoD will mirror the race and ethnicity information that would have been available in the VHA data if patients had provided this information.

Our finding that individuals with missing race and ethnicity data in VHA records were younger and more likely to be female, lack non-VA insurance coverage, and have missing information on marital status is consistent with the results of previous research that used data from the late 1990s.³⁴ However, unlike that earlier study, we found that the differences tended to decrease over time and, in some cases, disappear altogether.

Explanations for the large proportion of missing race and ethnicity values in VHA data are outside the scope of this report but this issue deserves some discussion. The small number of “Declined to Answer” responses in the data suggests that the reason for the high number of missing race values is not primarily that enrollees actively refused to provide the information. Prior to FY2003, missing race values were much more common in outpatient than inpatient records. At that time, the VA did not require VHA outpatient facilities to collect race information and the software for transmitting VistA inpatient data to the AITC blocked records that lacked a race value. VistA-to-AITC data transmission procedures implemented to accommodate the race and ethnicity data that began to be collected in FY2003 had the unintended consequence of blocking the transmission of values in the pre-FY2003 RACE field. This is evident in the marked dip in the percentage of records with a usable race value in FY2003.

Although this decrease was followed by a partial recovery after the problem was partially corrected in FY2004, outpatient records continued to have more missing race values than inpatient records. The reasons for this disparity are unclear. We are not aware, for example, of any incentives that might have motivated inpatient but not outpatient clerical personnel to complete all fields in VistA files. Differences in the settings in which the inpatient admission and outpatient registrations were conducted may have contributed to the disparity. For example, inpatient admission interviews may be held in locations that offer greater privacy than outpatient registrations and, consequently, may be more conducive to collecting race and ethnicity information, which patients might perceive as sensitive. Perhaps the longer duration of inpatient stays allowed more opportunities to collect information not obtained during the initial interview. The finding that Veterans with more outpatient visits were more likely to have usable race and ethnicity values supports this notion.

The problem of incomplete race information in VHA data could be resolved. Several automated consistency checks—prompting the user to fill in required fields when the user leaves them blank, for example - are now built into the VistA enrollment package. If such checks could be

implemented for the race and ethnicity fields, they could dramatically reduce the extent of missing race and ethnicity data in the Medical SAS[®] Datasets. In the absence of automated checks, improved staff training and providing appropriate environments for patient enrollment interviews are potentially useful approaches to reducing the rate of missing VA race and ethnicity data.

We found that race data completeness for single-year cohorts of VHA users could be improved by drawing on information in VA data files from other fiscal years. In general, this strategy was more successful for earlier than the more recent single fiscal year cohorts in our sample, probably because of the larger number of visits—and, thus, more opportunities to fill in missing information—among Veterans who had enrolled earlier than those whose records had been entered into the database more recently. In addition, the VA has recently improved its strategies for collecting race and ethnicity data, increasing the completeness of more recent records.

We also succeeded in increasing the completeness of ethnicity data by using additional years of VA data. However, again, combining data from before and after FY2003 to obtain ethnicity information requires certain assumptions about the validity of the pre-FY2003 data such as the comparability of observer- and self-reported information. Data from FY2003 and later fiscal years can be used with greater confidence to fill in missing ethnicity information than data from before FY2003 because these data are self-reported and collected separately from race data. Completeness of the ETHNIC data field improved dramatically between FY2003 and FY2005, and if the same is true for data from FY2006 and later, these data will be even more useful for filling in missing ethnicity data from earlier years.

Filling in missing race data using Medicare data is very effective for individuals who are older than 65. This is not surprising, given that most Veterans in this age group are enrolled in Medicare and almost all Medicare records have race data. Medicare data also have some utility for increasing the completeness of race data for individuals younger than 65. We found a Medicare record for 10% of the 247,476 individuals under age 65 for whom we had no VA usable race.

The match rate for DoD records was much lower than for Medicare records. When we did find a match, the DoD data were much more useful for filling in missing race data for individuals younger than 65 than for those older than 65. Combining data from Medicare and DoD proved to be a highly effective strategy for filling in missing race data on Veterans of all ages, although it was particularly effective for those older than 65.

Medicare records were not helpful for increasing the completeness of ethnicity data in VA records because Medicare does not have complete ethnicity information. DoD records were useful for increasing the completeness of VHA ethnicity data in those younger than 65, but these records had little value for filling in missing ethnicity among older individuals. Because Medicare data did not improve ethnicity record completeness, using both Medicare and DoD records to fill in missing ethnicity data was no more effective than using DoD records only for this purpose.

Our finding that agreement between data on self-reported North American Natives and APIOs in VA, Medicare, and DoD data was just fair suggests that researchers who use Medicare or DoD race data to supplement VA race data will under-identify members of non-African American minority groups and will misclassify a substantial proportion of these individuals as White. However, these groups represent less than 2% of the merged VA–Medicare sample and 3% of the merged VA–DoD sample. As a result, the likely impact on research results is small, except in studies focusing specifically on those groups.

As in our study, other studies have found that the discordance between data sources is greatest for non-African American minorities and, in most cases, these individuals are misclassified as White [17,34]. Our finding that Medicare data have much lower sensitivity and PPV for North American Natives and APIOs than for Whites and African Americans is also consistent with the results of other studies of Medicare race data validity (we know of no other studies of DoD race data quality) [35-38]. However, these earlier studies found much higher PPVs for North American Natives than our study. For example, a comparison of Medicare race data to self-reported race data in the Medicare Current Beneficiary Data found a PPV of 69.5 for Medicare North American Native patients, much higher than the PPV of 38.0 in our study [38].

The low PPV in our study is due to the large proportion of individuals identified as North American Native in Medicare data but as another race (primarily White) in VA data. Our findings are also supported by a VA study that compared survey results on self-reported race with race listed in the VA’s electronic health records; this study found a greater than 85% concordance for Whites and African Americans but only a 20% concordance for Native Americans [39].

Several factors might explain the high levels of discordance in our study data on North American Natives and APIOs. First, the VA’s data-entry system does not prevent personnel from entering values that are not self-reported (such as values based on observation by clinic staff members) and we could not verify the source of the information in patient records. A comparison of self-reported to observer-reported race in VA data from before FY2003 found that 58% of self-reported North American Natives were identified as White in observer-reported data [17]. In an effort to improve the quality of its race data, Medicare has been using data from the Indian Health Service to correctly characterize North American Native beneficiaries since 1999, increasing the identification of North American Natives by approximately 68% [32]. It is therefore possible that the race of North American Natives and APIOs is misclassified in VA and not Medicare data.

Because the VA, Medicare, and DoD collected data on race at different periods in Veterans’ lifetimes (and race is a social, rather than biological, construct), an additional explanation for discordance in the data for non-African American minority groups is that individuals’ racial self-identification could have changed over time. Studies show that racial self-identification does change over time among multiracial individuals and especially among Native Americans [41]. Since as many as 30% of Asians and 67% of American Indians marry outside their own racial group (compared to 7% of both Whites and African Americans), [40] a substantial proportion of APIO and American Indian Veterans may be multiracial. So, it is possible that a majority of Veterans of North American Native or APIO heritage reported their race as White to the VA.

Thirdly, the discordance could be at least partially due to changes in the likelihood that patients will identify their race as part of a minority group or to different preferences regarding the revelation of racial or ethnic affiliation in the VA, Medicare, and DoD settings [23,41]. As noted above, Medicare race data are supplemented by data obtained from the Indian Health Service (IHS). In order to obtain IHS benefits, American Indian Veterans must identify themselves as such while there is no such need to do so in VA or DoD registrations. There may be other examples of incentives that drive differences in race identification across institutions. The best approaches to resolve the problems associated with missing race information in VA data need further research. Researchers should evaluate whether Indian Health Service data and data obtained directly from the DoD's Defense Manpower Data Center might help reduce the extent of missing data on race and ethnicity. Studies are also needed on the completeness of DoD data if researchers are to use these data to reduce the extent of missing race and ethnicity data in studies using VHA data.

Caveats and Remaining Questions

We used conservative criteria to match VA and Medicare records based on date of birth, gender, and SSN (Appendix C). We excluded a small percentage of patients with matching SSN but not gender, date of birth, or both in the VA and Medicare files. As a result, we might have excluded a small number of true matches from our analysis. In addition, we did not exclude non-Veterans from our sample. Due to the large sample size, it is unlikely that these exclusions have affected our conclusions.

We combined the Asian, Native Hawaiian, Other Pacific Islander, and Other categories into a single category to ensure that the race categories in the three datasets were compatible. According to previous studies, Medicare race data have much higher sensitivity and PPV for Asians than for Other races, so we may have underestimated the agreement between Medicare (and perhaps DoD) data and VA data for people listed as Asian in VA data [35,36,38].

Recommendations

Many investigators have chosen to exclude data on patients whose race is not known from sample analyses. Our investigation suggests that this approach results in changes in important sample characteristics and therefore has the potential to bias study results, especially if the study is using older data.

The best approach to managing the problem of missing race and ethnicity information in VA datasets may depend on the study. This study demonstrated that potentially useful approaches include searching multiple years of VA data, particularly if more recent data are available, and supplementing VA data with Medicare data, DoD data, or both. Another approach is to create an "unknown" category for the study's race and ethnicity variable, allowing the investigators to analyze data separately on patients whose race or ethnicity is not known [16,42].

Researchers using the VA's VSF to supplement race and ethnicity data should identify patient matches using a combination of date of birth, gender, and SSN. This will maximize the likelihood of identifying the right individuals [43].

Because so many non-African American minorities in VA datasets are listed as White in Medicare and DoD data, researchers who supplement VA data with Medicare or DoD data should use a dichotomous grouping of African Americans and non-African Americans. This will result in higher rates of accurate classifications than adding categories for North American Natives and APIOs.

Researchers studying non-African American minorities might consider using other data sources, such as the Indian Health Service data, or survey data to supplement the VHA data on race. When supplementing missing ethnicity data in VHA datasets using VHA data from additional years, researchers may be more successful using VHA data from before FY2004 rather than Medicare data [17]. Researchers cannot use Medicare data to identify Hispanics accurately or completely. Researchers should also consider using data from DoD and special surveys to supplement the VHA's ethnicity data.

Reference List

- (1) *Research on Health Disparities and Minority Health*. VA Research & Development (Web site online). Available at: <http://www.research.va.gov/about/research-health-disparities.cfm>. Accessed January 3, 2011.
- (2) Banez LL, Terris MK, Aronson WJ, et al. Race and time from diagnosis to radical prostatectomy: does equal access mean equal timely access to the operating room?-- Results from the SEARCH database. *Cancer Epidemiol Biomarkers Prev*. 2009;18(4):1208-1212.
- (3) Hausmann LR, Mor M, Hanusa BH, et al. The effect of patient race on total joint replacement recommendations and utilization in the orthopedic setting. *J Gen Intern Med*. 2010;25(9):982-988.
- (4) Burgess DJ, van RM, Grill J, et al. Presence and Correlates of Racial Disparities in Adherence to Colorectal Cancer Screening Guidelines. *J Gen Intern Med*. 2011 Mar;26(3):251-8.
- (5) Hinojosa MS, Rittman M, Hinojosa R, Rodriguez W. Racial/ethnic variation in recovery of motor function in stroke survivors: role of informal caregivers. *J Rehabil Res Dev*. 2009;46(2):223-232.
- (6) Cannon KT, Sarrazin MV, Rosenthal GE, Curtis AE, Thomas KW, Kaldjian LC. Use of mechanical and noninvasive ventilation in black and white chronic obstructive pulmonary disease patients within the Veterans Administration health care system. *Med Care*. 2009;47(1):129-133.
- (7) Johnson PJ, Carlson KF, Hearst MO. Healthcare disparities for American Indian veterans in the United States: a population-based study. *Med Care*. 2010;48(6):563-569.
- (8) *HSR&D Studies and QUERI Projects: Special Projects*. VA Research & Development (Web site online). Available at: [http://www.hsrdr.research.va.gov/research/dra.cfm#Special_\(Underserved,_High_Risk\)_Populations](http://www.hsrdr.research.va.gov/research/dra.cfm#Special_(Underserved,_High_Risk)_Populations) Accessed January 3, 2011.
- (9) 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):410-416.
- (10) Chakkera HA, O'Hare AM, Johansen KL, et al. Influence of race on kidney transplant outcomes within and outside the Department of Veterans Affairs. *J Am Soc Nephrol*. 2005;16(1):269-277.
- (11) Petersen LA, Wright SM, Peterson ED, Daley J. Impact of race on cardiac care and outcomes in veterans with acute myocardial infarction. *Medical Care*. 2002;40(1 Suppl):I86-I96.

- (12) Rosenheck R, Fontana A. Black and Hispanic Veterans in Intensive VA Treatment Programs for Posttraumatic Stress Disorder. [Article]. *Medical Care*. 2002;40(1):I.
- (13) Volpp KG, Stone R, Lave JR, et al. Is thirty-day hospital mortality really lower for black veterans compared with white veterans? *Health Serv Res*. 2007;42(4):1613-1631.
- (14) FY2003 & FY2004 Medical SAS[®] Datasets Missing Race & Ethnicity Data. *VIReC Data Issues Brief*, Mar.-2004.
- (15) Jia H, Zheng YE, Cowper DC, et al. Race/ethnicity: who is counting what? *J Rehabil Res Dev*. 2006;43(4):475-484.
- (16) 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):128-140.
- (17) Sohn MW, Zhang H, Arnold N, et al. Transition to the new race/ethnicity data collection standards in the Department of Veterans Affairs. *Popul Health Metr*. 2006;4:7.
- (18) U.S. Department of Veterans Affairs. *VHA Directive 2004-041, Implementation of the VA Form 10-10EZ, Revised Application for Health Benefits, and VA Form 10-10EZR, Health Benefits Renewal Form*. Washington, DC: U.S. Department of Veterans Affairs; 2004 Aug 4.
- (19) U.S. Department of Veterans Affairs. *VHA Directive 2006-002, Data Entry Requirements for Administrative Data*. Washington, DC: U.S. Department of Veterans Affairs; 2006 Jan 19.
- (20) U.S. Department of Veterans Affairs. *VHA Directive 2009-021, Data Entry Requirements for Administrative Data*. Washington, DC: U.S. Department of Veterans Affairs; 2009 Apr 13.
- (21) Office of Management and Budget. *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity, Notice of Decision*. 1997 Oct 30. Report No.: 62. Available at: <http://www.whitehouse.gov/omb/fedreg/1997standards.html>.
- (22) U.S. Department of Veterans Affairs. *VHA Directive 2003-027, Capture of Race and Ethnicity Categories*. Washington, DC: U.S. Department of Veterans Affairs; 2003 May 29.
- (23) Citro CF, Cork DL, Norwood JL. Citro CF, Cork DL, Norwood JL, eds. *Counting under adversity*. Washington, DC: National Academies Press; 2004.
- (24) Ford ME, Kelly PA. Conceptualizing and categorizing race and ethnicity in health services research. *Health Serv Res*. 2005;40(5 Pt 2):1658-1675.
- (25) 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):381-393.

- (26) ver Ploeg M, Perrin E. *National Research Council Panel on DHHS Collection of Race and Ethnicity Data. Eliminating health disparities: Measurement and data needs.* Washington, DC: National Academies Press; 2004.
- (27) VA Information Resource Center. *VIReC Research User Guide: VHA Medical SAS[®] Outpatient Datasets.* Edward J. Hines, Jr. VA Hospital, Hines, IL: U.S. Department of Veterans Affairs, VA Health Services Research and Development Service; 2004.
- (28) VA Information Resource Center. *VIReC Research User Guide: VHA Medical SAS[®] Inpatient Datasets.* Edward J. Hines, Jr. VA Hospital, Hines, IL: U.S. Department of Veterans Affairs, VA Health Services Research and Development Service; 2004.
- (29) VA Information Resource Center. *VA Vital Status Files* (Web site online). U.S. Department of Veterans Affairs, VA Health Services Research and Development Service; Available at: <http://www.virec.research.va.gov/DataSourcesName/VitalStatus/VitalStatus.htm>. Accessed February 1, 2011.
- (30) *VHA Cohort.* VA Information Resource Center (Web site online). Available at: <http://www.virec.research.va.gov>. Accessed August 1, 2009.
- (31) 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):712-716.
- (32) McBean AM. *Improving Medicare's Data on Race and Ethnicity.* National Academy of Social Insurance; 2006 Oct. Medicare Brief, No. 15.
- (33) Privacy Act; Systems of Records, Veterans Affairs Department of Defense Identity Repository (VADIR)-VA. *Federal Register*, July-27-2009, 37093.
- (34) Kressin NR, Chang BH, Hendricks A, Kazis LE. Agreement between administrative data and patients' self-reports of race/ethnicity. *Am J Public Health.* 2003;93(10):1734-1739.
- (35) 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):107-116.
- (36) Eicheldinger C, Bonito A. More accurate racial and ethnic codes for Medicare administrative data. *Health Care Financ Rev.* 2008;29(3):27-42.
- (37) McBean AM. *Medicare Race and Ethnicity Data. Prepared for the Study Panel on Sharpening Medicare's Tools to Reduce Racial and Ethnic Disparities National Academy of Social Insurance.* 2004 Dec.
- (38) Waldo DR. Accuracy and Bias of Race/Ethnicity Codes in the Medicare Enrollment Database. *Health Care Financing Review.* 2004;26(2):61-72.

- (39) 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.
- (40) Waters M. Immigration, intermarriage, and the challenges of measuring racial/ethnic identities. *AJPH.* 2000;90(11):1735-1737.
- (41) 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 Dec. Report No.: 79.
- (42) 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):781-796.
- (43) 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):377-391.

Appendix A

Form 10-10EZ, *Department of Veterans Affairs, Application for Health Benefits*, may be viewed on the VA healthcare eligibility website at <http://www.va.gov/healtheligibility/application/>.

OMB Approved No. 2900-0091
Estimated Burden Avg. 45 min.

 Department of Veterans Affairs		APPLICATION FOR HEALTH BENEFITS			
SECTION I - GENERAL INFORMATION					
Federal law provides criminal penalties, including a fine and/or imprisonment for up to 5 years, for concealing a material fact or making a materially false statement. (See 18 U.S.C. 1001)					
1. VETERAN'S NAME (Last, First, Middle Name)		2. OTHER NAMES USED	3. MOTHER'S MAIDEN NAME	4. GENDER <input type="checkbox"/> MALE <input type="checkbox"/> FEMALE	
5. ARE YOU SPANISH, HISPANIC, OR LATINO? <input type="checkbox"/> YES <input type="checkbox"/> NO	6. WHAT IS YOUR RACE? (You may check more than one.) (Information is required for statistical purposes only.) <input type="checkbox"/> AMERICAN INDIAN OR ALASKA NATIVE <input type="checkbox"/> BLACK OR AFRICAN AMERICAN <input type="checkbox"/> ASIAN <input type="checkbox"/> WHITE <input type="checkbox"/> NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER				
7. SOCIAL SECURITY NUMBER	8. VA CLAIM NUMBER		9. DATE OF BIRTH (mm dd/yyyy)		
9A. PLACE OF BIRTH (City and State)			10. RELIGION		
11. PERMANENT ADDRESS (Street)		11A. CITY	11B. STATE	11C. ZIP CODE	
11D. COUNTY	11E. HOME TELEPHONE NUMBER (Include area code)		11F. E-MAIL ADDRESS		
11G. CELLULAR TELEPHONE NUMBER (Include area code)		12. TYPE OF BENEFIT(S) APPLYING FOR (You may check more than one) <input type="checkbox"/> ENROLLMENT/HEALTH SERVICES <input type="checkbox"/> DENTAL			
13. WHICH VA MEDICAL CENTER OR OUTPATIENT CLINIC DO YOU PREFER? (for listing of facilities visit www.va.gov/directory)		14. DO YOU WANT AN APPOINTMENT WITH A VA DOCTOR OR PROVIDER AS SOON AS ONE BECOMES AVAILABLE? <input type="checkbox"/> YES <input type="checkbox"/> NO I am only enrolling in case I need care in the future.			
15. CURRENT MARITAL STATUS (Check one) <input type="checkbox"/> MARRIED <input type="checkbox"/> NEVER MARRIED <input type="checkbox"/> SEPARATED <input type="checkbox"/> WIDOWED <input type="checkbox"/> DIVORCED <input type="checkbox"/> UNKNOWN					
16. NAME, ADDRESS AND RELATIONSHIP OF NEXT OF KIN			16A. NEXT OF KIN'S HOME TELEPHONE NUMBER (Include area code)		
			16B. NEXT OF KIN'S WORK TELEPHONE NUMBER (Include area code)		
17. NAME, ADDRESS AND RELATIONSHIP OF EMERGENCY CONTACT (if different than 16)			17A. EMERGENCY CONTACT'S HOME TELEPHONE NUMBER (Include area code)		
			17B. EMERGENCY CONTACT'S WORK TELEPHONE NUMBER (Include area code)		
SECTION II - INSURANCE INFORMATION (Use a separate sheet for additional information)					
1. ENTER HEALTH INSURANCE COMPANY NAME, ADDRESS AND TELEPHONE NUMBER (include coverage through spouse or other person)					
2. NAME OF POLICY HOLDER		3. POLICY NUMBER	4. GROUP CODE	5. ARE YOU ELIGIBLE FOR MEDICAID? <input type="checkbox"/> YES <input type="checkbox"/> NO	5A. EFFECTIVE DATE (mm dd/yyyy)
6. ARE YOU ENROLLED IN MEDICARE HOSPITAL INSURANCE PART A? <input type="checkbox"/> YES <input type="checkbox"/> NO		6A. EFFECTIVE DATE (mm dd/yyyy)			
7. ARE YOU ENROLLED IN MEDICARE HOSPITAL INSURANCE PART B? <input type="checkbox"/> YES <input type="checkbox"/> NO		7A. EFFECTIVE DATE (mm dd/yyyy)			
8. NAME EXACTLY AS IT APPEARS ON YOUR MEDICARE CARD			9. MEDICARE CLAIM NUMBER		

VA FORM
FEB 2011 **10-10EZ**

PREVIOUS EDITIONS OF THIS FORM ARE NOT TO BE USED

PAGE 1

Appendix B

Race and Ethnicity Codes Stored in Medical SAS[®] Datasets Before and After Implementation of VHA Directive 2003-027

Before Directive

Table B1. RACE Variable Values

Value	Description	Usable*
(blank)	Missing	N
1	Hispanic, White	Y
2	Hispanic, Black	Y
3	American Indian	Y
4	Black	Y
5	Asian	Y
6	White	Y
7	Unknown value	N

*Y: Yes, N: No

After Directive

Table B2. RACE1-RACE7 Variable Values

Value	Description	Usable*
(blank)	Missing	N
3	American Indian Or Alaska Native-Missing	Y
3O	American Indian Or Alaska Native-Observer	Y
3P	American Indian Or Alaska Native-Proxy	Y
3S	American Indian Or Alaska Native-Self-identification	Y
3U	American Indian Or Alaska Native-Unknown By Patient	Y
8	Asian-Missing	Y
8O	Asian-Observer	Y
8P	Asian-Proxy	Y
8S	Asian-Self-identification	Y
8U	Asian-Unknown By Patient	Y
9	Black or African American-Missing	Y
9O	Black or African American-Observer	Y
9P	Black or African American-Proxy	Y
9S	Black or African American-Self-identification	Y
9U	Black or African American-Unknown By Patient	Y
A	Native Hawaiian or Other Pacific Islander-Missing	Y
AO	Native Hawaiian or Other Pacific Islander-Observer	Y
AP	Native Hawaiian or Other Pacific Islander-Proxy	Y
AS	Native Hawaiian or Other Pacific Islander-Self-identification	Y
AU	Native Hawaiian or Other Pacific Islander-Unknown By Patient	Y
B	White-Missing	Y
BO	White-Observer	Y
BP	White-Proxy	Y
BS	White-Self-identification	Y
BU	White-Unknown By Patient	Y
C	Declined to Answer-Missing	N
CO	Declined to Answer-Observer	N
CP	Declined to Answer-Proxy	N
CS	Declined to Answer-Self-identification	N
CU	Declined to Answer-Unknown By Patient	N
D	Unknown-Missing	N
DO	Unknown-Observer	N
DP	Unknown-Proxy	N
DS	Unknown-Self-identification	N
DU	Unknown-Unknown By Patient	N

*Y: Yes, N: No

Table B3. ETHNIC Variable Values

Value	Description	Usable*
(blank)	Missing	N
D	Declined To Answer—Missing	N
DO	Declined To Answer—Observer	N
DP	Declined To Answer—Proxy	N
DS	Declined To Answer—Self-identification	N
DU	Declined To Answer—Unknown by Patient	N
H	Hispanic or Latino—Missing	Y
HO	Hispanic or Latino—Observer	Y
HP	Hispanic or Latino—Proxy	Y
HS	Hispanic or Latino—Self-identification	Y
HU	Hispanic or Latino—Unknown by Patient	Y
N	Not Hispanic or Latino—Missing	Y
NO	Not Hispanic or Latino—Observer	Y
NP	Not Hispanic or Latino—Proxy	Y
NS	Not Hispanic or Latino—Self-identified	Y
NU	Not Hispanic or Latino—Unknown by Patient	Y
U	Unknown—Missing	N
UO	Unknown—Observer	N
UP	Unknown—Proxy	N
US	Unknown—Self-identification	N
UU	Unknown—Unknown by Patient	N

*Y: Yes, N: No

Appendix C

Table C1. ETHNICITY Variable Values (DoD Data)

Value	Description
1	Hispanic
2	U.S./Canadian Indian tribes
3	Other Asian
4	Puerto Rican
5	Filipino
6	Mexican
7	Eskimo
8	Aleut
9	Cuban
D	Indian
E	Melanesian
G	Chinese
H	Guamanian
J	Japanese
K	Korean
L	Polynesian
Q	Other Pacific Islander
S	Latin American with Hispanic descent
V	Vietnamese
W	Micronesian
X	Other

Appendix D

Table D1. Comparison of Study Sample to All VA Users*

Characteristics	Study Sample N=939,021		All VA Users N=9,405,890	
	n	(%)	n	(%)
Age[†]				
< 65 years	499,677	(53.2)	4,999,505	(53.2)
≥65 years	439,344	(46.8)	4,406,385	(46.8)
Mean (SD)	62.1	(17.4)	62.1	(17.4)
Gender				
Male	829,291	(88.3)	8,307,324	(88.3)

* A unique combination of SSN, date of birth, and gender defines an individual. The study sample was drawn from the FY1997-2005 Medical SAS[®] Acute Inpatient Main files and Outpatient Visit[†] files. To examine characteristics of all VA users, we used information from all of the FY1997-2005 Medical SAS[®] Inpatient 'Main' files (acute care, extended care, non-VA care and observation care (not available in FY1997), as well as the outpatient 'Visit' files from the same years.

[†] Age on December 31, 2004.

Appendix E

Table E1. Characteristics of Individuals With and Without a Usable Race Value in VA Data, 1997-2005*

Usable Race Value	1997		1998		1999		2000	
	Yes	No	Yes	No	Yes	No	Yes	No
n	223,106	77,166	239,217	88,833	252,100	90,425	267,607	99,397
Age								
Mean (SD)	66 (15)	60 (18)	66 (15)	60 (18)	65 (15)	60 (18)	65 (14)	61 (18)
≥ 65 years	54	41	54	41	53	43	53	45
Gender %								
Male	94	77	94	78	94	79	94	81
Marital Status %								
Married	51	41	52	44	53	47	54	50
Unmarried	43	25	43	25	44	27	43	27
Missing	6	34	5	31	3	26	3	23
Insurance Coverage %								
Insured	32	20	36	21	41	27	45	33
Uninsured	68	80	64	79	49	73	55	67
Unknown	<1	<1	<1	<1	<1	<1	<1	<1
Geographic Region %								
Northeast	18	19	18	19	18	19	18	19
South	38	32	41	34	38	31	39	32
Midwest	24	24	21	22	23	25	23	24
West	20	25	20	25	20	26	20	25

(Table E1 Continued)

Usable Race Value	2001		2002		2003		2004		2005	
	Yes	No								
n	290,503	117,490	309,624	136,792	318,195	153,599	334,743	154,204	345,787	156,418
Age										
Mean (SD)	65 (14)	62 (18)	65 (14)	63 (17)	64 (14)	62 (17)	63 (14)	62 (17)	63 (15)	61 (17)
≥ 65 years	53	50	52	51	51	50	48	48	46	46
Gender %										
Male	94	84	94	86	94	87	94	87	94	87
Marital Status %										
Married	56	55	57	58	57	59	57	59	57	58
Unmarried	42	26	41	26	41	26	42	27	42	27
Missing	2	19	1	16	2	15	2	15	1	15
Insurance Coverage %										
Insured	48	39	50	44	53	47	54	49	56	51
Uninsured	52	61	50	56	47	52	46	51	44	49
Unknown	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
Geographic Region %										
Northeast	18	19	18	18	17	17	17	16	17	16
South	39	34	39	33	40	34	40	34	41	34
Midwest	23	24	24	25	24	25	24	25	24	25
West	19	24	19	24	19	24	19	25	19	25

Note: SD - standard deviation

*The combination of SSN, date of birth and gender identified the individual. Variable definitions are in Table 4.

Appendix F

Table F1. Results of Match Between Medical SAS[®] and Medicare Records Shown Separately for the Full Study Sample and Individuals Age 65 and Older*

Match Criteria	Study Sample N=416,611		Individuals Age 65 and Older N = 169,135	
	Records	(%)	Records	(%)
Valid Match				
SSN + 3 Parts DOB + Gender	169,467	(40.7)	145,358	(86)
SSN + 2 Parts DOB + Gender	10,904	(2.6)	9,347	(6)
SSN + 3 Parts DOB Only	602	0.0	446	(0.3)
Subtotal	180,973	(43.3)	155,151	(92)
Non- Matching Records				
SSN Match Only	10,923	(2.6)	7,443	(4)
No SSN Match	224,715	(53.9)	6,541	(4)
Subtotal	235,638	(56.5)	13,984	(8)
Total	416,611	(100)	169,135	(100)

*Acceptable matches are shown in the first 3 rows, under "Valid Match." Although we considered Social Security Number (SSN) and 3 parts date of birth (DOB) a valid match, regardless of a match on gender, we show results with and without the gender match.

Table F2. Results of Match Between Medical SAS[®] and DoD Records Shown Separately for the Full Study Sample and Individuals Age 64 and Younger*

Match Criteria	Study Sample N = 416,611		Individuals Age 64 and Younger N = 247,476	
	Records	(%)	Records	(%)
Valid Match				
SSN + 3 Parts DOB + Gender	281,487	(68)	166,872	(67)
SSN + 2 Parts DOB + Gender	8,613	(2)	5,318	(2)
SSN + 3 Parts DOB Only	2,018	0	1,353	(1)
Subtotal	292,118	(70)	173,543	(70)
Non- Matching Records				
SSN Match Only	8,054	(2)	4,556	(2)
No SSN Match	116,439	(28)	69,377	(28)
Subtotal	124,493	(30)	73,933	(30)
Total	416,611	(100)	247,476	(100)

Note: DoD - Department of Defense

*Acceptable matches are shown in the first 3 rows, under "Valid Match." Although we considered Social Security Number (SSN) and 3 parts date of birth (DOB) a valid match, regardless of a match on gender, we show results with and without the gender match.