

## DATA QUALITY UPDATE: RACE

**November 2009**

### Background

Concerns regarding VHA race and ethnicity data quality include missing values and inconsistency across time due to changes in data collection method and in allowable response categories and format.<sup>1,2</sup> Race information from Medicare files has recently been added to the [VA Vital Status Master File \(VSF\)](#)<sup>3</sup>.

VIReC conducted a study of race and ethnicity data quality. The objectives of this study were 1) to estimate the extent race data completeness in the Medical SAS Datasets could be improved by supplementing with Medicare race information and 2) to evaluate the agreement in race information in the two data sources. This update provides a summary of findings from that study.

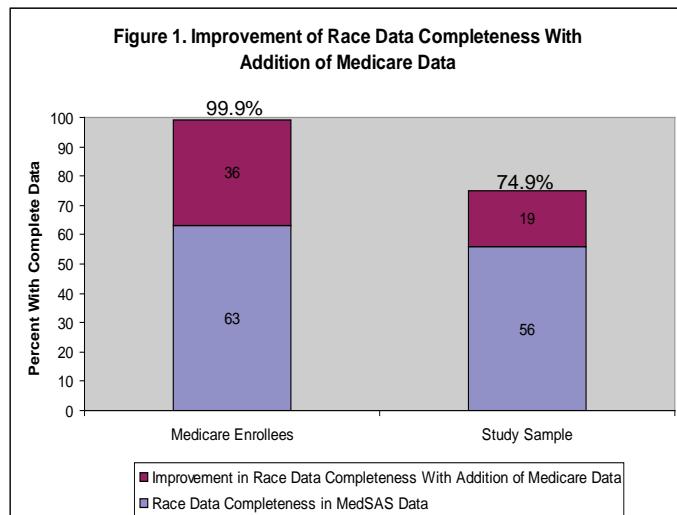
We identified a 10% random sample of patients who used VA acute inpatient or outpatient care during fiscal years 1997 through 2005, comprising 939,021 individuals. The results for the first study objective, along with a brief summary of the methods used, were published in the *Race Data Quality Update* of June, 2008. They are briefly recapped here.

### Improvement in Race Completeness

- Among Medicare enrollees lacking a usable race value in the Medical SAS datasets, nearly 100% had a race value in Medicare data (Figure 1).
- The addition of Medicare data improved race data completeness in the overall study sample of VA patients to 75%.

### Race Consistency

Using 2004 and 2005 VA data for our study sample, we evaluated agreement between Medicare race<sup>4</sup> and VA race/ethnicity, using VA self-reported race and ethnicity as the gold standard.



- Among patients self-identified as White or Black or African American in VA data, 98% and 95%, respectively, had concordant information in Medicare records.
- Medicare race values had positive predictive values (PPV)<sup>5</sup> of 95% and 94%, respectively, for the White and African-American categories.

- Concordance between Medicare and VA data for the Asian (41%) and North American Native (25%) groups were much lower.
- Among patients who reported Hispanic or Latino heritage in VA, 25% were identified as Hispanic<sup>4</sup> in Medicare.
- 64% of both self-identified North American Native and Hispanic VA patients were classified as White in Medicare.

## **Conclusions**

Supplementing VA data with Medicare data substantially improves VA race data completeness. The degree of improvement will vary for each study, depending on the proportion of Medicare enrollees in the study cohort. There are high rates of agreement between VA and Medicare in categorizing African-Americans and Whites but poor agreement in categorizing Asians, North American Natives, and individuals of Hispanic heritage.

## **Recommendations**

Researchers are advised:

- Use of Medicare race information, now more readily accessible to VA researchers in the VA VSF, will reduce the problem of missing race in VA studies using administrative data.
- Although the Medicare White and African-American categories both had high predictive values, given the frequency with which VA North American Natives and Hispanics were classified as White (and Non-Hispanic) in Medicare, the use of a dichotomous race classification of Black/African American and All Other in VA studies using Medicare race information will result in higher rates of accurate classification than other groupings.
- Medicare data cannot be used to identify Hispanics with any degree of accuracy or completeness.
- The VA VS Master File contains a record for each unique combination of SSN, date-of-birth, and gender. When researchers seek race information for individuals in their study cohort from the VSF, they will be most likely to identify the right individuals in the VSF if they use all three elements when conducting the VSF-study cohort record match.

A full technical report containing more details about this ViReC study will be available shortly on the VA Information Resource Center website.

---

<sup>1</sup> US Department of Veterans Affairs VA Information Resource Center (ViReC), *ViReC Research User Guide: VHA Medical SAS® Outpatient Datasets FY2006*. Edward J. Hines, Jr. VA Hospital, Hines, IL, September 2007. Available: <http://www.virec.research.va.gov/References/RUG/RUG-Outpatient06er.pdf>.

<sup>2</sup> Sohn MW, Zhang H, Arnold N, Stroupe K, Taylor BC, Wilt TJ, Hynes DM, "Transition to the new race/ethnicity data collection standards in the Department of Veterans Affairs." *Population Health Metrics*, no. 4 (2006). Available: <http://www.pophealthmetrics.com/content/4/1/7>.

<sup>3</sup> The VA VSF, constructed and maintained by National Data Systems, a division of VHA Office of Information, Health Informatics and Data, combines information from multiple VA databases and is the most complete source of information on veterans' vital status and date of death.

<sup>4</sup> 'Hispanic' is a race category in Medicare. Since only one race may be recorded, individuals identified as White or African-American cannot also be Hispanic.

<sup>5</sup> PPV indicates the proportion of those with a particular race value in Medicare data who were correctly identified (using VA data as the gold standard).