

RACE DATA QUALITY UPDATE

June 2008

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.^{1,2} VIReC is conducting a study of race and ethnicity data quality. The objectives of this study are to characterize the completeness of race and ethnicity data in the VHA Medical SAS datasets and quantify the improvement in race and ethnicity data completeness that can be expected when information from Medicare is added to VHA data. This update provides a summary of findings from that study.

Study Methods

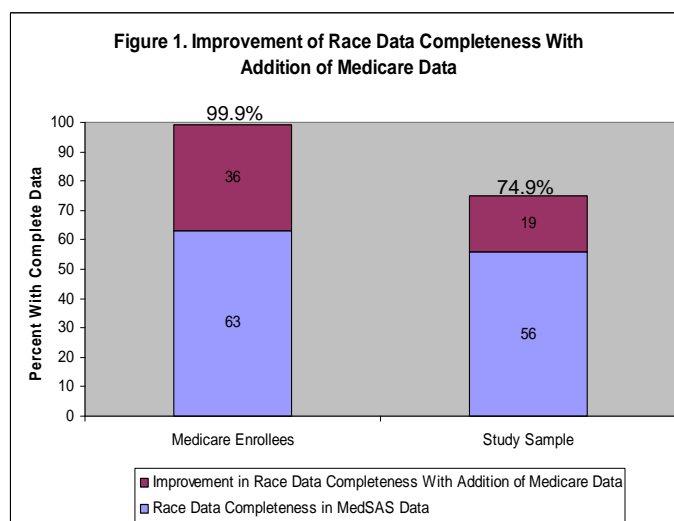
In order to assess the proportion of patients who have a valid value (i.e., not 'unknown' or missing) for race and/or ethnicity in the Medical SAS datasets, we identified a 10% random sample of veterans who used VA acute inpatient or outpatient care during fiscal years 1997 through 2005.^{2,3} We obtained and analyzed all race and ethnicity data elements for this sample. We then matched these records with Medicare records on the same individuals using scrambled SSN, date of birth and gender. We measured the completeness of race information that was achieved after linking race/ethnicity information from Medicare files with VHA data, for those veterans for whom a valid race value in VHA data was missing.

Results

Race

The study sample comprised 939,021 veterans. Overall, 56% had at least one valid race value in the VHA data. Race data completeness in fiscal-year data files ranged from a low of 49% in 2003 to a high of 67% in 2005 and averaged 58% during FY 1997 to FY 2005. Inpatient and outpatient data completeness averaged 90% and 57%, respectively.

We were able to match 52% (490,197) of our study sample with Medicare records. Among these enrollees, the proportion having a valid race value rose from 63% using VHA data alone to nearly 100% when Medicare information was added. (Figure 1)



Of the 416,611 veterans for whom we found no valid race information in the Medical SAS datasets, 43% (180,973) had a race value in the Medicare data. Thus, the addition of Medicare data improved race data completeness in the overall sample to 75%.

Ethnicity

The VA began collecting information on Hispanic ethnicity distinct from race late in FY 2003. The percentage of individuals for whom ethnicity (Hispanic or not) could be ascertained using the new ethnicity variable in MedSAS data in the period FY 2004 to 2005 was 49%. The addition of Medicare data improved ethnicity information completeness minimally.

Conclusions

In our sample, race could not be ascertained from VHA data for 44% of patients. Among Medicare-enrolled VHA users, the addition of Medicare data reduced race data missingness to less than one percent but did not contribute substantially to information on ethnicity.

For research focusing on the VHA senior population, race can be determined for nearly 100% of Medicare enrollees once the data is supplemented. Race data completeness in VHA inpatient files is much better than it is in outpatient files. Investigators should be aware that the number of missing race values in VHA outpatient files may be substantial and pose a problem in particular for those groups for whom supplementing with Medicare data is not an option (i.e. users under 65 who are not disabled and users over 65 not enrolled in Medicare). Even more extensive are the challenges presented by missing ethnicity data.

A technical report containing more details about the study and VA race data quality will be available shortly on the VA Information Resource Center [website](#). In the second part of this study, we will examine the consistency of race and ethnicity values across VA and Medicare records.

¹ US Department of Veterans Affairs VA Information Resource Center (VIREC), "VIREC Exploring Options for Researchers Needing Data on Race." *VIREC Data Issue Brief* (March, 2004) <http://www.virec.research.va.gov/References/DataIssuesBrief/2004/DIB-0403.pdf>.

² VIREC Documentation of VHA Medical SAS Datasets: SAS Proc Contents Outpatient Datasets, Visit (SF) and VHA Medical SAS Datasets: SAS Proc Contents Inpatient Datasets, Acute Care, Main (PM), *US Department of Veterans Affairs VA Information Resource Center (VIREC)*, on the VIREC Website. <http://www.virec.research.va.gov/DataSourcesName/Medical-SAS-Datasets/SASdocumentation.htm>.