Assessing Race and Ethnicity

February 2, 2014

Presented by: Maria K. Mor, PhD

Center for Health Equity Research and Promotion
VA Pittsburgh Healthcare System
Session Outline

- Introduction
- Locating race and ethnicity in VA data
- Locating race and ethnicity in Medicare data
- Quality of VA race/ethnicity data
- Examples of VA studies that have assessed race and ethnicity
- Recommendations to address data quality issues
- Where to go for more help
Audience Poll

• Have you ever used VA Race/Ethnicity Data?
  ▫ Yes
  ▫ No
Session Outline

• Introduction
• Locating race and ethnicity in VA data
• Locating race and ethnicity in Medicare data
• Quality of VA race/ethnicity data
• Examples of VA studies that have assessed race and ethnicity
• Recommendations to address data quality issues
• Where to go for more help
Introduction

- Racial/ethnic disparities in health and health care are well-documented and persistent in the US
  - Root causes and solutions are not well understood
  - While overall quality is improving, access is getting worse and disparities are not changing (AHRQ 2012)
- Racial/ethnic disparities also exist in VHA, where financial barriers to receiving care are minimized
  - Although quality has improved, significant within-facility disparities observed in clinical outcomes (Trivedi 2011)
- More research to detect, understand, and address disparities in health and health care is needed
Introduction

- Accurate race/ethnicity data are essential to disparities research and research on clinical factors associated with race/ethnicity

- Problems with race/ethnicity data in the VA
  - Incomplete
  - Inaccuracies
  - Inconsistent over time
Racial/Ethnic Distribution of Veterans

• Approximately 79% of all Veterans are White
  ▫ 0.7% American Indian or Alaska Native
  ▫ 1.4% Asian
  ▫ 11.1% Black
  ▫ 6.2% Hispanic
  ▫ 1.4% Two or more races

• Use of VA health care differs by race
  ▫ Asian Veterans less likely to use (23.3 %)
  ▫ Black, AIAN, 2+ races more likely to use (>30%)

VA Race and Ethnicity Categories
VHA Handbook 1601A.01 (2009)

- Ethnicity
  - Spanish, Hispanic, or Latino
- Race (>1 may be selected)
  - American Indian or Alaska Native
  - Asian
  - Black or African American
  - Native Hawaiian or Other Pacific Islander
  - White
  - Unknown by Patient
- Current reporting method
  - 2 question format: ethnicity, race
  - Self-reported
## Acquisition of Race/Ethnicity Data in VHA

<table>
<thead>
<tr>
<th>WHO:</th>
<th>WHAT:</th>
<th>WHEN:</th>
<th>WHERE:</th>
<th>HOW:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information</td>
<td>- VA Form 10-10EZ, Application for Health Benefits (online, paper, or</td>
<td>- Enrollment</td>
<td>- Online</td>
<td>- VHA Facility Enrollment Coordinator or designee (e.g., Admission Interview Clerk,</td>
</tr>
<tr>
<td>Source:</td>
<td>by interview)</td>
<td>- Hospital admission</td>
<td>- Telephone call from local VHA facility</td>
<td>Enrollment Specialist) or Outpatient clinic personnel</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Outpatient visit or</td>
<td>- In-person visit to local VHA facility</td>
<td>Collects the information and enters into VistA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pre-registration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient (self-report)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proxy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Session Outline

• Introduction
• Locating race and ethnicity in VA data
• Locating race and ethnicity in Medicare data
• Quality of VA race/ethnicity data
• Examples of VA studies that have assessed race and ethnicity
• Recommendations to address data quality issues
• Where to go for more help
Audience Poll

• What sources of VA race/ethnicity data have you used (check all that apply)?
  ▫ 1 Never Used
  ▫ 2 MedSAS files
  ▫ 3 CDW
  ▫ 4 VistA or regional warehouse
  ▫ 5 Other VA data sources
# Sources of Race and Ethnicity Data in VA

## Medical SAS Datasets

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>MedSAS Dataset</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RACE</td>
<td>Inpatient (PTF Main File)</td>
<td>FY 1970 - present</td>
</tr>
<tr>
<td></td>
<td>Outpatient (Visit File)</td>
<td>FY 1997 - present</td>
</tr>
<tr>
<td></td>
<td>Outpatient (Event File)</td>
<td>FY 1998 - present</td>
</tr>
<tr>
<td>RACE1-RACE6</td>
<td>Inpatient (PTF Main)</td>
<td>FY 2003 - present</td>
</tr>
<tr>
<td>RACE1-RACE7</td>
<td>Outpatient (Visit, Event)</td>
<td>FY 2004 - present</td>
</tr>
<tr>
<td>ETHNIC</td>
<td>Inpatient (PTF Main)</td>
<td>FY 2003 - present</td>
</tr>
<tr>
<td></td>
<td>Outpatient (Visit, Event)</td>
<td>FY 2004 - present</td>
</tr>
</tbody>
</table>
Race/Ethnicity Variables in MedSAS

- Prior to FY 2003
  - Race and ethnicity captured jointly in the variable RACE
  - Single value allowed for race/ethnicity
- After FY 2003
  - Multiple races captured in RACE1-RACE7
  - Single value for ethnicity captured in ETHNIC
  - RACE1-RACE7 and ETHNIC have a length of 2 characters
    - First character has race or ethnicity
    - Second character has method of data collection
    - Common format used for method of data collection
Medical SAS Datasets: Race/Ethnicity Values (Pre 2003)

- **RACE**: Single value for race and ethnicity

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hispanic, white</td>
</tr>
<tr>
<td>2</td>
<td>Hispanic, black</td>
</tr>
<tr>
<td>3</td>
<td>American Indian</td>
</tr>
<tr>
<td>4</td>
<td>Black</td>
</tr>
<tr>
<td>5</td>
<td>Asian</td>
</tr>
<tr>
<td>6</td>
<td>White</td>
</tr>
<tr>
<td>7 or missing</td>
<td>Unknown</td>
</tr>
</tbody>
</table>
Medical SAS Datasets: Race Values (Post 2003)

- **RACE1-RACE7**: Race and method of data collection
- The first character specifies race

<table>
<thead>
<tr>
<th>1st Character</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>American Indian Or Alaska Native</td>
</tr>
<tr>
<td>8</td>
<td>Asian</td>
</tr>
<tr>
<td>9</td>
<td>Black or African American</td>
</tr>
<tr>
<td>A</td>
<td>Native Hawaiian or Other Pacific Islander</td>
</tr>
<tr>
<td>B</td>
<td>White</td>
</tr>
<tr>
<td>C</td>
<td>Declined to Answer</td>
</tr>
<tr>
<td>D</td>
<td>Unknown</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td>Missing</td>
</tr>
</tbody>
</table>
Medical SAS Datasets: Ethnicity Values (Post 2003)

- ETHNIC: Ethnicity and method of data collection
- The first character captures ethnicity

<table>
<thead>
<tr>
<th>1st Character</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>Declined To Answer</td>
</tr>
<tr>
<td>H</td>
<td>Hispanic or Latino</td>
</tr>
<tr>
<td>N</td>
<td>Not Hispanic or Latino</td>
</tr>
<tr>
<td>U</td>
<td>Unknown</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td>Missing</td>
</tr>
</tbody>
</table>
Medical SAS Datasets: Race and Ethnicity Source (Post 2003)

- RACE1-RACE7, ETHNIC
- The second character specifies method of data collection

<table>
<thead>
<tr>
<th>2nd Character</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Missing</td>
</tr>
<tr>
<td>O</td>
<td>Observer</td>
</tr>
<tr>
<td>P</td>
<td>Proxy</td>
</tr>
<tr>
<td>S</td>
<td>Self-identification</td>
</tr>
<tr>
<td>U</td>
<td>Unknown By Patient</td>
</tr>
</tbody>
</table>
Corporate Data Warehouse (CDW)

- National repository of data from VistA Patient File with race and ethnicity data from October 1999 -
- Contains 1 demographic record for each VA station a Veteran has visited
- Contains standard and nonstandard race values
- Stored in a view called PatSub.PatientRace
- Documentation: Best Practices Guide Race Data
Nonstandard Race Values in CDW

- 26 of 31 nonstandard race values can be mapped to 4 standard races

Examples

<table>
<thead>
<tr>
<th>Nonstandard Race</th>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMER INDIAN OR ALASKAN NATIVE, AMERICAN INDIAN, AMERICAN INDIAN / ALASKAN NATIVE</td>
<td>AMERICAN INDIAN OR ALASKA NATIVE</td>
</tr>
<tr>
<td>BLACK; BLACK NOT OF HISP ORIG; BLACK, NON HISPANIC; HISPANIC BLACK</td>
<td>BLACK OR AFRICAN AMERICAN</td>
</tr>
<tr>
<td>WHITE NOT OF HISP ORIG; WHITE, NOT HISPANIC; HISPANIC WHITE; CAUCASIAN;</td>
<td>WHITE</td>
</tr>
<tr>
<td>PACIFIC ISLANDER</td>
<td>NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER</td>
</tr>
</tbody>
</table>
Non-Mapped Values (CDW)

- 5 values are not mapped to standard values
  - Asian or Pacific Islander
  - Asian Pacific Islander
  - Asian/Pacific Islander
  - Mexican American
  - Unknown
- 4.6% of data values fall into 1 of these 5 categories
Multiple Race Values (CDW)

- Approximately 1.7% of patients linked to a standard race have more than 1 standard race
- Not possible to identify the most recent record for a patient
- Recommendation for multiple values
  - Use only self-identified races, if any are recorded
  - Use all recorded races for patients without self-identified race
- Documentation: CDW Race Data and Multiple Races
  [Link](https://vaww.vha.vaco.portal.va.gov/sites/HDI/DQ/WebDQ/PublicFolder/DataQualityAnalysis/CDW%20Race%20Data%20and%20Multiple%20Races.pdf) (VA Intranet only)
Ethnicity (CDW)

- Ethnicity found in two CDW tables
  - PatSub.PatientEthnicity – new method
    - ‘HISPANIC OR LATINO’
    - ‘NOT HISPANIC OR LATINO’
  - PatSub.PatientRace – old method (Appendix A)
    - HISPANIC, WHITE
    - WHITE NOT OF HISP ORIG
    - HISPANIC, BLACK
    - BLACK NOT OF HISP ORIG

- Documentation: CDW Ethnicity Data
  https://vaww.vha.vaco.portal.va.gov/sites/HDI/DQ/WebDQPublicFolder/DataQualityAnalysis/CDW_Ethnicity_Data_Analysis_V5.pdf (VA Intranet only)
Session Outline

• Introduction
• Locating race and ethnicity in VA data
• Locating race and ethnicity in Medicare data
• Quality of VA race/ethnicity data
• Examples of VA studies that have assessed race and ethnicity
• Recommendations to address data quality issues
• Where to go for more help
Sources of Medicare Race Data in VA

- VA Vital Status File
- CMS_RACE
- Race is in Master File only
  - Master File contains one record for each SSN-date of birth (DOB)-gender combination found in VA data
  - Some SSNs have more than one record
- Denominator file from Medicare
- RACE (same as CMS_RACE)
- RTI_RACE
Medicare Race/Ethnicity Data

- Potentially useful source for Veterans in Medicare
  - Age 65 and older (>95% of VA elderly)
  - Disabled (~20% of VA patients <65 years)
  - End stage renal disease
- Derived primarily from Social Security Administration
  - Obtained at the time of application for SSN and/or replacement card
  - Reporting sources: Usually self or family
- Distinctions from current VA race/ethnicity data
  - ‘Hispanic’ is a race category
  - No multiple race reporting
Medicare Race Data from SSA

- Until 1980, 4 categories only
  - White
  - Black
  - Other
  - Unknown
- In 1980, ‘Other’ replaced by
  - Asian, Asian American or Pacific Islander
  - Hispanic
  - American Indian or Alaskan Native
RTI Race in Medicare

- Research Triangle Institute created and implemented an algorithm to increase accuracy of race variable, especially for Hispanic and Asian individuals
- RTI_RACE available in Medicare Denominator File
- Algorithm uses first name, last name, preferred language, place of residence
- Improvement in sensitivity of racial codes
  - Increased from 30% to 77% for Hispanic
  - Increased from 55% to 80% for Asian/Pacific Islander
Medicare Race Data Summary

- Medicare race data quality issues
  - Information on most enrollees (those who obtained SSN prior to 1980) limited to original 4 categories
  - SSN application form – single question format and no multiple race reporting

- Initiatives to improve quality of race/ethnicity data
  - Periodic updates on American Indians and Alaskan Natives from Indian Health Service
  - 1997 survey of enrollees classified as ‘Other’, ‘Unknown’, or with Spanish surname, requesting race/ethnicity self-report
  - RTI Race Algorithm
Session Outline

• Introduction
• Locating race and ethnicity in VA data
• Locating race and ethnicity in Medicare data
• Quality of VA race/ethnicity data
• Examples of VA studies that have assessed race and ethnicity
• Recommendations to address data quality issues
• Where to go for more help
Medical SAS Datasets: Completeness of Race Data

- A substantial portion of patients do not have a “usable” race value* in the VA Medical SAS Inpatient and Outpatient Datasets

<table>
<thead>
<tr>
<th>FY</th>
<th>Usable Race, %</th>
<th>FY</th>
<th>Usable Race, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>57</td>
<td>2005</td>
<td>67</td>
</tr>
<tr>
<td>1998</td>
<td>57</td>
<td>2006</td>
<td>72</td>
</tr>
<tr>
<td>1999</td>
<td>58</td>
<td>2007</td>
<td>75</td>
</tr>
<tr>
<td>2000</td>
<td>58</td>
<td>2008</td>
<td>76</td>
</tr>
<tr>
<td>2001</td>
<td>56</td>
<td>2009</td>
<td>78</td>
</tr>
<tr>
<td>2002</td>
<td>55</td>
<td>2010</td>
<td>80</td>
</tr>
<tr>
<td>2003</td>
<td>49</td>
<td>2011</td>
<td>83</td>
</tr>
<tr>
<td>2004</td>
<td>62</td>
<td>2012</td>
<td>85</td>
</tr>
</tbody>
</table>

* A usable race value is any value that is not ‘missing’ or ‘unknown’ or ‘declined’
Medical SAS Datasets: Completeness by Source

- Visit-level data usability*

<table>
<thead>
<tr>
<th>FY</th>
<th>Race, %</th>
<th>Ethnicity, %</th>
<th>Race, %</th>
<th>Ethnicity, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>67.9</td>
<td>46.3</td>
<td>72.0</td>
<td>75.7</td>
</tr>
<tr>
<td>2008</td>
<td>51.3</td>
<td>37.3</td>
<td>75.3</td>
<td>80.3</td>
</tr>
<tr>
<td>2009</td>
<td>43.6</td>
<td>32.1</td>
<td>78.0</td>
<td>83.6</td>
</tr>
<tr>
<td>2010</td>
<td>40.7</td>
<td>31.4</td>
<td>80.1</td>
<td>86.2</td>
</tr>
<tr>
<td>2011</td>
<td>40.8</td>
<td>31.8</td>
<td>82.3</td>
<td>88.5</td>
</tr>
<tr>
<td>2012</td>
<td>41.7</td>
<td>31.8</td>
<td>84.9</td>
<td>90.9</td>
</tr>
<tr>
<td>2013</td>
<td>41.2</td>
<td>32.2</td>
<td>86.2</td>
<td>92.1</td>
</tr>
<tr>
<td>2014</td>
<td>60.7</td>
<td>46.7</td>
<td>88.5</td>
<td>93.6</td>
</tr>
</tbody>
</table>

* A usable race value is any value that is not ‘missing’ or ‘unknown’ or ‘declined’
Medical SAS Datasets: Completeness of Ethnicity Data

• 90% of visits in FY 2012 have a “usable” ethnicity value* in the VA Medical SAS Inpatient and Outpatient Datasets

• Completeness of ethnicity in the VA Medical SAS Inpatient Datasets is low (32% for FY 2012)
  ▫ Half of inpatient facilities have blank ethnicity data for at least 98% of inpatient records
  ▫ 36% of facilities have blank ethnicity data for all inpatient records
  ▫ Improvements in FY 2014, half of inpatient facilities have usable ethnicity for >40% of patients

  * A usable ethnicity value is any value that is not ‘missing’ or ‘unknown’ or ‘declined’
CDW Completeness of Race Data

- Percent of patients with a standard race in the CDW varies by year of most recent healthcare activity

<table>
<thead>
<tr>
<th>FY</th>
<th>Standard Race, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999*</td>
<td>39.0</td>
</tr>
<tr>
<td>2000</td>
<td>42.6</td>
</tr>
<tr>
<td>2001</td>
<td>43.5</td>
</tr>
<tr>
<td>2002</td>
<td>44.1</td>
</tr>
<tr>
<td>2003</td>
<td>48.2</td>
</tr>
<tr>
<td>2004</td>
<td>53.8</td>
</tr>
<tr>
<td>2005</td>
<td>58.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FY</th>
<th>Standard Race, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>63.0</td>
</tr>
<tr>
<td>2007</td>
<td>65.9</td>
</tr>
<tr>
<td>2008</td>
<td>66.6</td>
</tr>
<tr>
<td>2009</td>
<td>67.2</td>
</tr>
<tr>
<td>2010</td>
<td>68.5</td>
</tr>
<tr>
<td>2011</td>
<td>70.2</td>
</tr>
<tr>
<td>2012</td>
<td>84.6</td>
</tr>
</tbody>
</table>

* No activity after FY 1999
CDW Completeness of Ethnicity Data

• Results
  ▫ 61% of all patients have ethnicity recorded
  ▫ 88% of those with healthcare activity in FY 2012
  ▫ 78% with one standard category are self-identified
  ▫ 1% have conflicting ethnicity categories

• Recommendations
  ▫ Only use ethnicity captured through self-identification, if available
  ▫ Otherwise, use ethnicity captured through new recording method (Patsub.PatientEthnicity)
  ▫ Use older collection methods (Patsub.PatientRace) when no other data are available
CDW Summary

• 8.3 million unique patient records with standard race values
• 2.3 million patient records with nonstandard race values that can be mapped to standard values
• Can contain multiple records per patient if patient visited more than 1 facility
• Sample queries for CDW data contained in Best Practices Guide: Race Data (link provided on slide 18)
• When multiple values are present
  ▫ Use self-identified race and ethnicity
  ▫ Otherwise use new collection methods (not self-identified)
  ▫ Only use older collection methods if no other data
Session Outline

• Introduction
• Locating race and ethnicity in VA data
• Locating race and ethnicity in Medicare data
• Quality of VA race/ethnicity data
• Examples of VA studies that have assessed race and ethnicity
• Recommendations to address data quality issues
• Where to go for more help
Aims

1. To estimate the extent to which missing “usable” race data in VA MedSAS files can be reduced by using non-VA data sources
   a. Medicare
   b. DoD

2. To evaluate the agreement between VA self-reported race data in MedSAS files and
   a. Medicare race data
   b. DoD race data
Use of Medicare and DoD Data for Improving VA Race Data Quality
Stroupe, et al. (2010) *Journal of Rehabilitation Research & Development*

- **Patient cohort**
  - 10% representative sample of VA patients who obtained services during FY 2004-2005
  - \( N = 570,018 \)
- **Medicare race data were obtained from Medicare Vital Status file**
- **DoD race data were obtained from the VA/DoD Identity Repository (VADIR) database for individuals <65 years**
  - VA/DoD data-sharing agreement
  - Self-reported race/ethnicity obtained from service members
Use of Medicare and DoD Data for Improving VA Race Data Quality
Stroupe, et al. (2010) *Journal of Rehabilitation Research & Development*

<table>
<thead>
<tr>
<th>Patients with &amp; without Usable Race Data in VA MedSAS</th>
<th>Usable Race Value (n=275,008)</th>
<th>No Usable Race Value (n=295,010)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, ≥65 years, %</td>
<td>43</td>
<td>45</td>
</tr>
<tr>
<td>Male, %</td>
<td>94</td>
<td>89</td>
</tr>
<tr>
<td>Married, %</td>
<td>56</td>
<td>56</td>
</tr>
<tr>
<td>Geographic Region, %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>South</td>
<td>44</td>
<td>37</td>
</tr>
<tr>
<td>Midwest</td>
<td>23</td>
<td>21</td>
</tr>
<tr>
<td>West</td>
<td>17</td>
<td>25</td>
</tr>
</tbody>
</table>
Aim 1a: Improvement in Race Completeness with Addition of Medicare Data from 2004-2005

Figure 3. Adding Medicare Data Improves Race Data Completeness

- 24.5% Missing Race
- 27.2% Medicare Usable Race
- 48.3% VA Usable Race
- 41.9% Missing Race
- 9.0% Medicare Usable Race
- 49.1% VA Usable Race
- 2.4% Missing Race
- 50.4% Medicare Usable Race
- 47.2% VA Usable Race

* Sample sizes: All 570,018; Age <65 319,756; Age >=65 250,262
Aim 1b: Improvement in Race Completeness with *Addition of Medicare and DoD Data from 2004-2005*

**Figure.** Adding DoD or DoD and Medicare Data Improves Race Data Completeness Among Non-Elderly (N = 319,756)
### Aim 2: Comparison of Medicare with VA and DoD Data, 2004-2005

<table>
<thead>
<tr>
<th>VA Race</th>
<th>Medicare Race</th>
<th>DoD Data</th>
<th>Classification Constructed for Consistency Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>White</td>
<td>White</td>
<td>White</td>
</tr>
<tr>
<td>Black or African American</td>
<td>Black</td>
<td>Black</td>
<td>Black or African American</td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>North American Native</td>
<td>American Indian or Alaska Native</td>
<td>North American Native</td>
</tr>
<tr>
<td>Asian</td>
<td>Asian</td>
<td>Asian or Pacific Islander</td>
<td>Asian, Pacific Islander, or Other</td>
</tr>
<tr>
<td>Native Hawaiian or Other Pacific Islander</td>
<td>Other</td>
<td>Other</td>
<td></td>
</tr>
</tbody>
</table>
Aim 2a: Comparison of VA with Medicare Data, 2004-2005

Figure 7. Concordance Between VA Race Values and Race Values from External Sources

7a. Medicare\(^a\) (N=151,721)

- White: 99%
- Black: 96%
- North American Native: 36% (5% Asian, Pacific Islander, or Other)
- Asian, Pacific Islander, or Other: 47% (6% Black, 47% White)

\(^a\) Data values not shown for percentages of less than 2.5
Aim 2a: Comparison of VA with Medicare Data, 2004-2005

Figure 8. Medicare Race Among VA Self-Reported Hispanics
Aim 2b: Comparison of VA with DoD Data, 2004-2005

Figure 7. Concordance Between VA Race Values and Race Values from External Sources

7b. DoD\textsuperscript{a} (N=56,434)

- **White**: 93%
- **Black**: 95%
- **North American Native**: 39%
- **Asian, Pacific Islander, or Other**: 65%

- **DoD Race**
  - Asian, Pacific Islander, or Other: 5%
  - American Indian: 46%
  - Black: 8%
  - White: 27%

\textsuperscript{a}Data values not shown for percentages less than 2.5
Conclusion

- Supplementing VA with Medicare and DoD data improves VA race data completeness substantially.
- More study is needed to understand poor rates of agreement between VA and external sources in identifying non-African-American minority individuals.
Improving Patient Demographic Information in VA Databases

- T21-funded collaboration between Center for Health Equity Research and Promotion and Veterans Engineering Resource Center

- Patients (n=173) with missing or declined race at VA Pittsburgh Healthcare System were surveyed by telephone regarding their:
  - Comfort with being asked to provide race/ethnicity, address/telephone, and insurance when coming to the VA
  - Preference for providing that information to a clerk or computer kiosk (n=48 seen at site with kiosks)
Comfort in Providing Race/Ethnicity Data

Comfort in Providing Requested Information (N=173)

Level of Comfort*
- Completely Uncomfortable
- Somewhat Uncomfortable
- Neither
- Somewhat Comfortable
- Completely Comfortable

Race/Ethnicity: 86% (5% Completely Uncomfortable, 93% Somewhat Uncomfortable, 85% Neither, 8% Somewhat Comfortable, 4% Completely Comfortable)

Address/Telephone Information Type: 93% (4% Completely Uncomfortable, 93% Somewhat Uncomfortable, 85% Neither, 8% Somewhat Comfortable, 4% Completely Comfortable)

Insurance: 85% (8% Completely Uncomfortable, 93% Somewhat Uncomfortable, 85% Neither, 8% Somewhat Comfortable, 4% Completely Comfortable)

*Data values not shown for percentages less than 4

02/2015
Preference for Collection of Race/Ethnicity Data

Preferred Mode for Updating Information (N=48)

- **Race/Ethnicity**
  - Kiosk: 10%
  - No Preference: 42%
  - Clerk: 44%

- **Address/Insurance**
  - Kiosk: 17%
  - No Preference: 33%
  - Clerk: 46%

Information Type:
- Race/Ethnicity
- Address/Insurance
Session Outline

• Introduction
• Locating race and ethnicity in VA data
• Locating race and ethnicity in Medicare data
• Quality of VA race/ethnicity data
• Examples of VA studies that have assessed race and ethnicity
• Recommendations to address data quality issues
• Where to go for more help
Recommendations

- Supplement inpatient race and ethnicity with outpatient data when using MedSAS files
- When multiple sources of race and ethnicity exist
  - Use self-identified race and ethnicity, if available
  - Use data from the old collection method, only if data from the new collection method are not available
- Use the RACE variable to obtain ethnicity and race collected by the old method (MedSAS)
- Use Sub.PatientRace to obtain ethnicity collected by the old method (CDW)
Recommendations

• Use of Medicare race information will reduce the problem of missing race in VA studies using administrative data

• When using VA VSF, match on date of birth and gender, in addition to (scrambled) SSN
  • Researchers will be most likely to identify the right individuals in the VSF if they use all 3 elements when conducting their VSF-study cohort record match
Recommendations

• Classifying minorities as either Black/African American or Other in VA studies using Medicare race information results in higher rates of accurate classification than other groupings
  ▫ VA North American Natives and Hispanics frequently misclassified as White (and Non-Hispanic) in Medicare
  ▫ Medicare White and African-American categories, both had high predictive values for VA race
Recommendations

- Medicare data cannot be used to identify Hispanics with any degree of accuracy or completeness
- RTI_RACE in the Medicare Denominator file can increase the identification of Hispanics and Asians
- Consider other supplementary data sources
  - Department of Defense
  - Special surveys
Session Outline

• Introduction
• Locating race and ethnicity in VA data
• Locating race and ethnicity in Medicare data
• Quality of VA race/ethnicity data
• Examples of VA studies that have assessed race and ethnicity
• Recommendations to address data quality issues
• Where to go for more help
VIReC Help

• VIReC Webpage- [http://vaww.virec.research.va.gov](http://vaww.virec.research.va.gov) (VA Intranet only)
  ▫ Information on VA data sources and how to access data
  ▫ Documentation on some VA datasets, e.g., MedSAS datasets

• HSRData Listserv
  ▫ Discussion among 1,000+ data stewards, managers, and users
  ▫ Past messages in archive (on intranet)

• VIReC Help Desk [VIReC@va.gov](mailto:VIReC@va.gov)
  ▫ VIReC staff will answer your question and/or direct you to available resources on topics


Selected Recent References on Race/Ethnicity Data


Selected Recent References, cont’d


Selected Recent References, cont’d


Questions?
Next VIReC Database & Methods Seminar

*Extracting Data from the EHR Using CAPRI and VistAWeb*

March 2, 2015

Linda Williams, MD