Assessing Race and Ethnicity

Maria K. Mor, PhD
Center for Health Equity Research and Promotion
VA Pittsburgh Healthcare System
Database & Methods Cyberseminar Series

Informational seminars to help VA researchers understand how to use VA and non-VA data in research and quality improvement

Topics

• Application of VA and non-VA data to research and quality improvement questions
• Limitations of secondary data use
• Resources to support VA data use
**FY ‘17 Database & Methods Schedule**

First Monday of the month*  |  1:00pm-2:00pm ET

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/3/16</td>
<td>Overview of VA Data &amp; Research Uses</td>
</tr>
<tr>
<td>11/7/2016</td>
<td>Requesting Access to VA Data</td>
</tr>
<tr>
<td>12/5/2016</td>
<td>Healthcare Utilization with MedSAS &amp; CDW</td>
</tr>
<tr>
<td>1/9/2017</td>
<td>VA Medicare Data (VA/CMS)</td>
</tr>
<tr>
<td>2/6/2017</td>
<td>Assessing Outpatient Utilization with VA Data</td>
</tr>
<tr>
<td>3/6/2017</td>
<td>Mortality Ascertainment &amp; Cause of Death</td>
</tr>
<tr>
<td>4/3/2017</td>
<td>Assessing Race &amp; Ethnicity</td>
</tr>
<tr>
<td>6/5/2017</td>
<td>Pharmacy Data</td>
</tr>
<tr>
<td>7/10/2017*</td>
<td>CAPRI/VistAWeb for EHR Access</td>
</tr>
<tr>
<td>8/7/2017</td>
<td>Comorbidity Measures Using VA and CMS Data</td>
</tr>
<tr>
<td>8/21/2017</td>
<td>Advanced Topics in Comorbidity Measures</td>
</tr>
<tr>
<td>9/11/2017*</td>
<td>CDW microbiology, lab, &amp; pharmacy domains</td>
</tr>
</tbody>
</table>

*Schedule shifts by one week in event of VA holiday.

Visit VIReC’s [Cyberseminar page](http://www.virec.research.va.gov) for more information and to register for upcoming sessions.

**Select a link to view archived cyberseminars.**

www.virec.research.va.gov
Database & Methods Cyberseminar Series

Assessing Race and Ethnicity

Maria K. Mor, PhD
Center for Health Equity Research and Promotion
VA Pittsburgh Healthcare System
By the end of this session, attendees will be able to:

- Locate race and ethnicity in VA and Medicare data
- Assess the quality of VA race and ethnicity data
- Create SQL code to use race and ethnicity data
Session Outline

• Introduction
• Locating race and ethnicity in VA data
• Locating race and ethnicity in Medicare/Medicaid
• Quality of VA race/ethnicity data
• Examples
• Recommendations to address data quality issues
• Where to go for more help
Session Outline

• Introduction
• Locating race and ethnicity in VA data
• Locating race and ethnicity in Medicare/Medicaid
• Quality of VA race/ethnicity data
• Examples
• Recommendations to address data quality issues
• Where to go for more help
Poll Question #1

I am interested in VA data primarily due to my role as:

a. Principal investigator/Co-PI
b. Research staff (Project coordinator, data manager, programmer)
c. Clinical Staff
d. Operations Staff
e. Other—Please describe via the Q & A function
Poll Question #2:
Have you ever used VA Race/Ethnicity Data?

- Yes
- No
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
  • Most minorities groups experience negative disparities in Access and Quality of care (AHRQ 2016)

• 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 78% of all Veterans are White
  - 0.6% American Indian or Alaska Native
  - 1.6% Asian
  - 11.2% Black
  - 6.6% Hispanic
  - 1.4% Two or more races

- Use of VA health care differs by race
  - Asian Veterans less likely to use (25.4 %)
  - Black, AIAN, 2+ races more likely to use (>36%)

- National Center for Veterans Analysis and Statistics
- 2014 Minority Report
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

• How
  • Patient (self-report)
  • Proxy
  • VHA Enrollment Coordinator or clerk

• When
  • VA Form 10-10EZ Application for Health Benefits (on-line, paper, interview)
  • Inpatient or outpatient visit to VHA facility

• Data are entered directly into VistA
Session Outline

• Introduction

• Locating race and ethnicity in VA data

• Locating race and ethnicity in Medicare/Medicaid

• Quality of VA race/ethnicity data

• Examples

• Recommendations to address data quality issues

• Where to go for more help
Poll Question #3:
What sources of VA race/ethnicity data have you used?
(check all that apply)

- Never used race/ethnicity data
- CDW
- MedSAS files
- VistA or regional warehouse
- 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 1976 - 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>(blank)</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>(blank)</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>(blank)</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 to present
- Contains 1 demographic record for each VA station a Veteran has visited
- Contains standard and nonstandard race values
- Racial data available in 2 views
  - PatSub.PatientRace (newer collection standards)
  - SPatient.SPatient or Patient.Patient (older collection standards)
  - Use both views to obtain all available race data

Reference

*Best Practices Guide: Race Data* (Data Quality Report)

Race Tables in CDW

- PatSub.PatientRace
  - RACE contains patient race
  - COLLECTIONMETHOD contains method of data collection
  - Patient/STA3N level
    - Most recent data available for the patient
    - Multiple records if more than one race identified

- SPatient.SPatient or Patient.Patient
  - RACESID contains the SID for the patient race
  - Link to CDWWork.Dim.Race to map to race
  - Contains race data collected under the old collection methods
  - Does not allow for multiple races
Non-standard Race Values in CDW

26 of 31 non-standard races can be mapped to 4 standard races

Examples

<table>
<thead>
<tr>
<th>Non-standard Race</th>
<th>Standard Race</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-standard values rarely used in PatSub.PatientRace (<1%)
- Standard values rarely used in SPatient.SPatient/Patient.Patient (<1%)
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 fall into 1 of these 5 categories (2012)
Multiple Race Values (CDW)

- Approximately 1.7% of patients linked to a standard race have more than 1 standard race (2013)
- 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

Reference

*CDW Race Data and Multiple Races* (Data Quality Report)

http://vaww.vhadataportal.med.va.gov/Portals/0/DataQualityProgram/Reports/CDW_Race_Data_and_Multiple_Races.pdf  (VA Intranet only)
Ethnicity (CDW)

- Ethnicity found in 3 CDW tables
  - PatSub.PatientEthnicity – new method
    - ‘HISPANIC OR LATINO ‘NOT HISPANIC OR LATINO’
  - PatSub.PatientRace or SPatient.SPatient/Patient.Patient
    - Hispanic race/ethnicity (e.g., HISPANIC, WHITE; HISPANIC, BLACK)
    - Non Hispanic race/ethnicity (e.g., WHITE NOT OF HISP ORIG; BLACK NOT OF HISP ORIG)
    - Not all race/ethnicity values indicate ethnicity (e.g., ASIAN, BLACK)

Reference

*CDW Ethnicity Data (Data Quality Report)*

http://vaww.vhadataportal.med.va.gov/Portals/0/DataQualityProgram/Reports/CDW_Ethnicity_Data.pdf (VA Intranet only)
Session Outline

• Introduction
• Locating race and ethnicity in VA data
• Locating race and ethnicity in Medicare/Medicaid
• Quality of VA race/ethnicity data
• Examples
• Recommendations to address data quality issues
• Where to go for more help
Sources of Medicare/Medicaid Race 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
• Medicaid Personal Summary (Enrollment)
  • EL_RACE_ETHNCY_CD
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, only 4 categories collected
  • 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 (RTI) 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
## Medicaid Race/Ethnicity

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>White</td>
</tr>
<tr>
<td>2</td>
<td>Black or African American</td>
</tr>
<tr>
<td>3</td>
<td>American Indian or Alaskan Native</td>
</tr>
<tr>
<td>4</td>
<td>Asian</td>
</tr>
<tr>
<td>5</td>
<td>Hispanic or Latino – No race information available</td>
</tr>
<tr>
<td>6</td>
<td>Native Hawaiian or Other Pacific Islander</td>
</tr>
<tr>
<td>7</td>
<td>Hispanic or Latino and one or more races</td>
</tr>
<tr>
<td>8</td>
<td>More than one race</td>
</tr>
<tr>
<td>9</td>
<td>Unknown</td>
</tr>
</tbody>
</table>
Medicaid Race/Ethnicity Summary

• Medicaid race/ethnicity variables
  • Summary variable: EL_RACE_ETHNCY_CD
  • Individual variables:
    • ETHNICITY_CODE
    • RACE_CODE_1 – RACE_CODE_5
  • Can identify multiple races and/or race and ethnicity
Medicaid Race/Ethnicity Summary (cont.)

• Data Issues
  • Availability lags behind both VA and Medicare
  • Fewer enrollees than Medicare (~10%)  
    • Data collection changes over time
      • October 1998 many changes/additions
Session Outline

• Introduction
• Locating race and ethnicity in VA data
• Locating race and ethnicity in Medicare/Medicaid
• Quality of VA race/ethnicity data
• Examples
• Recommendations to address data quality issues
• Where to go for more help
Medical SAS Datasets: Completeness by Source

Visit-level data usability*

<table>
<thead>
<tr>
<th>FY</th>
<th>Inpatient Race, %</th>
<th>Inpatient Ethnicity, %</th>
<th>Outpatient Race, %</th>
<th>Outpatient Ethnicity, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>49.4</td>
<td>27.9</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>2005</td>
<td>82.4</td>
<td>51.7</td>
<td>57.8</td>
<td>58.6</td>
</tr>
<tr>
<td>2007</td>
<td>67.9</td>
<td>46.3</td>
<td>72.0</td>
<td>75.7</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>2011</td>
<td>40.8</td>
<td>31.8</td>
<td>82.3</td>
<td>88.5</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>2015</td>
<td>91.9</td>
<td>69.2</td>
<td>91.1</td>
<td>94.8</td>
</tr>
</tbody>
</table>

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

• Prior to FY 2003, < 60% of patients had usable race/ethnicity

• Completeness of data has improved dramatically from about 50% complete in FY 2003 to > 90% complete in FY 2015

• Completeness varies between inpatient and outpatient files
  • Inpatient ethnicity data completely missing for many facilities and FY years
  • Substantial differences in completeness of race also exist
  • Always use both the inpatient and outpatient data to capture race/ethnicity in the MedSAS files

* A usable race 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 Race Data FY 2016

- 91% of Veterans have standard usable race data available from the new collection methods
- 2% only have older race data
- Within each collection method < 1% of those with data have conflicting values
- Unique Veterans with $\geq 1$ outpatient visit (NoncountClinicFlag = ‘N’) in FY 2016
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
CDW Completeness of Ethnicity Data (cont.)

• Recommendations

  • Only use ethnicity captured through self-identification, if available

  • Otherwise, use ethnicity captured through new recording method (Patsub.PatientEthnicity)

  • Use older collection methods (SPatient.SPatient/Patient.Patient or Patsub.PatientRace) when no other data are available
Comparison to Non-VA Data Sources

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 (Medicare and DoD)
2. To evaluate the agreement between VA self-reported race data in MedSAS files and Medicare and DoD race data

Cohort

10% representative sample of VA patients obtaining services during FY 2004-2005 (N=570,018)

Reference

Reduction in Missing Data

- 52% were missing usable race from VA data sources

- Age ≥ 65
  - 53% missing usable VA race data
  - 95% of those with missing VA race data had usable Medicare data

- Age < 65
  - Of the 51% missing usable VA race data:
    - 18% had usable Medicare data
    - 37% had usable DOD data
    - 52% had usable data from Medicare and/or DOD data
Concordance with Non-VA Data Sources

Compared to self-reported VA race/ethnicity data

• Agreement was good (93-99%) for White and African-American for both non-VA data Sources

• Agreement was poor for non-African American minorities (27-55%) for both Medicare and DoD

• Most Hispanics were classified as White (64%) rather than Hispanic (25%) in the Medicare data

• Asian, Pacific Islanders, and other minorities had to be collapsed into one category for comparisons
Session Outline

- Introduction
- Locating race and ethnicity in VA data
- Locating race and ethnicity in Medicare/Medicaid
- Quality of VA race/ethnicity data
- Examples
- Recommendations to address data quality issues
- Where to go for more help
SQL Examples in CDW

- **Getting Started with Using CDW** (cyberseminar archive)
  - Several archived seminars on using SQL to join and manipulate CDW data

- **Race Data Best Practices Guide**
  - Several SQL examples for multiple tasks utilizing race and ethnicity data

- **Researcher’s Notebook: Using SQL to “Sort Out” Race in CDW**

- Connected to server vhacdwaw01.vha.med.va.gov
Example: Patsub.PatientRace

```
SELECT Race, Count(Race) AS Freq
FROM CDWWork.PatSub.PatientRace
GROUP BY Race
ORDER BY Freq DESC;
```

<table>
<thead>
<tr>
<th>Race</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHITE</td>
<td>15828774</td>
</tr>
<tr>
<td>BLACK OR AFRICAN AMERICAN</td>
<td>3417339</td>
</tr>
<tr>
<td>DECLINED TO ANSWER</td>
<td>739573</td>
</tr>
<tr>
<td>UNKNOWN BY PATIENT</td>
<td>638644</td>
</tr>
<tr>
<td>AMERICAN INDIAN OR ALASKA NATIVE</td>
<td>233700</td>
</tr>
<tr>
<td>NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER</td>
<td>232330</td>
</tr>
<tr>
<td>ASIAN</td>
<td>219544</td>
</tr>
<tr>
<td>WHITE NOT OF HISP ORIG</td>
<td>52128</td>
</tr>
<tr>
<td><em>Missing</em></td>
<td>5</td>
</tr>
<tr>
<td><em>Unknown at this time</em></td>
<td>1</td>
</tr>
</tbody>
</table>
Example: Mapping to Standard Race Values

- Create a table that maps between non-standard and standard values
- Code is on p. 10 of Race Data Best Practices Guide
- Additional entries to map to “Unable to Map”
  - “*Unknown at this time*”
  - “*Missing*”
  - “ASIAN/PACIFIC ISLANDER”
- Change mapped categories to match project needs
- See Researcher’s Notebook: Using SQL to “Sort Out” Race in CDW for alternate method for programming standard race values
Example: Race Translation Table

```sql
if OBJECT_ID('tempdb..#RaceTranslationTable') is not null
drop table #RaceTranslationTable
create table #RaceTranslationTable
(InboundRace varchar(50),
StandardRace varchar(50));
insert into #RaceTranslationTable
values('NULL', 'Unable to Map')
insert into #RaceTranslationTable
values('AMER INDIAN OR ALASKAN NATIVE', 'AMERICAN INDIAN OR ALASKA NATIVE')
insert into #RaceTranslationTable
values('AMERICAN INDIAN', 'AMERICAN INDIAN OR ALASKA NATIVE')
insert into #RaceTranslationTable
values('AMERICAN INDIAN / ALASKAN NATIVE', 'AMERICAN INDIAN OR ALASKA NATIVE')
insert into #RaceTranslationTable
values('AMERICAN INDIAN OR ALASKA NATIVE', 'AMERICAN INDIAN OR ALASKA NATIVE')
```

Text ‘NULL’ ≠ null value
Delete table if it already exists
Use # to create temporary tables

*See page 10 of Race Data Best Practices Guide for the remaining code*
Example: Convert to Standard Values

```sql
SELECT b.StandardRace, Count(b.StandardRace) as Freq
FROM CDWWork.PatSub.PatentRace as a left join #RaceTranslationTable as b
ON a.Race=b.InboundRace
GROUP BY b.StandardRace
ORDER BY Freq;
```

<table>
<thead>
<tr>
<th>StandardRace</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unable to Map</td>
<td>6</td>
</tr>
<tr>
<td>ASIAN</td>
<td>219544</td>
</tr>
<tr>
<td>NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER</td>
<td>232330</td>
</tr>
<tr>
<td>AMERICAN INDIAN OR ALASKA NATIVE</td>
<td>233700</td>
</tr>
<tr>
<td>UNKNOWN BY PATIENT</td>
<td>638644</td>
</tr>
<tr>
<td>DECLINED TO ANSWER</td>
<td>739573</td>
</tr>
<tr>
<td>BLACK OR AFRICAN AMERICAN</td>
<td>3417339</td>
</tr>
<tr>
<td>WHITE</td>
<td>15880902</td>
</tr>
</tbody>
</table>
Example: Patsub.PatientEthnicity

```
SELECT Ethnicity, FORMAT(COUNT(Ethnicity), '####,###,###') AS Freq
FROM CDWWork.PatSub.PatientEthnicity
GROUP BY Ethnicity
ORDER BY Freq;
```

Format to show commas

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Missing</em></td>
<td>1</td>
</tr>
<tr>
<td><em>Unknown at this time</em></td>
<td>1</td>
</tr>
<tr>
<td>HISPANIC OR LATINO</td>
<td>1,198,471</td>
</tr>
<tr>
<td>NOT HISPANIC OR LATINO</td>
<td>19,013,031</td>
</tr>
<tr>
<td>DECLINED TO ANSWER</td>
<td>420,184</td>
</tr>
<tr>
<td>UNKNOWN BY PATIENT</td>
<td>833,011</td>
</tr>
</tbody>
</table>
Example: Collection Method

```
SELECT CollectionMethod, FORMAT(COUNT(CollectionMethod), '###,###,###') AS Freq
FROM CDWWork.PatSub.PatientRace
GROUP BY CollectionMethod
ORDER BY COUNT(CollectionMethod);
```

<table>
<thead>
<tr>
<th>CollectionMethod</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Unknown at this time</em></td>
<td>1</td>
</tr>
<tr>
<td><em>Missing</em></td>
<td>3</td>
</tr>
<tr>
<td>PROXY</td>
<td>471</td>
</tr>
<tr>
<td>OBSERVER</td>
<td>1,865</td>
</tr>
<tr>
<td>UNKNOWN</td>
<td>227,838</td>
</tr>
<tr>
<td>SELF IDENTIFICATION</td>
<td>21,137,564</td>
</tr>
</tbody>
</table>

Default Value, rarely changed
Example: Patient.Patient

```sql
SELECT b.Race, FORMAT(COUNT(a.PatientSID), '###,###,###') as Freq
FROM Patient.Patient as a
LEFT JOIN Dim.Race as b
ON a.Sta3n = b.Sta3n AND a.RaceSID = b.RaceSID
GROUP BY b.race
ORDER BY COUNT(a.PatientSID) DESC;
```

COUNT requires variables without Null values

<table>
<thead>
<tr>
<th>Race</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>NULL</td>
<td>30,029,914</td>
</tr>
<tr>
<td>WHITE, NOT OF HISPANIC ORIGIN</td>
<td>4,388,429</td>
</tr>
<tr>
<td>UNKNOWN</td>
<td>1,307,415</td>
</tr>
<tr>
<td>BLACK, NOT OF HISPANIC ORIGIN</td>
<td>885,944</td>
</tr>
<tr>
<td>HISPANIC, WHITE</td>
<td>310,461</td>
</tr>
<tr>
<td>WHITE, NOT OF HISPANIC ORIGIN</td>
<td>225,785</td>
</tr>
<tr>
<td>CAUCASIAN</td>
<td>121,945</td>
</tr>
<tr>
<td>BLACK</td>
<td>85,572</td>
</tr>
</tbody>
</table>
Example: Patient.Patient (Standard Values)

Null Values will not link to #RaceTranslationTable

```
SELECT c.StandardRace, FORMAT(COUNT(a.PatientSID), '####,###,###') as Freq
FROM Patient.Patient as a
LEFT JOIN Dim.Race as b
ON a.Sta3n = b.Sta3n and a.RaceSID = b.RaceSID
LEFT JOIN #RaceTranslationTable as c
ON b.Race = c.InboundRace
GROUP BY c.StandardRace
ORDER BY count(a.PatientSID) DESC;
```
Example: Linking Null Values

```
SELECT c.StandardRace, FORMAT(COUNT(a.PatientSID), '#####') as Freq
FROM Patient.Patient as a left join Dim.Race as b
ON a.Sta3n = b.Sta3n and a.RaceSID = b.RaceSID
LEFT JOIN #RaceTranslationTable as c
ON COALESCE(b.Race,'NULL')=c.InboundRace
GROUP BY c.StandardRace
ORDER BY count(a.PatientSID) DESC;
```

‘NULL’ will link to #RaceTranslationTable
Example: Multiple Sources (Long Format)

```sql
if OBJECT_ID('tempdb..#RandomPatients') is not null drop table #RandomPatients
SELECT TOP 100 PatientSID, Sta3n, RaceSID
   INTO #RandomPatients
   FROM CDWWork.Patient.Patient;

SELECT c.PatientSID, c.Sta3N, c.Race, c.CollectionMethod
FROM #RandomPatients as a INNER JOIN cdwwork.PatSub.PatientRace AS c
ON a.Sta3n=c.Sta3n and a.PatientSID = c.PatientSID
UNION ALL

SELECT d.PatientSID, d.Sta3N, b.Race, NULL as CollectionMethod
FROM #RandomPatients AS d
LEFT JOIN cdwwork.dim.Race AS b
ON d.Sta3n = b.Sta3n and d.RaceSID = b.RaceSID
WHERE b.Race is not Null
ORDER BY 1;
```

Can select different value for CollectionMethod but must have the same # of columns for each table.

Sorts by the 1\textsuperscript{st} column
Session Outline

• Introduction
• Locating race and ethnicity in VA data
• Locating race and ethnicity in Medicare/Medicaid
• Quality of VA race/ethnicity data
• Examples
• Recommendations to address data quality issues
• Where to go for more help
Recommendations: VA Data

• When multiple sources of race and ethnicity exist
  • Use self-identified* race and ethnicity, if available
  • Otherwise, use new collection methods (not self-identified)
  • Use data from the old collection method (< FY 2003), only if data from the new collection method are not available
    • Use SPatient.SPatient/Patient.Patient and Sub.PatientRace to obtain race and ethnicity collected by the old method (CDW)
    • RACE variable contains ethnicity and race from the old method (MedSAS)
  • When using MedSAS obtain race and ethnicity from both the inpatient and outpatient files

*Given lack of variability, consideration of collection method is optional
Recommendations: Non-VA Data

- Use of non-VA race data can reduce missing data
- Carefully consider any potential bias (e.g., age or disability) in the outside data source
- Classifying non-Black minorities as “Other” results in better agreement with other data sources
- Potential supplementary data sources
  - Medicare
  - Department of Defense
  - Medicaid
  - Special Surveys
Recommendations: Medicare

- 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
- 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
Session Outline

• Introduction
• Locating race and ethnicity in VA data
• Locating race and ethnicity in Medicare/Medicaid
• Quality of VA race/ethnicity data
• Examples
• Recommendations to address data quality issues
• Where to go for more help
Race and Ethnicity overview:

http://vaww.virec.research.va.gov/RaceAndEthnicity/Overview.htm

(Intranet only)
### Quick links for VA data resources

<table>
<thead>
<tr>
<th>Resource</th>
<th>URL</th>
<th>Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIReC</td>
<td><a href="http://vaww.virec.research.va.gov/Index.htm">http://vaww.virec.research.va.gov/Index.htm</a></td>
<td>VA Intranet</td>
</tr>
<tr>
<td>CDW</td>
<td><a href="https://vaww.cdw.va.gov/Pages/CDWHome.aspx">https://vaww.cdw.va.gov/Pages/CDWHome.aspx</a></td>
<td>VA Intranet</td>
</tr>
</tbody>
</table>
## VIReC Options for Specific Questions

<table>
<thead>
<tr>
<th>HSRData Listserv</th>
<th>HelpDesk</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Community knowledge sharing</td>
<td>• Individualized support</td>
</tr>
<tr>
<td>• ~1,200 VA data users</td>
<td><a href="mailto:virec@va.gov">virec@va.gov</a></td>
</tr>
<tr>
<td>• Researchers, operations, data stewards, managers</td>
<td>(708) 202-2413</td>
</tr>
</tbody>
</table>

3/2017
Contact information

VA Information Resource Center
Hines VA Hospital
virec@va.gov
708-202-2413

Maria Mor
Maria.Mor@va.gov
Next session:
June 5, 2017
1 pm Eastern

Database & Methods Cyberseminar Series

VA Pharmacy Data

Bonnie Paris, PhD
Data Knowledge Analyst
VA Information Resource Center

Walid Gellad, MD, PhD
VA Pittsburgh Healthcare System
University of Pittsburgh Graduate School of Public Health
Selected Recent References on Race/Ethnicity Data


Selected Recent References on Race/Ethnicity Data


Selected Recent References, cont’d


Selected Recent References, cont’d


