



## Database & Methods Cyberseminar Series

---

# 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

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

Select a link to view archived cyberseminars.

Visit VIReC's [Cyberseminar page](#) for more information and to register for upcoming sessions.

[www.virec.research.va.gov](http://www.virec.research.va.gov)



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



## 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

Variable Name	MedSAS Dataset	Description
RACE	Inpatient (PTF Main File)	FY 1976 - present
	Outpatient (Visit File)	FY 1997 - present
	Outpatient (Event File)	FY 1998 - present
RACE1-RACE6	Inpatient (PTF Main)	FY 2003 - present
RACE1-RACE7	Outpatient (Visit, Event)	FY 2004 - present
ETHNIC	Inpatient (PTF Main)	FY 2003 - present
	Outpatient (Visit, Event)	FY 2004 - present

# 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

Value	Description
1	Hispanic, white
2	Hispanic, black
3	American Indian
4	Black
5	Asian
6	White
7 or missing	Unknown

# Medical SAS Datasets: Race Values (Post-2003)

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

1 <sup>st</sup> Character	Description
3	American Indian Or Alaska Native
8	Asian
9	Black or African American
A	Native Hawaiian or Other Pacific Islander
B	White
C	Declined to Answer
D	Unknown
(blank)	Missing

# Medical SAS Datasets: Ethnicity Values (Post-2003)

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

1 <sup>st</sup> Character	Description
D	Declined To Answer
H	Hispanic or Latino
N	Not Hispanic or Latino
U	Unknown
(blank)	Missing

# Medical SAS Datasets: Race and Ethnicity Source (Post-2003)

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

2 <sup>nd</sup> Character	Description
(blank)	Missing
O	Observer
P	Proxy
S	Self-identification
U	Unknown By Patient

# 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)

[http://vaww.vhadataportal.med.va.gov/Portals/0/DataQualityProgram/Reports/Best\\_Practices\\_Guide\\_Race\\_Data.pdf](http://vaww.vhadataportal.med.va.gov/Portals/0/DataQualityProgram/Reports/Best_Practices_Guide_Race_Data.pdf) (VA Intranet only)

# 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 CDWork.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

Non-standard Race	Standard Race
AMER INDIAN OR ALASKAN NATIVE, AMERICAN INDIAN, AMERICAN INDIAN / ALASKAN NATIVE	AMERICAN INDIAN OR ALASKA NATIVE
BLACK; BLACK NOT OF HISP ORIG; BLACK, NON HISPANIC; HISPANIC BLACK	BLACK OR AFRICAN AMERICAN
WHITE NOT OF HISP ORIG; WHITE, NOT HISPANIC; HISPANIC WHITE; CAUCASIAN;	WHITE
PACIFIC ISLANDER	NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER

- 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.vhadatportal.med.va.gov/Portals/0/DataQualityProgram/Reports/CDW\\_Race\\_Data\\_and\\_Multiple\\_Races.pdf](http://vaww.vhadatportal.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](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

## EL\_RACE\_ETHNCY\_CD

Value	Description
1	White
2	Black or African American
3	American Indian or Alaskan Native
4	Asian
5	Hispanic or Latino – No race information available
6	Native Hawaiian or Other Pacific Islander
7	Hispanic or Latino and one or more races
8	More than one race
9	Unknown

# 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\*

	Inpatient		Outpatient	
FY	Race, %	Ethnicity, %	Race, %	Ethnicity, %
2003	49.4	27.9	----	----
2005	82.4	51.7	57.8	58.6
2007	67.9	46.3	72.0	75.7
2009	43.6	32.1	78.0	83.6
2011	40.8	31.8	82.3	88.5
2013	41.2	32.2	86.2	92.1
2015	91.9	69.2	91.1	94.8

\* 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

FY	Standard Race, %
1999*	39.0
2000	42.6
2001	43.5
2002	44.1
2003	48.2
2004	53.8
2005	58.7

FY	Standard Race, %
2006	63.0
2007	65.9
2008	66.6
2009	67.2
2010	68.5
2011	70.2
2012	84.6

\* 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

---

Stroupe, et al. (2010). Use of Medicare and DoD Data for Improving VA Race Data Quality. *Journal of Rehabilitation Research & Development*.

# Reduction in Missing Data

- 52% were missing usable race from VA data sources
- Age  $\geq$  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)
  - <http://vaww.virec.research.va.gov/CDW/Documentation.htm>
  - Several archived seminars on using SQL to join and manipulate CDW data
- Race Data Best Practices Guide
  - [http://vaww.vhadatportal.med.va.gov/Portals/0/DataQualityProgram/Reports/Best\\_Practices\\_Guide\\_Race\\_Data.pdf](http://vaww.vhadatportal.med.va.gov/Portals/0/DataQualityProgram/Reports/Best_Practices_Guide_Race_Data.pdf) (VA Intranet only)
  - Several SQL examples for multiple tasks utilizing race and ethnicity data
- Researcher's Notebook: Using SQL to "Sort Out" Race in CDW
  - <http://vaww.virec.research.va.gov/Notebook/RNB/RNB6-CDW-SQL-to-Sort-Out-Race-CY16.pdf> (VA intranet only)
- Connected to server `vhacdwa01.vha.med.va.gov`

## Example: PatSub.PatientRace

```

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

```

00 % <

Results Messages

	Race	Freq
1	WHITE	15828774
2	BLACK OR AFRICAN AMERICAN	3417339
3	DECLINED TO ANSWER	739573
4	UNKNOWN BY PATIENT	638644
5	AMERICAN INDIAN OR ALASKA NATIVE	233700
6	NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER	232330
7	ASIAN	219544
8	WHITE NOT OF HISP ORIG	52128
9	*Missing*	5
10	*Unknown at this time*	1

## 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

```
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')
```

Delete table if it already exists

Use # to create temporary tables

Text 'NULL' ≠ null value

\*See page 10 of Race Data Best Practices Guide for the remaining code

## Example: Convert to Standard Values

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

	StandardRace	Freq
1	Unable to Map	6
2	ASIAN	219544
3	NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER	232330
4	AMERICAN INDIAN OR ALASKA NATIVE	233700
5	UNKNOWN BY PATIENT	638644
6	DECLINED TO ANSWER	739573
7	BLACK OR AFRICAN AMERICAN	3417339
8	WHITE	15880902

## Example: PatSub.PatientEthnicity

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

Format to show commas

100 % <

Results Messages

	Ethnicity	Freq
1	*Missing*	1
2	*Unknown at this time*	1
3	HISPANIC OR LATINO	1,198,471
4	NOT HISPANIC OR LATINO	19,013,031
5	DECLINED TO ANSWER	420,184
6	UNKNOWN BY PATIENT	833,011

## Example: Collection Method

```

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

```

100 % <

Results Messages

	CollectionMethod	Freq
1	*Unknown at this time*	1
2	*Missing*	3
3	PROXY	471
4	OBSERVER	1,865
5	UNKNOWN	227,838
6	SELF IDENTIFICATION	21,137,564

← Default Value, rarely changed

## Example: Patient.Patient

```

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

100 % <

Results Messages

	Race	Freq
1	NULL	30,029,914
2	WHITE, NOT OF HISPANIC ORIGIN	4,388,429
3	UNKNOWN	1,307,415
4	BLACK, NOT OF HISPANIC ORIGIN	885,944
5	HISPANIC, WHITE	310,461
6	WHITE, NOT OF HISPANIC ORIGIN	225,785
7	CAUCASIAN	121,945
8	BLACK	85,572

## Example: Patient.Patient (Standard 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 b.Race=c.InboundRace
GROUP BY c.StandardRace
ORDER BY count(a.PatientSID) DESC;

```

Null Values will not link to #RaceTranslationTable

	StandardRace	Freq
1	NULL	30,029,914
2	WHITE	5,197,583
3	Unable to Map	1,353,508
4	BLACK OR AFRICAN AMERICAN	1,102,201
5	AMERICAN INDIAN OR ALASKA NATIVE	39,143
6	UNKNOWN BY PATIENT	195
7	DECLINED TO ANSWER	100

## 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

	StandardRace	Freq
1	Unable to Map	31,383,422
2	WHITE	5,197,583
3	BLACK OR AFRICAN AMERICAN	1,102,201
4	AMERICAN INDIAN OR ALASKA NATIVE	39,143
5	UNKNOWN BY PATIENT	195
6	DECLINED TO ANSWER	100
7	NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER	5

## Example: Multiple Sources (Long Format)

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

```
SELECT c.PatientSID, c.Sta3N, c.Race, c.CollectionMethod
FROM #RandomPatients as a INNER JOIN cdwork.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 cdwork.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<sup>st</sup> 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

**VIREC** INTRANET

Search All VA Web Pages  Search [Open Advanced Search](#)

**VA INFORMATION RESOURCE CENTER (VIREC)**

VIREC Home  
VA/CMS Home  
About Us  
New Users of VA Data  
FAQs  
Acronyms  
HelpDesk

**Race & Ethnicity**

**Overview**

Researchers often use race and ethnicity data in health services research. These data are available from multiple sources within the VA. Researchers should be aware of the quality and completeness of race and ethnicity data elements in each source.

There are five minimum categories for federal data on race: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White. Federal data on ethnicity has two categories: Hispanic or Latino and Not Hispanic or Latino. Additional information including standards for maintaining, collecting, and presenting these data is available from [Statistical Policy Directive No. 15](#) (Office of Business Management and Budget, 2003).

**Medical SAS Datasets**

The following variables in the [VHA Medical SAS Inpatient and Outpatient](#) datasets describe the race and ethnicity of Veterans who have used VA healthcare.

- RACE – Captures race and ethnicity data prior to FY 2003.
- RACE1-RACE7 – Captures multiple races from FY2003 to present.
- ETHNIC – Captures a single value for ethnicity from FY2003 to present.

A substantial portion of Veterans do not have a “usable” race value in the Medical SAS Inpatient and Outpatient Datasets. VIREC’s [Technical Report: VA Race Data Quality](#) (September 2011) is the result of an investigation on Veterans’ race and ethnicity data in the Medical SAS Datasets. This study also assessed the feasibility and utility of approaches to reduce missing race and ethnicity information in VHA data.

**Corporate Data Warehouse (CDW)**

The [VHA Corporate Data Warehouse \(CDW\)](#) is a national data repository. CDW Data are organized into domains. The following tables in the Patient Domain contain race and ethnicity data for Veterans who have used VA healthcare: Dim.CollectionMethod, Dim.Race,

**General Resources**

- Data Access
- Data Sources
- Data Tools
- Data Topics
- Products & Services
- Special Projects

Race and Ethnicity overview:

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

(Intranet only)

## Quick links for VA data resources

*Quick Guide: Resources for Using VA Data*

<http://vaww.virec.research.va.gov/Toolkit/QG-Resources-for-Using-VA-Data.pdf> (VA Intranet)

VIREC: <http://vaww.virec.research.va.gov/Index.htm> (VA Intranet)

VIREC Cyberseminars: <http://www.virec.research.va.gov/Resources/Cyberseminars.asp>

VHA Data Portal: <http://vaww.vhadatportal.med.va.gov/Home.aspx> (VA Intranet)

VINCI: <http://vaww.vinci.med.va.gov/vincicentral/> (VA Intranet)

CDW: <https://vaww.cdw.va.gov/Pages/CDWHome.aspx> (VA Intranet)

# VIReC Options for Specific Questions

## HSRData Listserv

- Community knowledge sharing
- ~1,200 VA data users
- Researchers, operations, data stewards, managers
- Subscribe by visiting <http://vaww.virec.research.va.gov/Support/HSRData-L.htm> (VA Intranet)



## HelpDesk

- Individualized support



[virec@va.gov](mailto:virec@va.gov)

(708) 202-2413

# Contact information

VA Information Resource Center

Hines VA Hospital

[virec@va.gov](mailto:virec@va.gov)

708-202-2413

Maria Mor

[Maria.Mor@va.gov](mailto: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

- AHRQ (Agency for Healthcare Research and Quality) (2016). 2015 National Healthcare Quality and Disparities Report and 5th Anniversary Update on the National Quality Strategy. (Rep. No. AHRQ Publication No. 16-0015). Rockville, MD: Agency for Healthcare Research and Quality.
- Baker DW, Cameron KA, Feinglass J, Thompson, JA, Georgas P, Foster S, et al. (2006). A system for rapidly and accurately collecting patients' race and ethnicity. *Am J Public Health*, 96, 532-537.
- Bertolli J, LeeLisa M, Sullivan PS. (2007). Racial Misidentification of American Indians/Alaska Natives in the HIV/AIDS Reporting Systems of Five States and One Urban Health Jurisdiction, U.S., 1984–2002. *Public Health Reports*, 122, 382-392.
- Blustein J. (1994). The Reliability of Racial Classifications in Hospital Discharge Abstract Data. *American Journal of Public Health*, 84, 1018-1021.
- Boehmer U, Kressin NR, Berlowitz DR, Christiansen CL, Kazis LE, Jones JA. (2002). Self-reported vs administrative race/ethnicity data and study results. *Am J Public Health*, 92, 1471-1472.
- Bonito AJ, Bann C, Eicheldinger C, Carpenter L. Creation of New Race-Ethnicity Codes and Socioeconomic Status (SES) Indicators for Medicare Beneficiaries. Final Report, Sub-Task 2. (Prepared by RTI International for the Centers for Medicare and Medicaid Services through an interagency agreement with the Agency for Healthcare Research and Policy, under Contract No.500-00-0024, Task No. 21) AHRQ Publication No. 08-0029-EF. Rockville, MD, Agency for Healthcare Research and Quality. January 2008
- Brahan D, Bauchner H. (2005). Changes in reporting of race/ethnicity, socioeconomic status, gender, and age over 10 years. *Pediatrics*, 115, e163-e166.
- Clegg LX, Reichman ME, Hankey BF, Miller BA, Lin YD, Johnson NJ, et al. (2007). Quality of race, Hispanic ethnicity, and immigrant status in population-based cancer registry data: implications for health disparity studies. *Cancer Causes Control*, 18, 177-187.

## Selected Recent References on Race/Ethnicity Data

- Eicheldinger C, Bonito A. (2008). More accurate racial and ethnic codes for Medicare administrative data. *Health Care Financ Rev*, 29, 27-42.
- Elliott MN, Fremont A, Morrison PA, Pantoja P, Lurie N. (2008). A new method for estimating race/ethnicity and associated disparities where administrative records lack self-reported race/ethnicity. *Health Serv Res*.
- Ford ME, Kelly PA. (2005). Conceptualizing and categorizing race and ethnicity in health services research. *Health Serv Res*, 40, 1658-1675.
- Friedman DJ, Cohen BB, Averbach AR, Norton JM. (2000). Race/ethnicity and OMB Directive 15: implications for state public health practice. *Am.J Public Health*, 90, 1714-1719.
- Gomez SL, Kelsey JL, Glaser SL, Lee MM, Sidney S. (2005). Inconsistencies between self-reported ethnicity and ethnicity recorded in a health maintenance organization. *Ann Epidemiol*, 15, 71-79.
- Gomez SL, Glaser SL. (2006). Misclassification of race/ethnicity in a population-based cancer registry (United States). *Cancer Causes Control*, 17, 771-781.
- Hahn RA. (1992). The state of federal health statistics on racial and ethnic groups. *JAMA*, 267, 268-271.
- Hahn RA, Stroup DF. (1994). Race and ethnicity in public health surveillance: criteria for the scientific use of social categories. *Public Health Rep*, 109, 7-15.
- Hamilton NS, Edelman D, Weinberger M, Jackson GL. (2009). Concordance between self-reported race/ethnicity and that recorded in a Veteran Affairs electronic medical record. *N C Med J*, 70, 296-300.
- Institute of Medicine. (2003). *Unequal treatment: Confronting racial and ethnic disparities in health care* Washington, DC: National Academies Press.
- Jones CP, Truman BI, Elam-Evans LD, Jones CA, Jones CY, Jiles R, et al. (2008). Using "socially assigned race" to probe white advantages in health status. *Ethn Dis*, 18, 496-504.
- Kashner TM. (1998). Agreement between administrative files and written medical records: a case of the Department of Veterans Affairs. *Med Care*, 36, 1324-1336.

## Selected Recent References, cont'd

- Kramer BJ, Wang M, Hoang T, Harker JO, Finke B, Saliba D. (2006). Identification of American Indian and Alaska Native veterans in administrative data of the Veterans Health Administration and the Indian Health.
- Laws MB, Heckscher RA. (2002). Racial and ethnic identification practices in public health data systems in New England. *Public Health Rep*, 117, 50-61.
- Long JA, Bamba MI, Ling B, Shea JA. (2006). Missing race/ethnicity data in Veterans Health Administration based disparities research: a systematic review. *J Health Care Poor Underserved*. 17(1):128-40. Review.
- Mays VM, Ponce NA, Washington DL, Cochran SD. (2003). Classification of race and ethnicity: implications for public health. *Annu Rev Public Health*, 24, 83-110.
- McAlpine DD, Beebe TJ, Davern M, Call K T. (2007). Agreement between self-reported and administrative race and ethnicity data among Medicaid enrollees in Minnesota. *Health Serv Res*, 42, 2373-2388.
- McBean AM. (2006). Improving Medicare's Data on Race and Ethnicity. National Academy of Social Insurance. Medicare Brief, No. 15.  
Ref Type: Serial (Book, Monograph).
- Morgan RO, Wei II, Virnig BA. (2004). Improving identification of Hispanic males in Medicare: use of surname matching. *Med Care*, 42, 810-816.
- Office of Management and Budget Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity, Notice of Decision (Rep. No. 62).
- Pan CX, Glynn RJ, Mogun H, Choodnovskiy I, Avorn J. (1999). Definition of race and ethnicity in older people in Medicare and Medicaid. *J Am Geriatr Soc*, 47, 730-733.
- Polednak AP. (2001). Agreement in race-ethnicity coding between a hospital discharge database and another database. *Ethn Dis*, 11, 24-29.

## Selected Recent References, cont'd

- Rhoades D. (2005). Racial Misclassification and Disparities in Cardiovascular Disease Among American Indians and Alaska Natives. *Circulation*, 111, 1250-1256.
- Saha S, Freeman M, Toure J, Tippens KM, Weeks C, Ibrahim S. (2008). Racial and ethnic disparities in the VA Health Care System: A Systematic Review. *Journal of General Internal Medicine*, 23, 654-671.
- Sohn M, Zhang H, Arnold N, Stroupe K, Taylor B, Wilt T, et al. (2006). Transition to the new race/ethnicity data collection standards in the Department of Veterans Affairs. *Population Health Metrics*, 4.
- Sondik EJ, Lucas JW, Madans JH, Smith, SS. (2000). Race/ethnicity and the 2000 census: implications for public health. *Am.J Public Health*, 90, 1709-1713.
- Stehr-Green P, Bettles J, Robertson LD. (2002). Effect of racial/ethnic misclassification of American Indians and Alaska Natives on Washington State death certificates, 1989-1997. *American Journal of Public Health*, 92, 443-444.
- Stroupe KT, Tarlov E, Zhang Q, Haywood T, Owens A, Hynes DM. Use of Medicare and DoD data for improving VA race data quality. *Journal of Rehabilitation Research & Development*. 2010;47(8):781-795.
- Sugarman J, Soderberg R, Gordon J, Rivara, FP. (1993). Racial misclassification of American Indians: its effect on injury rates in Oregon, 1989 through 1990. *Am J Public Health*, 83, 681-684.
- Sugarman J, Holliday M, Oss, A, Astorina J, Hui Y. (1996). Improving American Indian cancer data in the Washington State Cancer Registry using linkages with the Indian Health Service and Tribal Records. *Cancer*, 78, 1564-1568.
- The Joint Commission: Advancing Effective Communication, Cultural Competence, and Patient- and Family-Centered Care: A Roadmap for Hospitals. Oakbrook Terrace, IL: The Joint Commission, 2010.
- Thoroughman DA, Frederickson D, Cameron D, Shelby L, Cheek, JE. (2002). Racial misclassification of American Indians in Oklahoma State Surveillance Data for Sexually Transmitted Diseases. *American Journal of Epidemiology*, 155, 1137-1141.

## Selected Recent References, cont'd

- Trivedi AN, Grebla RC, Wright SM, Washington DL. (2011). Despite improved quality of care in the Veterans Affairs health system, racial disparity persists for important clinical outcomes. *Health Affairs*, 30, 707-715.
- US Department of Veterans Affairs (2003). VHA Directive 2003-027, Capture of Race and Ethnicity Categories Washington, DC: US Department of Veterans Affairs.
- US Department of Veterans Affairs (2009). VHA Handbook 1601A.01, Intake Registration Washington, DC: US Department of Veterans Affairs.
- Veterans Health Administration Decision Support Office (2009). National Data Extract Technical Guide Bedford, MA: US Department of Veterans Affairs.
- Wei II, Virnig BA, John DA, Morgan RO. (2006). Using a Spanish surname match to improve identification of Hispanic women in Medicare administrative data. *Health Serv Res*, 41, 1469-1481.