



Database & Methods Cyberseminar Series

Session #7: Assessing Race and Ethnicity in VA Data

April 1, 2019

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

Poll Question #1

What is your role in research and/or quality improvement?

- Investigator, PI, Co-I
- Data manager, analyst, or programmer
- Project coordinator
- Other – please describe via the Q&A function

Poll Question #2

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/Medicaid
- Quality of VA race/ethnicity data
- Examples
- Recommendations to address data quality issues
- Where to go for more help

Racial/ethnic disparities in health and health care persistent in US and in VHA

In US

- Disparities in measures for access and quality vary by racial/ethnic group
- Improved trends with disparities in some measures improving for Blacks, Asians, and Hispanics, **most disparities have not changed or have worsened for other racial/ethnic groups** (AHRQ 2018)

In VHA

- Racial/ethnic disparities persist even though 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

Problems with Race/Ethnicity Data in VA

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

77% White

11.8% Black

7.0% Hispanic

1.7% Asian

1.5% Two or
more races

0.7% American
Indian/Alaska Native

Use of VA health care differs by race and service connection

Asian Veterans less likely to use regardless of service connection

Black, and AIAN races more likely to use VHA services

VA Race and Ethnicity Categories

VHA Handbook 1601A.01 (2009)

Ethnicity	Spanish Hispanic Latino
Race <i>(>1 may be selected)</i>	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 are these data acquired?	Patient (self-report) Proxy VHA Enrollment Coordinator or Clerk
When are these data acquired?	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
- OMOP
- MedSAS files
- DOD (VADIR, DaVINCI)
- Other VA data sources

Race/Ethnicity Variables in MedSAS

Prior to FY2003 (old data collection methods)

- Race and ethnicity captured jointly in the variable RACE
- Single value allowed for race/ethnicity

After FY2003 (new data collection methods)

- 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

Location

- Inpatient: Main (PM) file, 1976-present
- Outpatient: Visit (SF) and Event (SE) files, 1997/1998- present

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
First character specifies race*

1st 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

1st 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

2nd 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 PatSub.PatientRace
 - Race (newer collection standards)
 - LegacyRace (older collection standards)
 - **Use both variables to obtain all available race data**

Patient 3.0 Release Documentation: <http://vaww.virec.research.va.gov/CDW/Factbook/FB-CDW-Patient-Domain.pdf> (VA Intranet only)

CDW Race Table Changes

The structure of the CDW data is subject to periodic changes.

As of January 2018, **none of the available CDW documentation for race and ethnicity match the current data structure.**

Patient 3.0 Domain VIREC Factbook **documents the current** data structure

Changes in the business rules for extraction have also led to some differences in the underlying race data stored in CDW.

CDW documentation may refer to race from older collection methods as being located in other CDW tables.

Patient.Patient or SPatient.SPatient tables	RaceSID contained the SID for the patient race Would link to CDWork.Dim.Race to map to race
Patsub.PatientRace	Currently contains the fields LegacyRace and LegacyRaceSID Previously, all race values were stored in the variable Race but those from older collection methods had a value of Null for CollectionMethod

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

http://vaww.vhadatportal.med.va.gov/Portals/0/DataQualityProgram/Reports/Best_Practices_Guide_Race_Data.pdf (VA Intranet only)

Race Tables in CDW

All race data are contained in PatSub.PatientRace

Data are at the Patient/STA3N level with the most recent data available for the patient

Race	Contains patient race from newer collection methods. Multiple records if more than one race identified.
CollectionMethod	Contains method of data collection for Race
LegacyRace	Contains patient race from the older collection methods <ul style="list-style-type: none"> - Does not allow for multiple races - The same value of LegacyRace will be contained on all records for a single PatientSID if that patient has multiple values of Race recorded. - Most patients have values of “*Missing*”, indicating the presence of no data on LegacyRace.

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 Race (<1%)

Current standard values rarely used in LegacyRace (<1%)

Non-mapped Values in CDW

5 values are not mapped to standard values

4.6% of data fall into 1 of these 5 categories (2012)

Non-mapped values

Asian or Pacific Islander
Asian Pacific Islander
Asian/Pacific Islander
Mexican American
Unknown

As of March 2019

- 17.6% of non-missing LegacyRace fall into 1 of these categories
- 96.7% of these non-mapped values are Unknown
- 3.0% of non-mapped values indicate Asian or Pacific Islander

Multiple Race Values in CDW

- Approximately 1.7% of patients linked to a standard race have **more than 1 standard race** (2013)
- Not possible to identify most recent record for a patient
- Recommendation for multiple values
 - Use only self-identified races (if recorded)
 - Use all recorded races for patients without self-identified race

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 in CDW

Ethnicity data found in 2 CDW tables

PatSub.PatientEthnicity - *new method*

'HISPANIC OR LATINO' / 'NOT HISPANIC OR LATINO'

PatSub.PatientRace (LegacyRace or rarely Race) - *old method*

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)

CDW Ethnicity Data (Data Quality Report)

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

VINCI OMOP Version 5

- VINCI Observational Medical Outcomes Partnership (OMOP) seeks to use a Common Data Model (CDM) to map and standardize data
- Data on Race and Ethnicity are contained in the **OMOPV5.Person table**
- Contains one standard value for Race and Ethnicity for each PERSON_ID
 - OMOPV5MAP.PERSON_SPatient_Spatient will link PERSON_ID to other CDW identifiers
 - See documentation regarding those without PatientICN or other potential linkage issues with patient identifiers
 - Excludes non-veterans, test patients, and possible test patients

Race in OMOP

OMOP CDM follows VA Data Quality Program's "Race Data and Multiple Races Report" and VIREC's Researcher's Notebook "Using SQL to "Sort Out" Race in CDW"

Source data	Source.SPatient_SPatient (now LegacyRace in Patsub.PatientRace) Source.Patsub_PatientRace
Six categories for race	White Black or African American Asian American Indian or Alaska Native Native Hawaiian or other Pacific Islander Unknown

"CDW Race Data and Multiple Races:"

http://vaww.vhadatportal.med.va.gov/Portals/0/DataQualityProgram/Reports/CDW_Race_Data_and_Multiple_Races.pdf

"VIREC 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>

Race Logic in OMOP

1. Identify records as self-report or non-self-report and count distinct values.
2. Select the most frequently occurring self-reported race value.
3. If no self-reported race or counts of self-reported race (not including unknown or null) are equal, then select the most frequent non-self-reported race.
4. If there isn't a most frequent value, then select the race value found on record at the patient's preferred institution.
5. If that is null, then select the value edited most recently as determined by ETLBatchID in the SPatient file.
6. If no most frequent or recent non-null value is available, then the value is "UNKNOWN"

Ethnicity in OMOP

OMOP CDM follows the “OMB Standards for Data on Race and Ethnicity” and the VA Data Quality Program’s “CDW Ethnicity Data Report.”

3 categories for ethnicity	Hispanic or Latino Not Hispanic or Latino Unknown
-----------------------------------	---

OMOP CDM Logic for Ethnicity

OMOP uses only the self-reported information provided under the new collection method, when available

Otherwise Ethnicity is captured from non-self-reported data provided by the new collection method

Ethnicity captured under the old collection methods is used when no data are available from the new recording method

Race in DaVINCI (Joint DoD and VA data)

race_cd	C=White M=Asian or Pacific Islander N=Black R=American Indian or Alaskan native X=Other Z=Unknown
race_ethnic_cd	A=American Indian/Alaskan Native B=Asian or Pacific Islander C=Black, not Hispanic D=white, not Hispanic E=Hispanic X=Other Z=Unknown

<https://vaww.vinci.med.va.gov/VinciCentral/DataSources/Index>

https://www.hsrd.research.va.gov/for_researchers/cyber_seminars/archives/video_archive.cfm?SessionID=244

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** (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

VA Medicare Data

- Denominator file from Medicare
- **RACE** (same as **CMS_RACE**)
- **RTI_RACE**

VA Medicaid Data

- Medicaid Personal Summary (Enrollment)
- **EL_RACE_ETHNCY_CD**

Medicare Race/Ethnicity Data

Potentially useful source of data for Veterans enrolled in Medicare, which generally means they are:

- Age 65 and older (>95% of VA elderly)
- Disabled (~20% of VA patients <65 years)
- Diagnosed with end stage renal disease

Derived primarily from Social Security Administration (SSA)

- 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:



In 1980, 'Other' replaced by:



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

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 data quality

- 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 Variables Summary

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

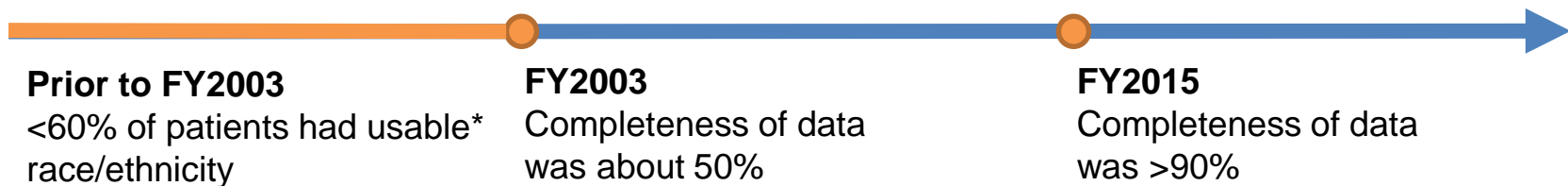
Medicaid Race/Ethnicity Data Issues

- 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 of Race and Ethnicity Data



Completeness varies between inpatient and outpatient files.

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 FY1999

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)

CDW Completeness of Race Data FY2018

New collection methods

92% of Veterans have standard usable race data available from these new methods

Almost **1%** with new data are coded as multiracial

0.3% have conflicting values

Old collection methods

<1% of Veterans only have older race data

1.3% of those have conflicting values

Unique Veterans with ≥ 1 outpatient visit (NoncountClinicFlag = 'N') in FY2018

CDW Completeness of Ethnicity Data

61% of all patients have ethnicity recorded

88% with healthcare activity in FY 2012

78% with one standard category are self-identified

1% have conflicting ethnicity categories

Recommendations for Using CDW Ethnicity Data

1. If available, use ethnicity captured through self-identification
2. Otherwise, use ethnicity captured through new recording method (**Patsub.PatientEthnicity**)
3. Use older collection methods (**Patsub.PatientRace**, **LegacyRace**, or **Race**) 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 FY2004-2005 (N=570,018)

Reduction in Missing Data

52% were missing usable race from VA data sources

Age \geq 65	Age < 65
<p>53% missing usable VA race data</p> <p><i>Of those...</i></p> <p>95% had usable Medicare data</p>	<p>51% missing usable VA race data</p> <p><i>Of those...</i></p> <p>18% had usable Medicare data</p> <p>37% had usable DoD data</p> <p>52% had usable data from Medicare and/or DoD data</p>

Concordance with Non-VA Data Sources

Table compares non-VA data sources to self-reported VA race/ethnicity data

Race/Ethnicity	--
White and African Americans	Agreement was good (93-99%) for both non-VA data Sources
Non-African American Minorities	Agreement was poor (27-55%) for both Medicare and DoD
Hispanics	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

Resources	Released
CDW: A Conceptual Overview 2017	2017/03/29
CDW: Locating Its Documentation 2017	2017/04/05
Getting the Information You Need From CDW: SQL Starter Language	2015/01/26
Building Your Dataset in CDW: Joining Tables within a Domain	2015/07/27
Getting CDW Back Together: Joining CDW Tables (Continued)	2016/10/24
Data Management in SQL: Selected Intermediate SQL Skills	2016/10/31

Getting Started with Using CDW

Includes several seminars on using SQL to join and manipulate CDW data

<http://vaww.virec.research.va.gov/CDW/Documentation.htm> (VA Intranet only)



Race Data Best Practices Guide

Several SQL examples for multiple tasks utilizing race and ethnicity data

http://vaww.vhadatportal.med.va.gov/Portals/0/DataQualityProgram/Reports/Best_Practices_Guide_Race_Data.pdf (VA Intranet only)

The Researcher's Notebook

Using SQL to "Sort Out" Race in CDW

A method for cleaning multiple values of race.

Introduction

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 `vhadwa01.vha.med.va.gov`

Please note that the location of race data is now different from what is in these guides.

Example: Patsub.PatientRace

```
SELECT Race, FORMAT(COUNT(Race), 'N0') as Freq
FROM CDWork.Patsub.PatientRace
GROUP BY Race
ORDER BY COUNT(Race)
```

Format to show commas

	Race	Freq
1	*Unknown at this time*	2
2	*Missing*	4
3	WHITE NOT OF HISP ORIG	74,231
4	ASIAN	269,125
5	NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER	273,076
6	AMERICAN INDIAN OR ALASKA NATIVE	281,759
7	UNKNOWN BY PATIENT	695,252
8	DECLINED TO ANSWER	871,746
9	BLACK OR AFRICAN AMERICAN	3,938,842
10	WHITE	18,054,251

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”
- Map these additional entries 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

<http://vaww.virec.research.va.gov/Notebook/RNB/RNB6-CDW-SQL-to-Sort-Out-Race-CY16.pdf> (VA intranet only)

Example: Race Translation Table

```

drop table if exists #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')
insert into #RaceTranslationTable
values('AMERICAN INDIAN OR ALASKAN NATIVE','AMERICAN INDIAN OR ALASKA NATIVE')
insert into #RaceTranslationTable
values('AMERICAN INDIAN/ALASKA NATIVE','AMERICAN INDIAN OR ALASKA NATIVE')
insert into #RaceTranslationTable
values('AMERICAN INDIAN/ALASKAN','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, FORMAT(COUNT(b.StandardRace), 'N0') as Freq
FROM CDWork.PatSub.PatientRace as a LEFT JOIN #RaceTranslationTable as b
ON a.Race = b.InboundRace
GROUP BY b.StandardRace
ORDER BY COUNT(b.StandardRace)
```

100 % <

Results

Messages

	StandardRace	Freq
1	Unable to Map	6
2	ASIAN	269,125
3	NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER	273,076
4	AMERICAN INDIAN OR ALASKA NATIVE	281,759
5	UNKNOWN BY PATIENT	695,252
6	DECLINED TO ANSWER	871,746
7	BLACK OR AFRICAN AMERICAN	3,938,842
8	WHITE	18,128,482

Example: PatSub.PatientEthnicity

```
SELECT Ethnicity, FORMAT(COUNT(Ethnicity), 'N0') as Freq
FROM CDWork.PatSub.PatientEthnicity as a
GROUP BY Ethnicity
ORDER BY COUNT(Ethnicity)
```

100 % <

Results Messages

	Ethnicity	Freq
1	*Missing*	1
2	*Unknown at this time*	1
3	DECLINED TO ANSWER	473,961
4	UNKNOWN BY PATIENT	989,292
5	HISPANIC OR LATINO	1,408,195
6	NOT HISPANIC OR LATINO	21,794,870

Example: Collection Method

```
SELECT CollectionMethod, FORMAT(COUNT(CollectionMethod), 'N0') 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	444
4	OBSERVER	1,832
5	UNKNOWN	218,142
6	SELF IDENTIFICATION	24,237,866

← Default Value, rarely changed

Example: LegacyRace

```
SELECT a.LegacyRace, FORMAT(COUNT(1), 'N0') as Freq
FROM (SELECT DISTINCT PatientSID, LegacyRace from PatSub.PatientRace) as a
GROUP BY a.LegacyRace
ORDER BY COUNT(1) DESC;
```

Need to remove duplicates

100 % <

Results Messages

	LegacyRace	Freq
1	*Missing*	20,170,636
2	WHITE, NOT OF HISPANIC ORIGIN	2,059,432
3	UNKNOWN	644,741
4	BLACK, NOT OF HISPANIC ORIGIN	496,330
5	HISPANIC, WHITE	171,881
6	WHITE, NOT OF HISPANIC ORIGIN	109,097
7	CAUCASIAN	67,308
8	BLACK	51,483

Example: LegacyRace (Standard Values)

```

SELECT b.StandardRace, FORMAT(COUNT(b.StandardRace), 'N0') as Freq
FROM (SELECT DISTINCT PatientSID, LegacyRace from PatSub.PatientRace) as a
LEFT JOIN #RaceTranslationTable as b
      ON a.LegacyRace=b.InboundRace|
GROUP BY b.StandardRace
ORDER BY COUNT(b.StandardRace) DESC;

```

100 % <

Results Messages

	StandardRace	Freq
1	Unable to Map	20,837,492
2	WHITE	2,481,836
3	BLACK OR AFRICAN AMERICAN	622,239
4	AMERICAN INDIAN OR ALASKA NATIVE	15,106
5	UNKNOWN BY PATIENT	329
6	DECLINED TO ANSWER	112
7	ASIAN	10
8	NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER	6

Example: Multiple Sources (Long Format)

```
DROP TABLE IF EXISTS #SampleCohort
SELECT TOP 100 PatientSID, Sta3n
    INTO #SampleCohort
    FROM CDWork.Patient.Patient;
```

```
SELECT a.PatientSID, a.Sta3N, b.Race, b.CollectionMethod
FROM #SampleCohort as a LEFT JOIN cdwork.PatSub.PatientRace AS b
ON a.PatientSID = b.PatientSID
```

```
UNION ALL
```

```
SELECT c.PatientSID, c.Sta3N, d.LegacyRace, NULL as CollectionMethod
FROM #SampleCohort AS c
LEFT JOIN (SELECT DISTINCT PatientSID, LegacyRace from PatSub.PatientRace) as d
ON c.PatientSID = d.PatientSID
```

```
ORDER BY 1;
```

Sorts by the 1st column

Names don't need to match
as long as data type and
column order are the same

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

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 data from newer collection methods, if available
- Only consider data from old collection methods (< FY 2003), if newer data are not available*
 - **LegacyRace** contains race and ethnicity in CDW
 - **RACE** contains race and ethnicity in MedSAS

If conflicting values are still present...

- Prioritize values from specific sites, if relevant for the project
- Consider using all recorded values

When using MedSAS...

- Obtain race and ethnicity from both inpatient and outpatient files

* Unless the cohort has less recent utilization, LegacyRace may not be helpful

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 most likely to identify the right individuals if they use all 3 elements when conducting their VSF-study cohort record match

Note that...

- Medicare data cannot be used to identify Hispanics with any degree of accuracy or completeness, but
- **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**

The screenshot shows the VIREC INTRANET website. At the top left is the VIREC logo, and at the top right is a search bar with the text "Search All VA Web Pages" and a "Search" button. Below the search bar is a link to "Open Advanced Search". The main header reads "VA INFORMATION RESOURCE CENTER (VIREC)".

On the left side, there is a navigation menu with the following items: VIREC Home, VA/CMS Home, About Us, New Users of VA Data, FAQs, Acronyms, and HelpDesk.

The main content area is titled "Race & Ethnicity" and has a sub-section "Overview". The text in the overview section reads: "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." It then lists 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. It also mentions that federal data on ethnicity has two categories: Hispanic or Latino and Not Hispanic or Latino. Additional information is available from *Statistical Policy Directive No. 15* (Office of Business Management and Budget, 2003).

Below the overview is a section titled "Medical SAS Datasets". It states that 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.

It then notes that 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.

Below this is a section titled "Corporate Data Warehouse (CDW)". It states that 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,

On the right side of the page, there is a "General Resources" section with a list of links: Data Access, Data Sources, Data Tools, Data Topics, Products & Services, and 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
- ~1,400 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

(708) 202-2413

Contact information

VA Information Resource Center

Hines VA Hospital

virec@va.gov

708-202-2413

Maria Mor

Maria.Mor@va.gov



**Next session:
May 6, 2019 at 1 pm Eastern**



Database & Methods Cyberseminar Series

Session #8: *VetPop: A New Data Source with Socioeconomic Variables*

R. George Hauser, MD & Ankur Bhargava
Veterans Affairs Connecticut Healthcare System

Selected Recent References on Race/Ethnicity Data

- AHRQ (Agency for Healthcare Research and Quality) (2018). 2017 National Healthcare Quality and Disparities Report. (Rep. No. AHRQ Publication No. 18-0033-EF). 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 on Race/Ethnicity Data

- 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 on Race/Ethnicity Data

- 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 on Race/Ethnicity Data

- 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.