DATABASE & METHODS CYBERSEMINAR SERIES

Session 7:

Assessing Race and Ethnicity in VA Data

April 3, 2023

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 access and use VA databases.

Sessions cover…

- VA data sources & data access systems

- Application of VA data to research and quality improvement questions

- Limitations of secondary data use

- Resources to support VA data use

UPCOMING DATABASE & METHODS SESSIONS

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

Date	Topic
5/1/2023	An Introduction to VA Pharmacy Data: Sources and Uses for Medication Information
7/10/2023	Joint Longitudinal Viewer (JLV) and Research, Part 1: Getting Started
7/24/2023	Joint Longitudinal Viewer (JLV) and Research Part 2: Uses in Research and a Peek at Cerner Data

Visit the VIReC Database & Methods Cyberseminar page for more information & registration links.

Visit HSR&D's VIReC Cyberseminar Archive page to watch previous sessions.

Where can I download a copy of the slides?

SAMPLE EMAIL

A Practical Approach to Working with VA-Purchased Community Care Data

Thursday, October 13, 2022
2:00 PM | (UTC-04:00) Eastern Time (US & Canada) | 1 hr

Please download today's slides
~~Please click here for today's live captions~~
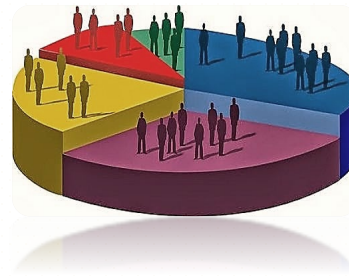
Join webinar

More ways to join:

Join from the webinar link
https://veteransaffairs.webex.com/veteransaffairs/j.php?

Poll #1:

*What is your primary **role** in projects using VA data?*

- Investigator, PI, Co-I

- Statistician, methodologist, biostatistician

- Data manager, analyst, or programmer

- Project coordinator
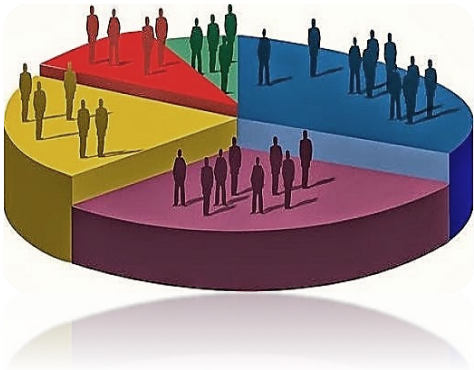
- Other – please describe via the chat function

Poll #2:

How many years of experience working with VA data?

- None – I'm brand new to this!

- One year or less

- More than 1, less than 3 years

- At least 3, less than 7 years

- At least 7, less than 10 years

- 10 years or more

VIReC

Poll #3:

What sources of VA race and ethnicity data have you used? (check all that apply)

- CDWWork (VistA)
- Cerner/Millennium
- OMOP
- MedSAS
- DOD (VADIR, DaVINCI)
- Other VA data sources

VIReC

By the end of this session, attendees will be able to:

- Locate race and ethnicity in VA and Medicare/Medicaid data

- Assess the quality of VA race and ethnicity data

- Create SQL code to use race and ethnicity data

Session roadmap

- Introduction
- Locating race and ethnicity in VA data
- Locating race and ethnicity in Medicare/Medicaid
- VA race and ethnicity data: Quality and Recommendations
- Examples

VIReC

Session roadmap

- **Introduction**
- Locating race and ethnicity in VA data
- Locating race and ethnicity in Medicare/Medicaid
- VA race and ethnicity data: Quality and Recommendations
- Examples

Guidance on Reporting Race and Ethnicity

- Race and ethnicity are social constructs

- Race and ethnicity should be reported with other sociodemographic factors and social determinants, including concerns about racism, disparities, and inequities

- The methods section should include an explanation of who identified race and ethnicity and the source of the classifications

- Specific racial and ethnic categories are preferred over collective terms; if used, clearly specify the categories included under "other"

- To the extent possible, report outcomes by specific racial and ethnic categories and consider analytic methods to accommodate smaller sample sizes in less frequently endorsed categories

Flanagin A, Frey T, Christiansen SL; AMA Manual of Style Committee. Updated Guidance on the Reporting of Race and Ethnicity in Medical and Science Journals. *JAMA*. 2021 Aug 17;326(7):621-627. doi: 10.1001/jama.2021.13304. PMID: 34402850

VIReC

Problems with Race and Ethnicity Data in VA

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

Problems with race and ethnicity data in the VA:

- **Incomplete**
- **Inaccuracies**
- **Inconsistent over time**
- **Inconsistent between sites**

VIReC

Racial and Ethnic Distribution of Veterans
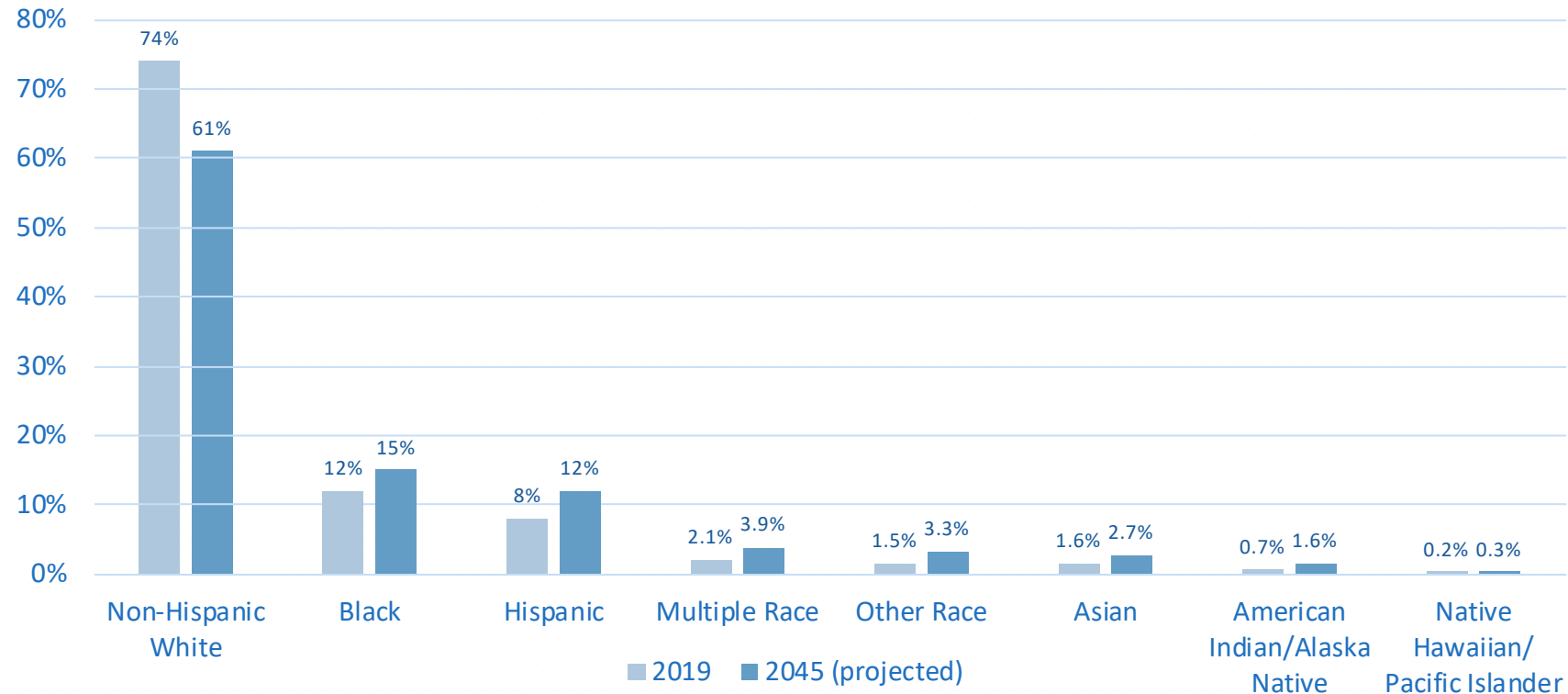
Office of Health Equity (https://www.va.gov/HEALTHEQUITY/Race_Ethnicity.asp)

VIReC

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

Ethnicity	Spanish
	Hispanic
	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

Race/Ethnicity Older Collection Methods (< FY2003)

- Prior to FY 2003 Race and Ethnicity captured in a single variable

- "Black" and "White" used for non-Hispanic Black and White

- Asian included Native Hawaiian or Other Pacific Islander

- No option for Ethnicity for Asian and American Indian races

- Could not select multiple races

Race/Ethnicity
Hispanic, white
Hispanic, black
American Indian
Black
Asian
White

VIReC

Acquisition of Race and 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 CPRS/Cerner

VIReC

Session roadmap

- Introduction
- **Locating race and ethnicity in VA data**
- Locating race and ethnicity in Medicare/Medicaid
- VA race and ethnicity data: Quality and Recommendations
- Examples

VIReC

Corporate Data Warehouse (CDW)

- National repository of data from VistA Patient File with race and ethnicity data from October 1999 to present

- 1 demographic record with the most recent value for each VA station (STA3N) a Veteran has visited

- Contains standard and nonstandard race values

- VistA racial data available in CDWWork.PatSub.PatientRace

  – Race (current collection standards)

  – LegacyRace (race and ethnicity from older collection standards < FY 2003)

  – **LegacyRace may be of limited utility**

- Cerner/Millennium data contained in CDWWork2 and CDWWork3

Patient 3.1 Release Documentation: https://vaww.virec.research.va.gov/CDW/Factbook/FB-CDW-Patient-Domain.pdf (VA Intranet only)

Race Tables in CDW

VistA race data are contained in CDWWork.PatSub.PatientRace

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

Race	Contains patient race from current collection methods. Multiple records if more than one race identified.
CollectionMethod	Contains method of data collection for Race
LegacyRace	Contains patient race/ethnicity from < FY 2003 − 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.

VIReC

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

  - Only use data from the current reporting standards, if available

  - Use all recorded races for patients from the utilized reporting standard

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 - *current method*

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

PatSub.PatientRace (LegacyRace) - *old method*

Hispanic ethnicity (e.g., HISPANIC, WHITE; HISPANIC, BLACK)

Non Hispanic ethnicity (e.g., WHITE NOT OF HISP ORIG; BLACK NOT OF HISP ORIG)

Not all LegacyRace values indicate ethnicity (e.g., ASIAN, BLACK)

CDW Ethnicity Data (Data Quality Report)
https://vaww.vhadataportal.med.va.gov/Portals/0/DataQualityProgram/Reports/CDW_Ethnicity_Data.pdf (VA Intranet only)

VIReC

Cerner/Millennium (CDWWork3)

- CDWWork3 contains standardized views with combined data from CDWWork and Millennium (CDWWork2 data model)

- Views have the same names as CDWWork with the suffix _EHR

- **Dim tables**

  - CDWWork3.Dim.Race_EHR

  - CDWWork3.Dim.Ethnicity_EHR

- **Patient tables**

  - CDWWork3.Patsub.PatientRace_EHR (Race)

    - LegacyRace is null for all millennium data (sta3n=200)

    - Only contains **one** value of Race per person (across all of VA)

  - CDWWork3.Patsub.PatientEthnicity_EHR (Ethnicity)

VIReC

CDWWork3.Dim.Race

- CDWWork3.Dim.Race contains the combined DIM tables with the Millennium and CDW values

- Sta3n = 200 for the data extracted from Cerner/Millennium

- In the Millennium data, race will be common across all stations

	RaceSID	RaceIEN	Sta3n	Race	RaceAbbrev
1	1800238528	309317	200	Asian	A
2	1800238529	309318	200	American Indian or Alaska Native	3
3	1800267576	309315	200	Black or African American	B
4	1800267577	309316	200	White	W
5	1800482022	4189861	200	Native Hawaiian or Pacific Islander	H
6	1800482023	4189862	200	Other Race	NULL
7	1	13	528	WHITE	W
8	2	12	528	UNKNOWN BY PATIENT	U
9	3	11	528	NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER	H
10	4	10	528	DECLINED TO ANSWER	D
11	5	9	528	BLACK OR AFRICAN AMERICAN	B

Race values in Millennium (subset of values, limited to those used in VA)

Site-specific race values from CDW

- VIReC has EHRM data integration resources on the home page (https://vaww.virec.research.va.gov)

Cerner/Millennium (CDWWork2)

- CDWWork2 is in the Millennium data model

- Used for operations and now available for research use

- CDWWork2.VeteranMill.Person or SVeteranMill.SPerson

  - Person-level race and ethnicity data with **one** value per person

  - Has both the display values and SIDs to link to CDWWork2.NDimMill.CodeValue

  - Race (RaceCD and RaceCodeValueSID)

  - EthnicGroup (EthnicGroupCD and EthicGroupCodeValueSID)

- CDWWork2.NDimMill.CodeValue

  - Essentially one giant Dim table with all the code value sets.

  - CodeValueSetID = 282 for race code values

  - CodeValueSetID = 27 for ethnicity code values

  - The code value sets contain many values for race and ethnicity not used in VA.

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

VINCI V5 OMOP DATABASE DATA SPECIFICATIONS
https://sps.vinci.med.va.gov/prod/vincipedia/OMOP%20Academy%20Training%20Materials/Forms/AllItems.aspx?RootFolder=%2fprod%2fvincipedia%2fOMOP%20Academy%20Training%20Materials%2fDocuments&FolderCTID=0x012000E41B19131A7A824CADA1762BC5E2A9BC

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:"
https://vaww.vhadataportal.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":
https://vaww.virec.research.va.gov/Notebook/RNB/RNB6-CDW-SQL-to-Sort-Out-Race-CY16.pdf

VIReC

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

Use the self-reported ethnicity provided under the new collection method, when available

Otherwise, non-self-reported ethnicity provided by the new collection method

Ethnicity from the old collection methods is used when no other data are available

"CDW Ethnicity Data:"
https://vaww.vhadataportal.med.va.gov/Portals/0/DataQualityProgram/Reports/CDW_Ethnicity_Data.pdf

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://www.hsrd.research.va.gov/for_researchers/cyber_seminars/archives/video_archive.cfm?SessionID=2441

USVETS (United States Veterans Eligibility Trends and Statistics)

- Combines VA, DOD, and commercial data for all Veterans (not just VA enrolled)

- Uses a proprietary algorithm and multiple sources to determine race and ethnicity

- Data should only be used at the aggregate level as an individual may have imputed data

- Race and ethnicity contained in the Veteran Static Data

- https://vaww.vhadataportal.med.va.gov/DataSources/USVETSData.

Hispanic: Hispanic indicator	
Value	**Description**
0	Non-Hispanic
1	Hispanic

VIReC

USVETS

Race	
Value	**Description**
1	White
2	Black/African American
3	American Indian/Alaska Native
4	Asian
5	Native Hawaiian/Pacific Islander
6	Asian or Pacific Islander, Unspecified
7	Other
8	Two or more races
Z	Unknown

Race and Ethnicity Variables in MedSAS

Prior to FY2003 (old data collection methods)

- Race and ethnicity captured jointly in a single variable RACE

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 (see Bonus Slides #76-79)
  - First character has race or ethnicity
  - Second character has method of data collection

Location (historical encounter-level race and ethnicity data)

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

Session roadmap

- Introduction
- Locating race and ethnicity in VA data
- **Locating race and ethnicity in Medicare/Medicaid**
- VA race and ethnicity data: Quality and Recommendations
- Examples

VIReC

Sources of Medicare/Medicaid Race in VA

VA Medicare Data

– Master Beneficiary Summary File (MBSF_BASE, 2011-2022)

– **RACE**

– **RTI_RACE_CD / RTI_RACE (2011-2015)**

VA Medicaid Data

– Medicaid Personal Summary (TAF_DE_BASE, 2014-2020)

– **RACE_ETHNTCY_CD**

– **ETHNTCY_CD**

VA Vital Status File (scheduled for retirement May 2023)

– **CMS_RACE** (Master File only)

– One record for each SSN-date of birth (DOB)-gender combination found in VA data

– Some SSNs have more than one record

Medicare Race and 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/replacement card

- Reporting sources: Usually self or family

Distinctions from current VA race and ethnicity data

- 'Hispanic' is a race category

- No multiple race reporting

VIReC

Medicare Race and Ethnicity 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 |

VIReC

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 Values

RACE

Value	Description
0	Unknown
1	White
2	Black
3	Other
4	Asian
5	Hispanic
6	North American Native

Medicare RTI Race Values

RTI_RACE / RTI_RACE_CD

Value	Description
0	Unknown
1	Non-Hispanic White
2	Black (or African-American)
3	Other
4	Asian/Pacific Islander
5	Hispanic
6	American Indian / Alaska Native

VIReC

Medicare Race and Ethnicity Data Summary

Data quality issues

- Information on many 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 and ethnicity self-report

- RTI Race Algorithm

Medicaid Race and Ethnicity

Data are collected at the state level

- Patients are asked to provide self-reported race and ethnicity

- States have flexibility in the implementation of data collection

- Categorization of data used for collection can vary state by state

- Amount of missing data varies widely by state

Current data does not include a separate race variable

- Detailed data may be collected on both race an ethnicity

- Ethnicity available as a separate variable

- Race is only contained in combined race/ethnicity variables that do not report race for those categorized as Hispanic.

Medicaid Ethnicity

ETHNCTY_CD

Value	Description
0	Not of Hispanic, or Latino/a, or Spanish origin
1	Mexican, Mexican American, Chicano/a
2	Puerto Rican
3	Cuban
4	Another Hispanic, Latino, or Spanish origin
5	Hispanic or Latino, Subcategory Unknown
Null/ missing	source value is missing or unknown

Medicaid Race and Ethnicity

RACE_ETHNCTY_CD

Value	Description
1	White, non-Hispanic
2	Black, non-Hispanic
3	Asian, non-Hispanic
4	American Indian and Alaska Native (AIAN), non-Hispanic
5	Hawaiian/Pacific Islander
6	Multiracial, non-Hispanic
7	Hispanic, all races
Null/ missing	source value is missing or unknown

Medicaid Race and Ethnicity Data Summary

Race only available in TAF_DE_BASE as a combined variable

– Priority given to Hispanic ethnicity

– Race is only coded for non-Hispanic Veterans

Additional detail may be available

– ETHNCTY_CD contains more specific ethnicity detail (e.g. Mexican, Cuban)

– RACE_ETHNCTY_EXP_CD – granular detail on race (e.g. Japanese, Korean)

  ▪ Only coded for non-Hispanic Veterans

  ▪ https://resdac.org/cms-data/variables/expanded-race-and-ethnicity-constructed-code-latest-year

– Prior data contained standardized race and ethnicity (MAX_PS, 1999-2015)

  ▪ ETHNICITY_CODE

  ▪ RACE_CODE_1 – RACE_CODE_5

VIReC

Medicaid Race and Ethnicity Data Issues

- Availability lags behind both VA and Medicare

- Only about 10% of Veterans are enrolled

- Data collection
  - Varies by state
  - Varies over time

- Definition of variables available for research has changed
  - See Bonus Slides 84-86

Session roadmap

- Introduction
- Locating race and ethnicity in VA data
- Locating race and ethnicity in Medicare/Medicaid
- **VA race and ethnicity data: Quality and Recommendations**
- Examples

VIReC

Completeness of CDW Race and Ethnicity

- **Completeness has varied over time**

  – Prior to FY 2003 data changes data were only about 50% complete

  – Completeness improved from 58.7 to 84.6* between (FY 2005 to 2012)

- **Completeness of Ethnicity in FY 2021**

  – **92%** of all patients have ethnicity recorded

  – Conflicting values are rare (<<1%)

CDW Race Data and Multiple Races (Data Quality Report):
https://vaww.vhadataportal.med.va.gov/Portals/0/DataQualityProgram/Reports/CDW_Race_Data_and_Multiple_Races.pdf (VA Intranet only)

VIReC

CDW Completeness of Race Data FY2021

New collection methods

Old collection methods

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

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

0.2% have conflicting values

1.1% of Veterans only have older race data

57.2% LegacyRace & MedSAS Race

33.2% MedSAS Race only

9.6% LegacyRace only

Unique Veterans with ≥ 1 outpatient visit (NoncountClinicFlag = 'N') in FY2021

VIReC

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)

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 ≥ 65	Age < 65
53% missing usable VA race data	**51% missing usable VA race data**
Of those...	*Of those...*
95% had usable Medicare 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

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

Race and 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

VIReC

Recommendations: VA Data

Currently Millennium has 1 value of race and ethnicity per person

When multiple sources of race and ethnicity exist...

→ Use data from current collection methods, if available

→ Only consider data < FY 2003 (LegacyRace in CDW or RACE in MedSAS), if newer data are not available*

When using MedSAS...

→ Obtain race and ethnicity from both inpatient and outpatient files

→ If using data < FY 2003 , use MedSAS instead of CDW as the primary source

If conflicting values are still present...

→ Prioritize values from specific sites, if relevant for the project

→ Consider using all recorded values

Unless the cohort has a long history of prior utilization, LegacyRace may not be helpful

Recommendations: Multiple Values

Analytic Choices and Considerations

→ Combine race and ethnicity (Hispanic, Non-Hispanic White, Non-Hispanic Black …)

→ Use of combined categories (multi-racial, other, etc.)

→ Exclusion of multiple-race, conflicting race, or groups with small sample sizes

→ Including individuals in all selected groups (White alone or in combination, …)

→ Bridge into a single category

  → Whole assignment, all individuals in a particular multiple-race category are assigned to the same single race ("White and Black" -> "Black")

  → Fractional assignment

    → Equal assignment ("White and Black" -> 0.5 "White" and 0.5 "Black")

    → Weights assigned by regression models or additional data (NCHS regression method)

Liebler CA, Halpern-Manners A. A practical approach to using multiple-race response data: a bridging method for public-use microdata. Demography. 2008 Feb;45(1):143-55. doi: 10.1353/dem.2008.0004. PMID: 18390296; PMCID: PMC2831381.

VIReC

Recommendations: Non-VA Data

- Use of non-VA race data can reduce missing data

- Carefully consider any potential bias (e.g., age or disability)

- Classifying non-Black minorities as "Other" results in better agreement with other data sources, but at the loss of potentially valuable detail

- **RTI_RACE** increases the identification of Hispanics and Asians (Medicare)

- Potential supplementary data sources

Medicare	Department of Defense	Medicaid	Special Surveys	USVETS

VIReC

Session roadmap

- Introduction
- Locating race and ethnicity in VA data
- Locating race and ethnicity in Medicare/Medicaid
- VA race and ethnicity data: Quality and Recommendations
- **Examples**

SQL Examples in CDW

Note:

Some guides were written before the current data structure in the Patient 3.1 domain. See slide #78 for an overview of these changes.

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

https://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

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

Researcher's Notebook: Using SQL to "Sort Out" Race in CDW

https://vaww.virec.research.va.gov/Notebook/RNB/RNB6-CDW-SQL-to-Sort-Out-Race-CY16.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

Example: Patsub.PatientRace

```
select Race, Format(count(1), 'N0') as Freq
from CDWWork.Patsub.PatientRace
group by Race
order by count(1)
```

Format to show commas

SQL won't count over NULL values

209 %

Results | Messages

	Race	Freq
1	NULL	42,451
2	WHITE NOT OF HISP ORIG	90,499
3	NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER	303,599
4	ASIAN	308,762
5	AMERICAN INDIAN OR ALASKA NATIVE	312,980
6	UNKNOWN BY PATIENT	771,938
7	DECLINED TO ANSWER	959,268
8	BLACK OR AFRICAN AMERICAN	4,294,819
9	WHITE	19,613,640

VIReC

Example: Patsub.PatientEthnicity

```sql
select Ethnicity, Format(count(1), 'N0') as Freq
from CDWWork.Patsub.PatientEthnicity
group by Ethnicity
order by count(1)
```

209 %

Results | Messages

	Ethnicity	Freq
1	*Unknown at this time*	1
2	*Missing*	2
3	DECLINED TO ANSWER	506,090
4	UNKNOWN BY PATIENT	1,103,104
5	HISPANIC OR LATINO	1,566,490
6	NOT HISPANIC OR LATINO	23,515,787

Example: Collection Method

```
select CollectionMethod, Format(count(1), 'N0') as Freq
from CDWWork.Patsub.PatientRace
group by CollectionMethod
order by count(1)
```

209 %

Results Messages

	CollectionMethod	Freq
1	*Unknown at this time*	1
2	*Missing*	3
3	PROXY	422
4	OBSERVER	1,663
5	UNKNOWN	213,474
6	SELF IDENTIFICATION	26,482,393

← **Default Value, rarely changed**

Example: Using a Race Lookup Table

- ## Create a lookup table for the variable Race

  - Standardize values (non-standard values or values from different data sources)

  - Create indicator variables

- ## Change categories to match project needs

- ## This example does not address the large number of non-standard values used in LegacyRace

  - To standardize LegacyRace use the "Race Data Best Practices Guide"

  - Code starting on p.10 is source for #RaceTranslationTable on slide #82.

    See Researcher's Notebook: Using SQL to "Sort Out" Race in CDW for alternate method for programming standard race values

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

Example: Race Values in CDWWork3

- CDWWork3 contains race values from VistA and the Cerner/Millennium data

- Many unused race code values in Cerner/Millennium, focus on utilized values

```sql
SELECT DISTINCT Race
FROM [CDWWork3].[PatSub].[PatientRace_EHR]
```

	Race
1	American Indian or Alaska Native
2	BLACK OR AFRICAN AMERICAN
3	Other Race
4	UNKNOWN BY PATIENT
5	NULL
6	WHITE
7	DECLINED TO ANSWER
8	*Implied NULL*
9	Native Hawaiian or Pacific Islander
10	NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER
11	WHITE NOT OF HISP ORIG
12	Asian

SQL is case insensitive. The difference between these entries is in the phrasing not the case.

VIReC

Example: Convert to Standard Values

```sql
SELECT DISTINCT Race
    ,CASE
        WHEN Race = 'AMERICAN INDIAN OR ALASKA NATIVE'
            THEN 'American Indian or Alaska Native'
        WHEN Race = 'UNKNOWN BY PATIENT' THEN NULL
        WHEN Race = 'DECLINED TO ANSWER' THEN NULL
        WHEN Race = '*Implied NULL*' then NULL
        WHEN Race = 'NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER'
            THEN 'Native Hawaiian or Pacific Islander'
        WHEN Race = 'WHITE NOT OF HISP ORIG' then 'White'
        ELSE Race
        END AS StandardRace
INTO #RaceLookup
FROM [CDWWork3].[PatSub].[PatientRace_EHR]
```

Example: Indicator Variables

```
,CASE
    WHEN StandardRace = 'Black or African American' THEN 1
    WHEN StandardRace is null then null
    ELSE 0
    END AS Black                    ⟶  Creates indicator: Black
,CASE
    WHEN StandardRace = 'White' THEN 1
    WHEN StandardRace is null then null
    ELSE 0
    END AS White                    ⟶  Creates indicator: White
,CASE
    WHEN Race = 'WHITE NOT OF HISP ORIG' THEN 0
    ELSE null
    END AS Hispanic
INTO #RaceLookup2
FROM #RaceLookup
```

Note:
For illustration purposes only. "NOT OF HISP ORIG", likely not truly informative of ethnicity for this category.

This non-standard value contains information for both race and ethnicity

VIReC

Example: Completed Lookup Table

Using the Lookup table

- SQL won't join on NULL values. May require additional programming if non-NULL values desired.
- SQL will join if linking variables have the same text but different case.

	Race	StandardRace	Asian	OtherRace	NativeHawaiian	AmericanIndian	Black	White	Hispanic
1	Asian	Asian	1	0	0	0	0	0	NULL
2	Other Race	Other Race	0	1	0	0	0	0	NULL
3	NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER	Native Hawaiian or Pacific Islander	0	0	1	0	0	0	NULL
4	*Implied NULL*	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
5	American Indian or Alaska Native	American Indian or Alaska Native	0	0	0	1	0	0	NULL
6	DECLINED TO ANSWER	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
7	Native Hawaiian or Pacific Islander	Native Hawaiian or Pacific Islander	0	0	1	0	0	0	NULL
8	Black or African American	Black or African American	0	0	0	0	1	0	NULL
9	UNKNOWN BY PATIENT	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
10	White	White	0	0	0	0	0	1	NULL
11	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
12	WHITE NOT OF HISP ORIG	White	0	0	0	0	0	1	0

THANK YOU!
Questions?

VIReC
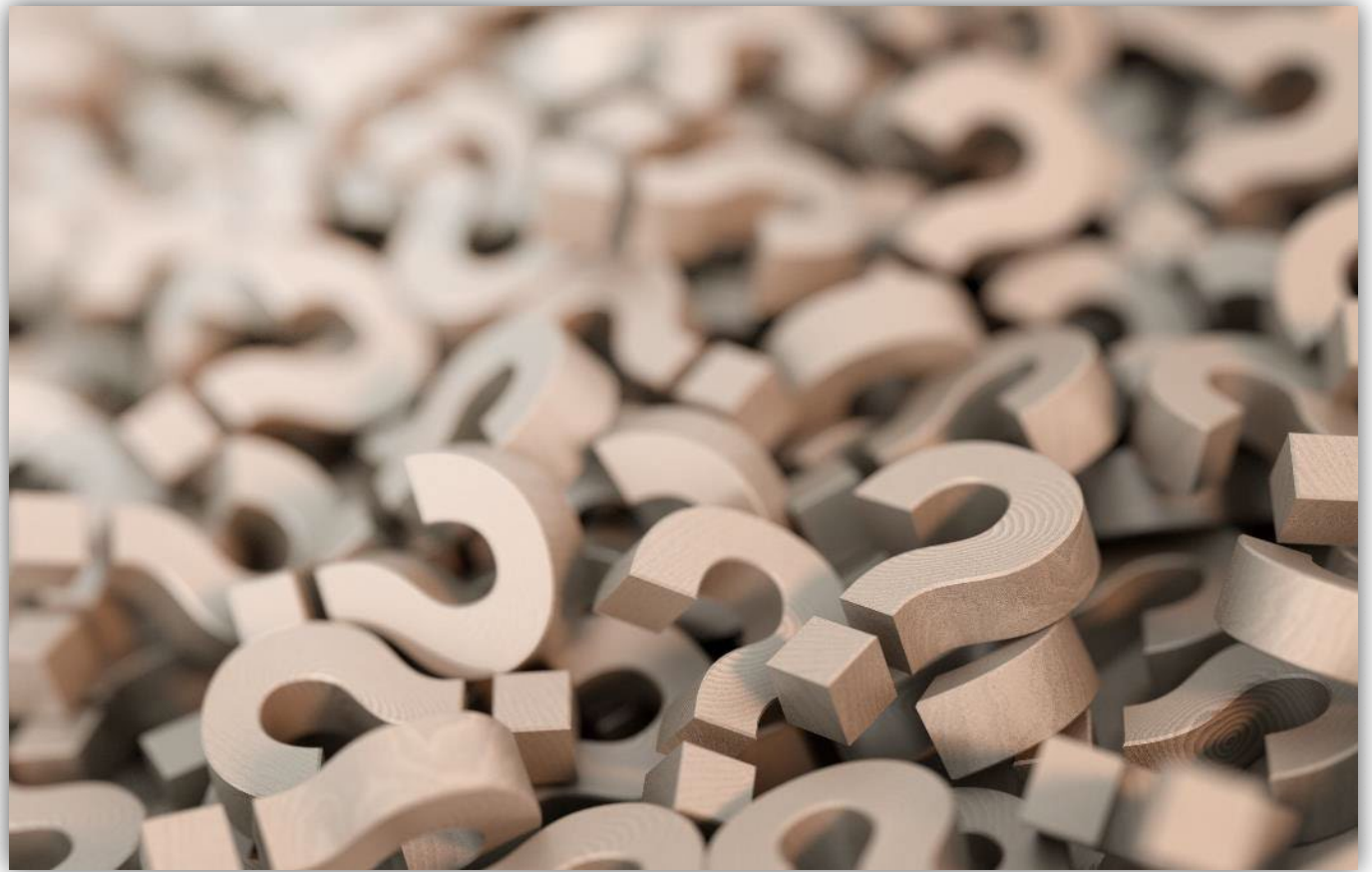
CONTACT INFORMATION

Maria Mor, PhD

Co-Director

Biostatistics, Informatics, and Computing Core

VA Center for Health Equity Research and Promotion

VA Pittsburgh Healthcare System

maria.mor@va.gov

DATABASE & METHODS CYBERSEMINAR SERIES

Next session:
May 1, 2023 at 1 pm Eastern

An Introduction to VA Pharmacy Data: Sources and Uses for Medication Information

VIReC

Race & Ethnicity
BONUS SLIDES

VIReC resources on Race and Ethnicity

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 in the *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity* (Office of Business Management and Budget,1997).

COVID-19 National Summary

The VA COVID-19 National Summary provides information on the status of COVID-19 patients who have been tested or treated at VA facilities. The "At a Glance" section includes a summary of VA COVID-19 cumulative cases by patient category or race and ethnicity.

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.

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

https://vaww.virec.research.va.gov/RaceAndEthnicity/Overview.htm (Intranet only)

Centralized Interactive Phenomics Resource

VA Phenomics Library

CIPHER

Centralized Interactive Phenomics Resource

CIPHER Goals
- Disseminate phenotype definitions and metadata
- Facilitate VA EHR-based phenotyping

CIPHER Content
- Phenotype definitions
- Data visualization tools
- Phenotyping resources

https://vhacdwdwhweb100.vha.med.va.gov/phenotype/index.php/VA_Phenomics_Library_-_Centralized_Interactive_Phenomics_Resource_(CIPHER) (Intranet only)

https://vhacdwdwhweb100.vha.med.va.gov/phenotype/index.php/Category:Demographics (Intranet only)

VIReC

Race and Ethnicity Data Sources

Source	Comment
CDWWork.Patsub.PatientRace	Most recent recorded race per person/station. LegacyRace contains race under the old collection standard
CDWWork.Patsub.PatientEthnicity	Most recent recorded ethnicity per person/station
CDWWork2.VeteranMill.Person	Millennium (Cerner) race and ethnicity
CDWWork.Patsub.PatientRace_EHR	Combined VistA and Millennium race data
CDWWork.Patsub.PatientEthnicity_EHR	Combined VistA and Millennium ethnicity data
MedSAS	Encounter-level race and ethnicity data. More complete source for older race data and contains historic values

VIReC

Race and Ethnicity Data Sources (continued)

Source	Comment
OMOP	Contains one value each for race and ethnicity, derived from CDW data
Medicare*	Separate application process data. Available for older and disabled Veterans CMS_RACE in VA Vital Status File (retiring in 2023)
Medicaid*	Tend to be available for lower income Veterans.
DOD*	Greater availability for younger, rather, than older Veterans.
USVETS*	Use only for aggregate inferences, may contain imputed valued.

Require specific process to apply for data access

Requesting Access to Race and Ethnicity Data

- Navigating VA Data Access: An Overview of the Process for Requesting Permission to Use VA Data

  - https://www.hsrd.research.va.gov/for_researchers/cyber_seminars/archives/video_archive.cfm?SessionID=4063

- CDW, MedSAS, and OMOP can be requested through:

  - DART data request for research

  - e-pas form for operations

- VHA Data Portal

  - http://vaww.vhadataportal.med.va.gov/DataSources/VACMSData.aspx (VA intranet only)

  - http://vaww.vhadataportal.med.va.gov/DataSources/USVETSData.aspx (VA intranet only)

  - http://vaww.vhadataportal.med.va.gov/DataSources/DAVINCI.aspx (VA intranet only)

CDW Race Table Changes

The structure of CDW data is subject to periodic changes.

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

VIReC's Patient 3.0 Domain Factbook **documents the current** data structure of race data in CDWWork.

CDW documentation may refer to race from older collection methods as being located in other CDW tables.	
Patient.Patient or SPatient.SPatient tables	Currently LegacyRace and LegacyRaceSID in Patsub.PatientRace Previously contained RaceSID for linking to CDWWork.Dim.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):
https://vaww.vhadataportal.med.va.gov/Portals/0/DataQualityProgram/Reports/Best_Practices_Guide_Race_Data.pdf (VA Intranet only)

VIReC

Medical SAS Datasets:
Completeness of Race and Ethnicity Data

Prior to FY2003
<60% of patients had usable* race/ethnicity

FY2003
Completeness of data was about 50%

FY2015
Completeness of data was >90%

Completeness varies between inpatient and outpatient files.

Always use both the inpatient and outpatient data to capture race and ethnicity in the MedSAS files.

A usable race value is any value that is not 'missing' or 'unknown' or 'declined'

VIReC

Medical SAS Datasets: Race and 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

VIReC

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

VIReC

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

VIReC

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

VIReC

Non-standard Race Values in CDW

- Non-standard values problematic in LegacyRace (>99%)

- May also need to standardize when using data from multiple sources

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;	**Black or African American**
White Not of Hisp orig; White, Not Hispanic; Hispanic White; Caucasian;	**White**
Pacific Islander	**Native Hawaiian or Other Pacific Islander**
Asian or Pacific Islander, Asian Pacific Islander, Mexican American, Unknown	**No direct map to a standard race**

VIReC

Example: LegacyRace

```sql
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

Table lists the most frequently occurring values out of 40 total

VIReC

Example: LegacyRace (Standard Values)

```sql
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;
```

See slide #60 for references to create this lookup table

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

VIReC

Medicare Race and Ethnicity: File Locations

- MBSF_BASE (2011-Forward)

- BENE_SUM (2009-2010)

- DENOM (1997-2008)

- Note, variable name for RTI race has changed over time
  - RTI_RACE_CD (2016 – Forward)
  - RTI_RACE (<2016)

Medicaid Race and Ethnicity: File Locations

- TAF_DE_BASE (2014-Forward)
  - ETHNCTY_CD
  - RACE_ETHNCTY_CD
  - RACE_ETHNCTY_EXP_CD
- MAX_PS (1999-2015)
  - EL_RACE_ETHNCY_CD
  - RACE_CODE1—RACE_CODE5
  - ETHNICITY_CODE

Medicaid Race and Ethnicity: Variable Format

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 and Ethnicity: Variable Description

Variable	Description
RACE_CODE1	Race - White
RACE_CODE2	Race - Black/ African American
RACE_CODE3	Race - Am Indian/Alaskan
RACE_CODE4	Race - Asian
RACE_CODE5	Race - Hawaiian/Pac. Islands
ETHNICITY_CODE	Ethnicity - Hispanic

Selected References on Race and Ethnicity Data

AHRQ (Agency for Healthcare Research and Quality) (2019). 2018 National Healthcare Quality and Disparities Report. (Rep. No. AHRQ Publication No. 19-0070-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.

VIReC

Selected Recent References on Race and Ethnicity Data

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.

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.

VIReC

Selected Recent References on Race and Ethnicity Data

Kashner TM. (1998). Agreement between administrative files and written medical records: a case of the Department of Veterans Affairs. Med Care, 36, 1324-1336.

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.

VIReC

Selected Recent References on Race and Ethnicity Data

Polednak AP. (2001). Agreement in race-ethnicity coding between a hospital discharge database and another database. Ethn Dis, 11, 24-29

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.

VIReC

Selected Recent References on Race and Ethnicity Data

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.

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.
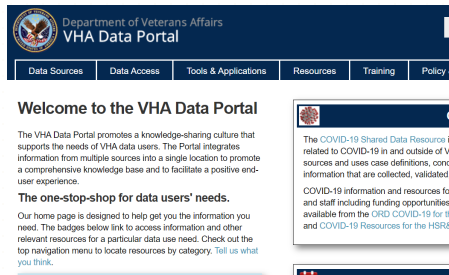
VIReC

Database & Methods
BONUS SLIDES

Resources for *VA Data Users*
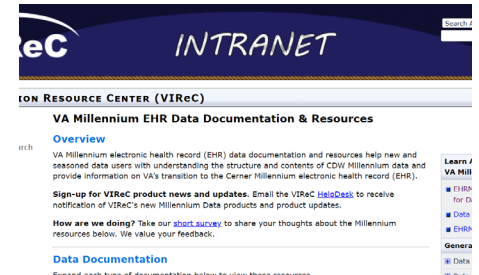
Select image to visit page

VA Information Resource Center (VIReC) (VA Intranet)

VHA Data Portal (VA Intranet)
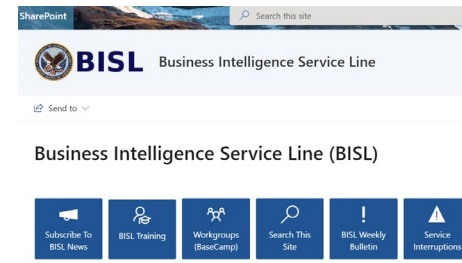

VIReC Cyberseminars

VA Millennium EHR Data Documentation (VA Intranet)

Quick Guide: Resources for Using VA Data (VA Intranet)

VA Informatics and Computing Infrastructure (VINCI) (VA Intranet)

BISL/CDW (VA Intranet)

Health Economics Resource Center (HERC) (VA Intranet)

HSRData Listserv

- Community knowledge sharing
- ~1,400 VA data users
- Researchers, operations, data stewards, managers
- Subscribe by visiting
  vaww.virec.research.va.gov/Support/HSRData-L.htm (VA Intranet)

VIReC HelpDesk

- Individualized support
- Request Form:
  varedcap.rcp.vaec.va.gov/redcap/surveys/?s=KXMEN77LXK (VA Intranet)

Questions about using VA Data?

VIReC