Database & Methods Cyberseminar Series

Session #8. *United States Veterans Eligibility Trends and Statistics (USVETS): A New Data Source with Socioeconomic Variables*

The objectives of this cyberseminar are to:

- To introduce the USVETS database to potential collaborators including addressing its accuracy, completeness, consistency, traceability, uniqueness, validity, and timeliness.

- Define socioeconomic data available in USVETS, purchased from third parties for all Veterans

- Discuss the process to obtain access to USVETS

05/2019

USVETS roadmap

- Story behind USVETS

- Quality

- Organization of USVETS data

- Specific data elements in USVETS

- Access to USVETS

USVETS roadmap

- **Story behind USVETS**
  - **What is USVETS?**
  - **Who produces USVETS?**
  - **Why is USVETS produced?**
  - **How is USVETS produced?**

- Quality
- Layout of USVETS data
- Specific fields in USVETS
- Access

What is USVETS?

- All US Veterans: Veteran military history, demographics, socioeconomics, and utilization of VA benefits and services
- Contains 250+ variables, integrates data from 35+ VA and federal data sources, and contains over 38 million records (and growing).

US Department of Veterans Affairs

Office of Enterprise Integration, Data Governance and Analytics

USVETS Data Quality Assessment

for
United States Veterans Eligibility Trends and Statistics (USVETS) Support
GS-10F-0062X VA119A-17-F-0295

Submitted: 2/20/2019

By: Clyde Blandford, Program Manager

Who Produces USVETS?

- Office of Enterprise Integration → Data Governance & Analytics

Staff Offices

- Acquisition, Logistics, and Construction
- Advisory Committee Management Office
- Board of Veterans' Appeals
- Center for Faith-Based and Neighborhood Partnerships
- Center for Minority Veterans
- Center for Women Veterans
- General Counsel
- Office of Accountability and Whistleblower Protection
- Inspector General
- Office of Congressional and Legislative Affairs
- Office of Employment Discrimination Complaint Adjudication
- Office of Human Resources and Administration
- Office of Information and Technology
- Office of Management
- Veterans Experience Office (VEO)
- Office of Operations, Security and Preparedness
- Office of Enterprise Integration
- Office of Public & Intergovernmental Affairs
- Office of Regulation Policy and Management
- Office of Small and Disadvantaged Business Utilization
- Office of Survivors Assistance
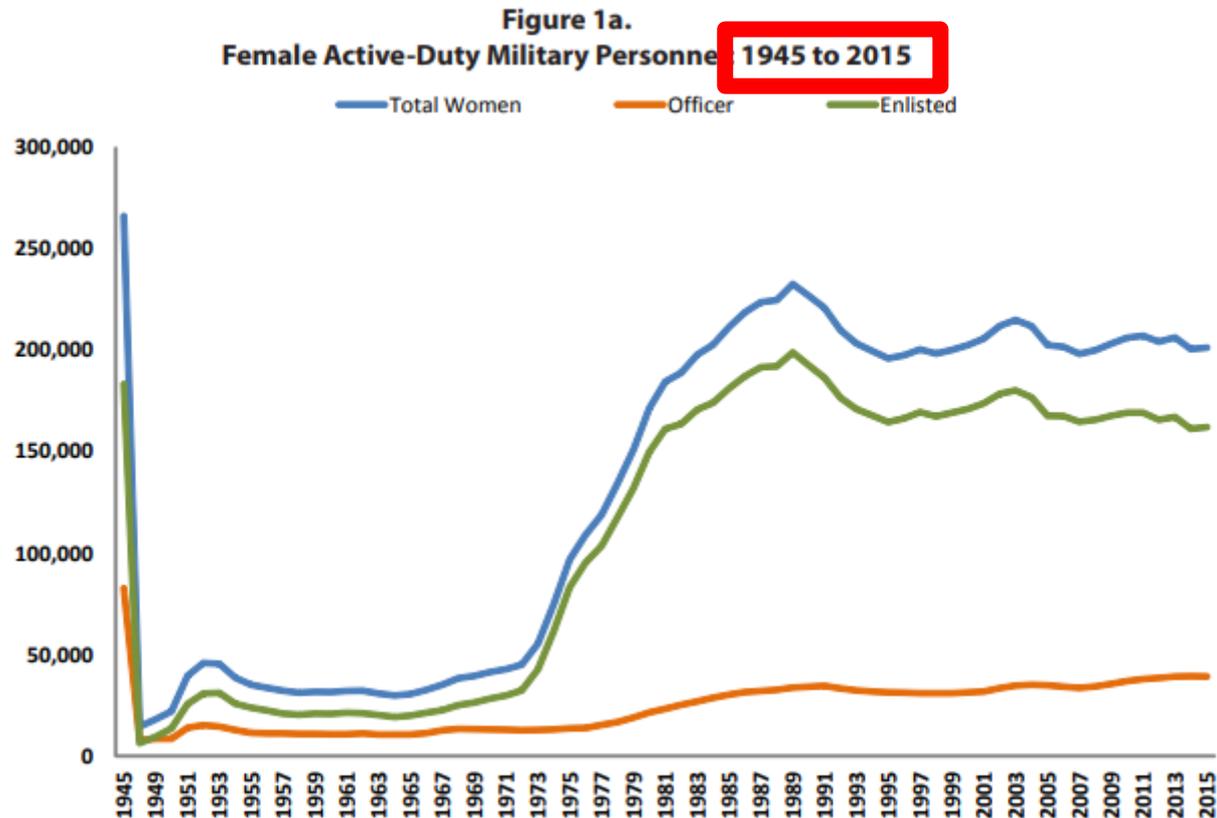- Veterans Service Organization Liaison
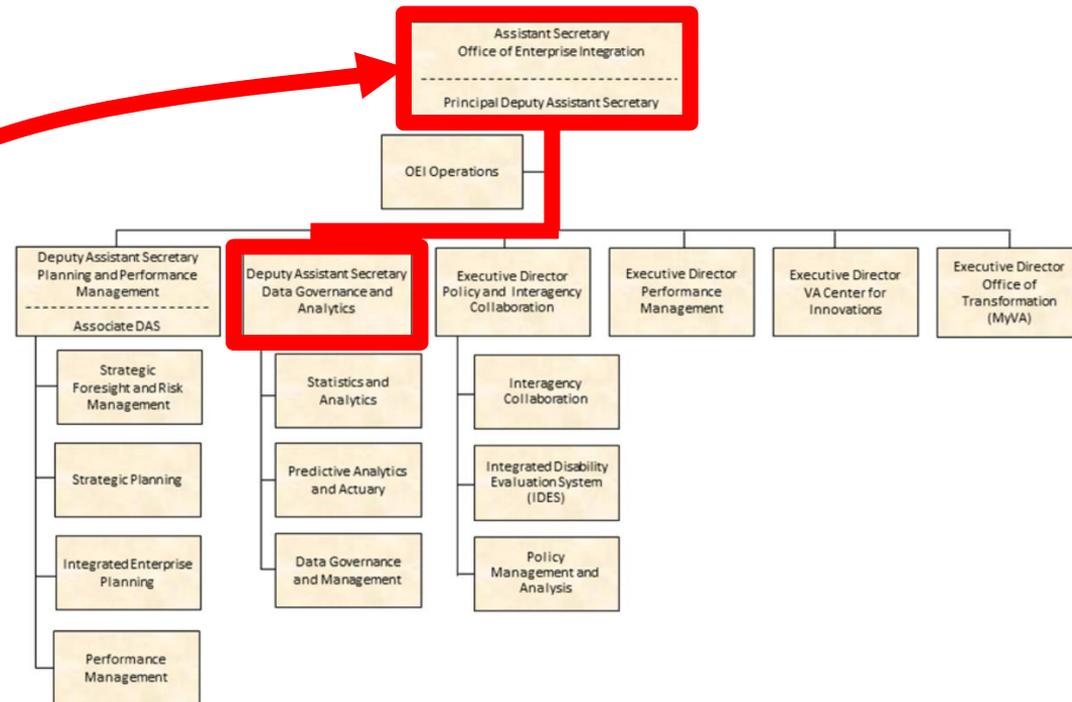
Administrations

- National Cemetery Administration
- Veterans Benefits Administration
- Veterans Health Administration

https://www.va.gov/oei/

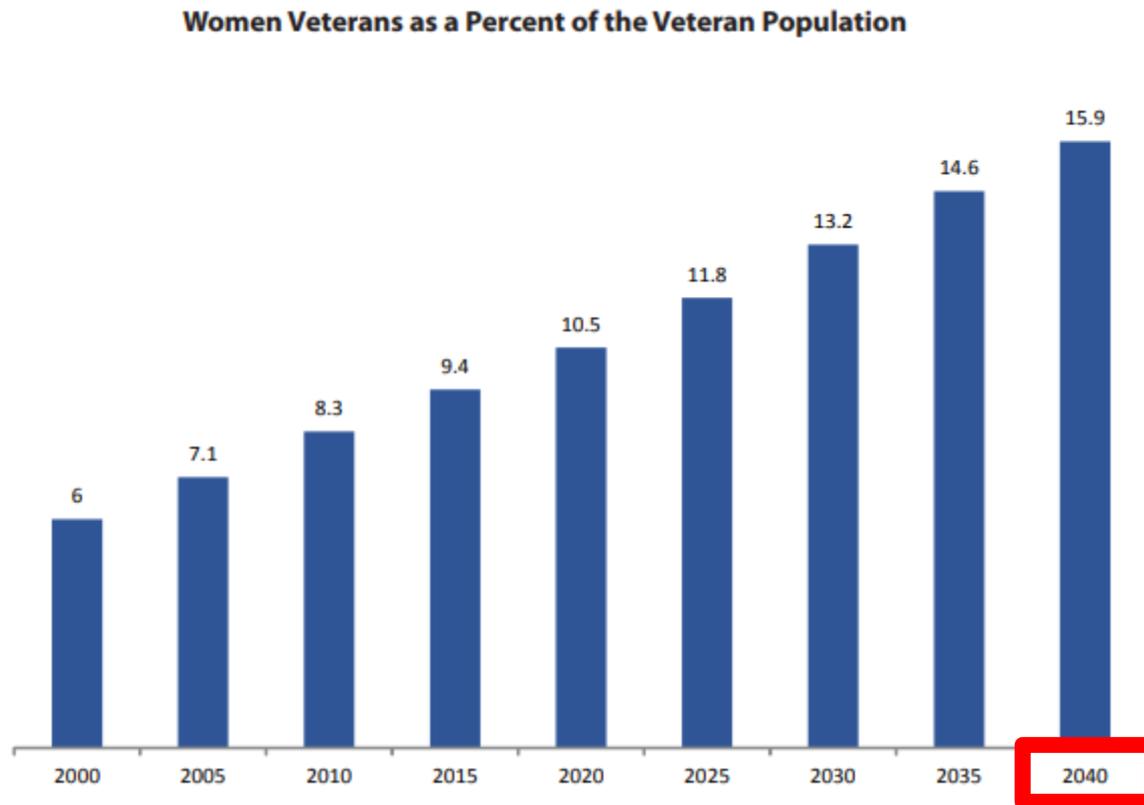05/2019

Why is USVETS produced?

- To support operational and policy issues throughout the Department of Veterans Affairs
  - Their focus is larger than the Veterans Health Administration
  - USVETS is their primary data source for the DGA
- Examples of their work:
  - VetPop: official Veteran population projection from the Department of Veterans Affairs (VA).   VetPop is used for strategic, policy planning, and budgeting within VA and by external organizations such as other federal agencies, Congress, state governments.
  - Women Veterans Report: This report summarizes the history of women in the military and as Veterans, who our women Veterans are, how their military service affects their post-military lives, and how they can be better served based on these insights.

Examples (cont.)

Figure 1a.
Female Active-Duty Military Personnel 1945 to 2015

Legend: Total Women, Officer, Enlisted

Source: Department of Defense, Defense Manpower Data Center, Statistical Information Analysis Division
Prepared by the National Center for Veterans Analysis and Statistics

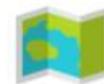https://www.va.gov/vetdata/docs/specialreports/women_veterans_2015_final.pdf   05/2019

Examples (cont.)

Women Veterans as a Percent of the Veteran Population

05/2019

https://www.va.gov/vetdata/

How is USVETS produced?

- Acquired from over 35 data sources including…
  - Department of Defense (DoD)
  - Veterans Health Administration (VHA)
  - National Cemetery Administration (NCA)
  - Veterans Benefits Administration (VBA)
  - Social Security Administration (SSA)
- Aggregated by a transparent data quality framework
- Hosted on the BISL SAS grid (currently non-research)

USVETS roadmap

- Story behind USVETS
- **Quality**
  - Accuracy
  - Validity
  - Timeliness
  - Completeness/Consistency/Traceability/Uniqueness
- Layout of USVETS data
- Specific fields in USVETS
- Access

Data Quality

- "USVETS must be considered the gold standard for Veteran data." **United States Veterans Eligibility Trends and Statistics (USVETS) Support USVETS Data Quality Assessment Version 3.1, 20 MAR 2019**

- Issues with aggregating multiple data sources

  - Multiple data sources contain the same data element → transparent adjudication process including limitations

  - Single data source contains unique data element → data may be unavailable for some of the study population

- Operate within a data quality framework: accuracy, completeness, consistency, traceability, uniqueness, validity, and timeliness

Accuracy: USVETS vs. sources

	USVETS Gender	Source 1 Gender	Source 2 Gender	Source 3 Gender	Source 4 Gender	USVETS Source 1 Match	USVETS Source 2 Match	USVETS Source 3 Match	USVETS Source 4 Match	Source Concordance
Veteran 1	M	M	M	Z	F	1	1	0	0	50%
Veteran 2	M	M	M	M	M	1	1	1	1	100%
Veteran 3	F	F	Z	F	F	1	0	1	1	75%
Veteran 4	Z	Z	Z	Z	F	1	1	1	0	75%
Veteran 5	M	M	Z	M	M	1	0	1	1	75%
								Average Concordance		75%

Variables	Average Concordance Score	Average Number of Sources	Average Number of Matches	Number of Sources
Gender	89.9%	4.5	4.3	22
Race	71.2%	3.7	3.2	11
Ethnicity	80.2%	3.7	3.7	9
DOB – day	96.9%	6.2	6.0	25
DOB – month	81.5%	6.2	5.2	25
DOB – year	97.6%	6.2	6.1	25
DOD – day	92.3%	1.6	3.0	10
DOD – month	99.6%	1.6	3.1	10
DOD – year	99.2%	1.6	3.1	10
First Name	97.5%	5.2	5.0	27
Middle Name	78.5%	5.2	4.3	27
Last Name	97.8%	5.2	5.1	27

- All categorical variables from multiple data sources in USVETS
- USVETS generally chooses the most representative value
- They may impute unknown values (e.g., name → gender)

05/2019

Validity

- Does the data make sense?
  - An address has a state abbreviation that does not exist
  - The Veterans activation date was prior to their birth
- Does the data conform to business rules?

3.2.3 Date of Birth Data Quality

TABLE 14. DATE OF BIRTH METADATA

USVETS Version	2017v3 E:
Date of Last Assessment	01-29-2019
USVETS Field	DOB
Valid Values	MM/DD/YYYY
Format	MMDDYY10
Data Source Selection Preference	SSA_V → MVI → Acxiom-Demo → VADIR → OEFOIF → GW1 → CP → VETSNET → ZERO → VRE → LGY → SAHSHA → VALI → VGLI → CH30 → CH33 → CH32 → BDN → NCA Demo → NCA User → VADS → BIRLS → SSAD → ENRL → DMDC
Business Rule	Use DOB to create DOB_YR, DOB_DY, and DOB_MO. (1) Use SSAV DOB if SSAV_verification_code in('V', '2') and DOB ne ''. (2) Else, most frequent DOB from raw files as done for name (at least 3 different sources). (3) Else, selection preferences. (4) Date imputations for valid years: (a) Day: when missing → '15', when greater than 31, 30, or 28/29 (depending on the month) → 'the latest proper day in the month' (b) Month: missing or invalid → '7' (c) Year: no imputation if not legitimate 4-digit → 'blank' (5) If age is too young, at time of running ETL from steps 2–4, then use next appropriate source for DOB.

Timeliness

USVETS is comprised of many different data sets sourced from VA, other Government agencies, and external organizations. These files are sent to or acquired by DGA following the schedule in Table 7.

TABLE 7. USVETS DATA AVAILABILITY SCHEDULE

USVETS Data Availability & ETL Schedule											
Quarter	**Month**	**BIRLS, VADS**	**SSA_DEATH, OEFOIF**	**C&P**	**VADIR**	**NCA, VBA**	**ENR, VHA, VGLI**	**CMS, MVI, ADR, NDI (SDR)**	**SSA Validation**	**3rd Party Data**	**AMF**	**ETL**
		Monthly	monthly	annual	annual	annual	annual	ad hoc		annual	annual	
Q1	October	All monthly files are available by October.	X	X								V1
	November		X		X							
	December		X			X						
Q2	January		X									V2 / V2.1
	February		X									
	March		X									
Q3	April		X				X					V2 / V2.1
	May		X						X*			
	June		X							X*		
Q4	July		X								X*	V3 (Final)
	August		X									
	September		X									
											* Desired Receipt Date	

The timely availability of source data is often out of the control of DGA and has been delayed several months in some cases. Currently, DGA does not officially measure the timeliness of source data.

Completeness, Consistency, Traceability, Uniqueness

- Completeness: contextual to the variables
  - A missing date of death could mean the Veteran is alive.
- Consistency
  - If the Veterans is known to be dead, but lacks a date of death
- Traceability: original data source
- Uniqueness: single record per SSN

Limitations

- USVETS is designed for population-based assessments, not individuals

  - Race and ethnicity may come from purchased data sources, if no better source exists

    - Researchers may prefer CDW for race and ethnicity

    - USVETS had Social Security Administration (SSA) review SSN, name, gender, and DOB

  - Gender may be determined from the name, if no data source exists

USVETS roadmap

- Story behind USVETS
- Quality
- **Layout of USVETS data**
- Specific fields in USVETS
- Access

How is USVETS data arranged?

- Statistical Analysis System (SAS) data sets
- One dynamic file per fiscal year (n=13; FY2005–2017)
- One master file
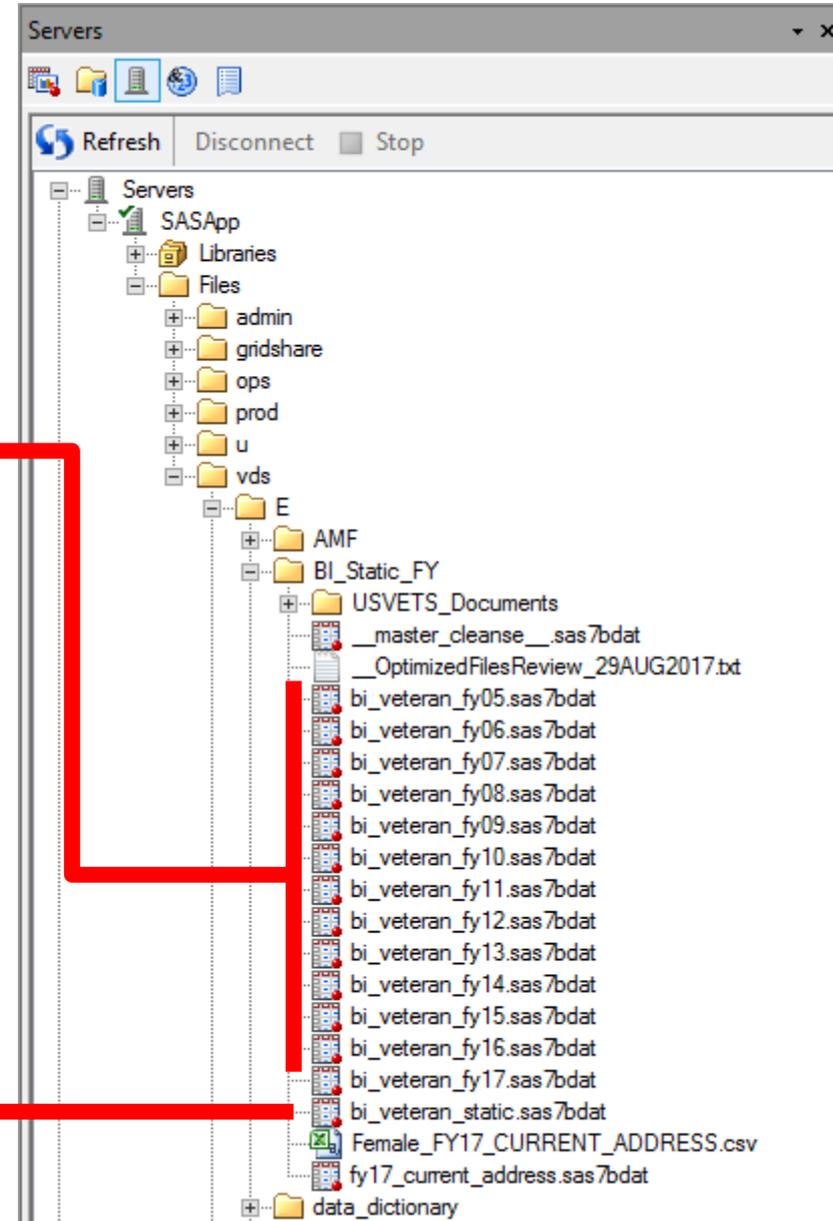
BI_VETERAN_STATIC

- Name (first, last)
- Date of birth
- Date of death
- Gender
- Race
- Ethnicity
- SSN
- "Data Indicators" – year of inclusion of data, by fiscal year

USVETS roadmap

- Story behind USVETS
- Quality
- Layout of USVETS data
- **Specific Fields in USVETS**
  - **Where Veterans Live**
  - **Who Lives with Veterans**
  - **Veteran Finances**
  - **Veteran Education and Employment**
  - **Health Insurance**
  - **Benefits & Utilization**
  - **Enrollment in Veteran-Specific Programs**
- Access

Where Veterans Live

- Where do they live?
  - State name, abbreviation, state FIPS code
  - County name, county FIPS code
  - ZIP & ZIP+4 codes
  - Latitude/longitude of their place of residence
- Community information
  - Rural-urban commuting area (RUCA) codes
  - Y/N – is their living area considered urban?
  - Congressional district of residence
  - VISN code/market/sector/submarket
  - Distance from VA inpatient and outpatient facilities

https://www.va.gov/vetdata/images/header004.png
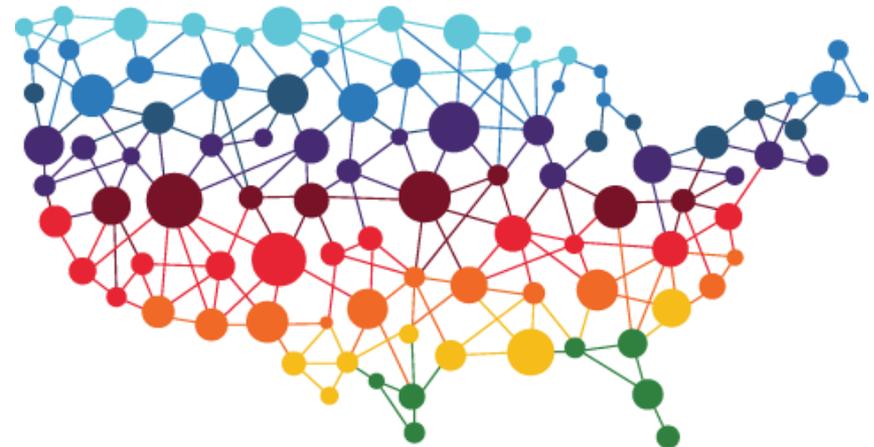
Who Lives with Veterans

- Home owner or renter
  - O = Owner
  - R = Renter
- Number of people in Household
  - Single parent?
  - Children?
  - Seniors?
- Age ranges of adults and children in household

https://101clipart.com/wp-content/uploads/01/Free%20Family%20Clipart%20Images%2001.jpg
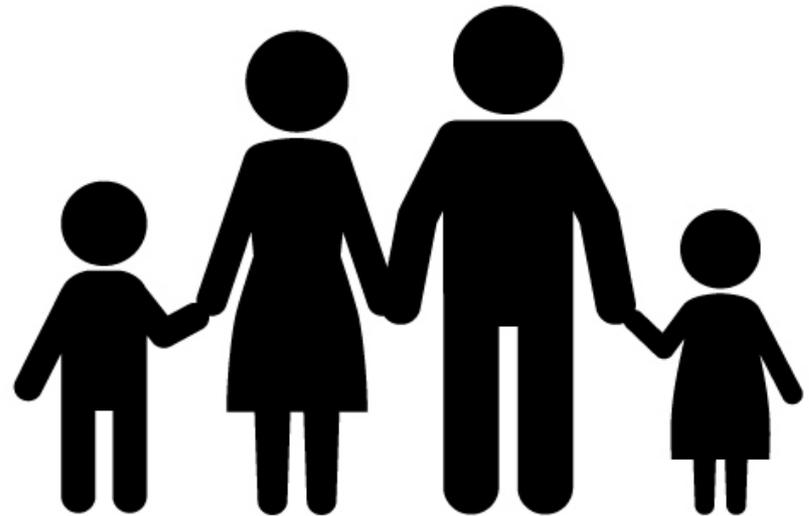
Percent of Veteran Household's with Children

Legend:
0% - 14.1%
14.2% - 25.3%
25.4% - 29.5%
29.6% - 32.9%
33% - 37%
37.1% - 80%

Alaska

Hawaii

Puerto Rico
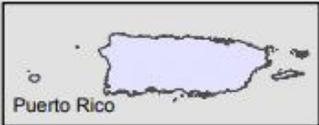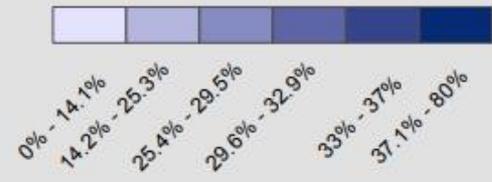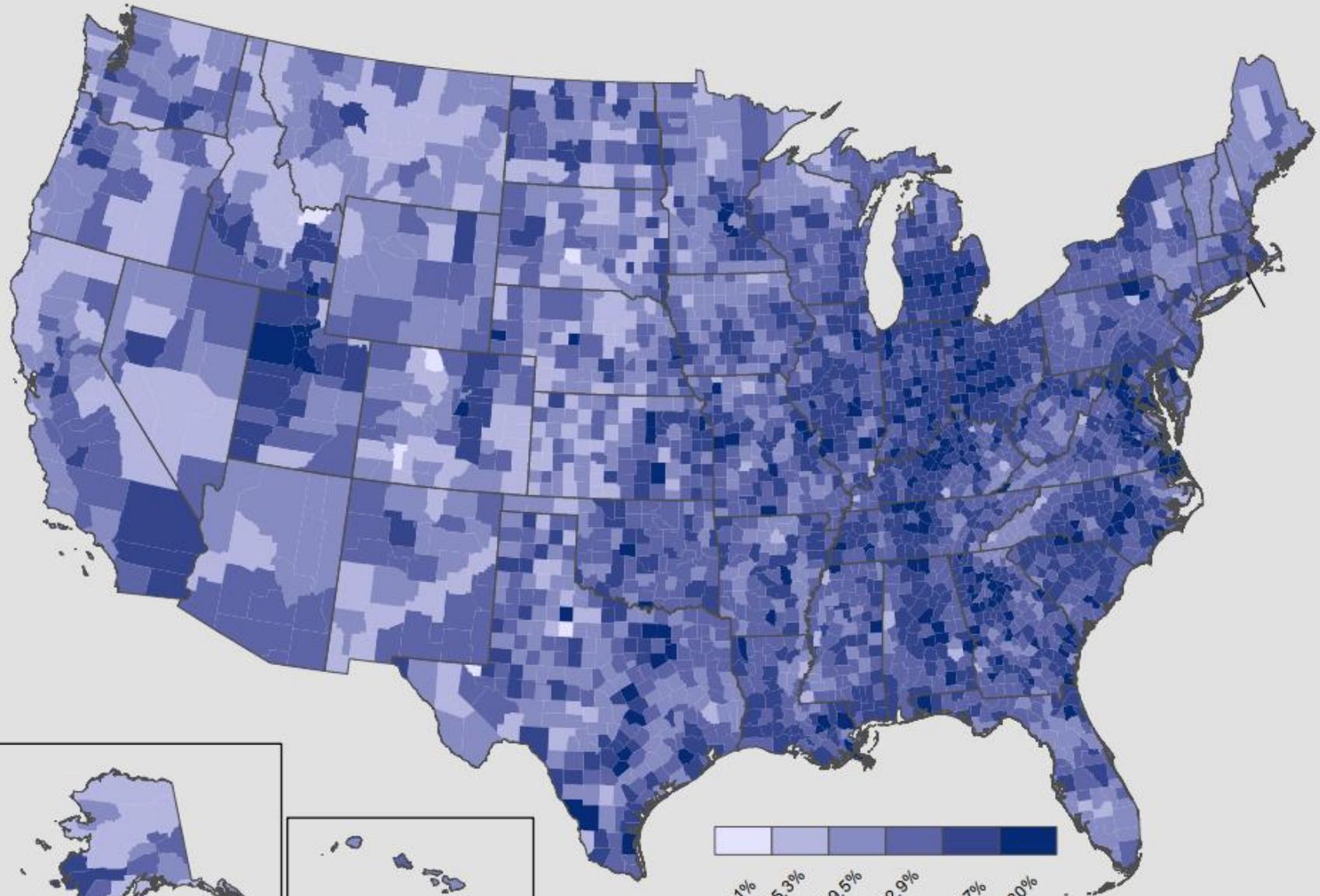
Veteran Finances

- Are they considered competent to receive payments directly? Y/N

- Do they receive compensation/pension from VA? Y/N

- Estimated household income – range and dollar amount

Veteran Education & Employment

- Highest education level:
  - 1 = Completed high school
  - 2 = Completed college
  - 3 = Completed graduate school
  - 4 = Attended vocational/technical school
- Does the Veteran own a business? If so, what type?
  - A, B, C, D, E, F, M, O, P, S – alpha codes for type
- Occupational information: alphanumeric code detailing the occupation of the individual
  - E.g. A000, G002

Health Insurance

- Health insurance coverage indicator – calculation based on multiple demographic variables to determine the propensity of having major medical insurance

- Active health insurance in FY, private or Medicare – Y/N

- Life insurance user indicator – Y/N if SSN in VGLI/VALI

- VA insurance policy amount – $$$ amount

- VHA active insurance – Y/N

- Medicare enrollment – Y/N if enrolled in parts A/B/C/D

Benefits & Utilization

- Number of service-connected conditions
- Did the veteran use one or more VA benefits or services?
- Date the veteran enrolled in healthcare benefits
- What kind of care did the Veteran receive in the FY?
  - Inpatient/outpatient, VA and non-VA
  - How many times and for how long?
- Care cost information:
  - Numerical value of VA inpatient and outpatient care
  - Numerical value of non-VA inpatient and outpatient care
- Pharmacy utilization
  - Did the veteran receive VA pharmacy care in the FY?
  - How much cost incurred by inpatient pharmacy?
  - How much cost incurred by outpatient pharmacy?

Enrollment in Veteran-Specific Programs

- Benefits Delivery Network (BDN)
- Montgomery GI Bill-Selected Reserve (MGIB-SR)
- Reserve Educational Assistance Program (REAP)
- Montgomery GI Bill – Active Duty (MGIB-AD)
- Post-9/11 Education Benefit (Chapter 33)
- Compensation & Pension (CP)
- Automated Enrollment Entry (ENR)

- Loan Guaranty (LGY)
- National Cemetery Administration (NCA)
- Special Adaptive Housing/Special Housing Allowance (SAH/SHA)
- Veterans Affairs Learning Infrastructure (VALI)
- Veterans Service Network (VETSNET)
- Veterans Group Life Insurance (VGLI)
- Veterans Health Administration (VHA)
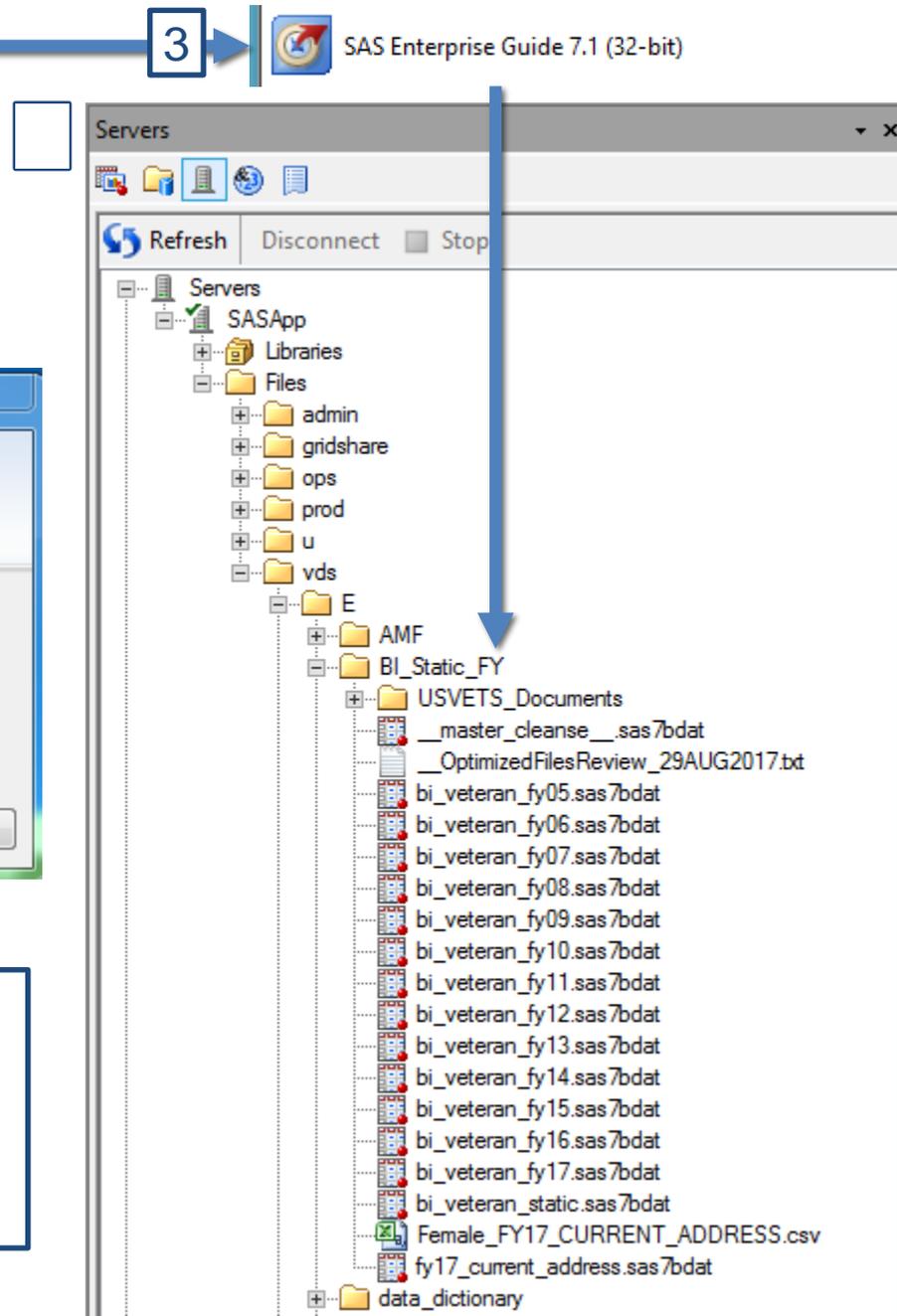- Vocational Rehabilitation and Employment (VR&E)

USVETS roadmap

- Story behind USVETS
- Layout of USVETS data
- Specific fields in USVETS
- **Access**

Access

1 Remote Desktop Connection ▶

2

Remote Desktop Connection

Remote Desktop Connection

Computer: vhacdwapp15.vha.med.va.gov

User name: None specified

You will be asked for credentials when you connect.

Show Options   Connect   Help

Hosted on the BISL SAS grid
(operational access)

3 ▶ SAS Enterprise Guide 7.1 (32-bit)

Servers

Refresh   Disconnect   Stop

- Servers
  - SASApp
    - Libraries
    - Files
      - admin
      - gridshare
      - ops
      - prod
      - u
      - vds
        - E
          - AMF
          - BI_Static_FY
            - USVETS_Documents
            - __master_cleanse__.sas7bdat
            - __OptimizedFilesReview_29AUG2017.txt
            - bi_veteran_fy05.sas7bdat
            - bi_veteran_fy06.sas7bdat
            - bi_veteran_fy07.sas7bdat
            - bi_veteran_fy08.sas7bdat
            - bi_veteran_fy09.sas7bdat
            - bi_veteran_fy10.sas7bdat
            - bi_veteran_fy11.sas7bdat
            - bi_veteran_fy12.sas7bdat
            - bi_veteran_fy13.sas7bdat
            - bi_veteran_fy14.sas7bdat
            - bi_veteran_fy15.sas7bdat
            - bi_veteran_fy16.sas7bdat
            - bi_veteran_fy17.sas7bdat
            - bi_veteran_static.sas7bdat
            - Female_FY17_CURRENT_ADDRESS.csv
            - fy17_current_address.sas7bdat
          - data_dictionary

Research Access

Email

- Pheakdey.Lim1@va.gov
- Ronald.Hauser@va.gov

Contact information

R. George Hauser, MD
Ronald.Hauser@va.gov
VA Connecticut Healthcare System

Ankur Bhargava, MD
Ankur.Bhargava@va.gov
VA Connecticut Healthcare System

Next session:
June 3rd @ 1 pm Eastern

Database & Methods Cyberseminar Series

Session #9: *Diagnoses, Labs, and Meds, Oh My! Using CDW to Define a Study Cohort by Multiple Criteria*

Peter A. Richardson, PhD
Kathryn E. Royse, PhD
Center for Innovations in Quality, Effectiveness and Safety (IQuESt)

VIReC
RESEARCHERS' GUIDE TO VA DATA

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

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.vhadataportal.med.va.gov/Home.aspx

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

Health Economics Resource Center (HERC): http://vaww.herc.research.va.gov (VA Intranet)

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

Archived cyberseminar: What can the HSR&D Resource Centers do for you?
http://www.hsrd.research.va.gov/for_researchers/cyber_seminars/archives/video_archive.cfm?SessionID=101