



New linked data from the Data Linkage Program: Veterans Health

Carla E. Zelaya, Health Statistician, Data Linkage Program
Data Linkage Methodology and Analysis Branch
Division of Analysis and Epidemiology

NCHS Webinar
September 19, 2022

National Center for Health Statistics (NCHS)

- Nation's principal health statistics agency
- One of 13 principal federal statistical agencies
- **Mission:** To provide timely, relevant statistical information that can be used to guide actions and policies to improve the health of the American people



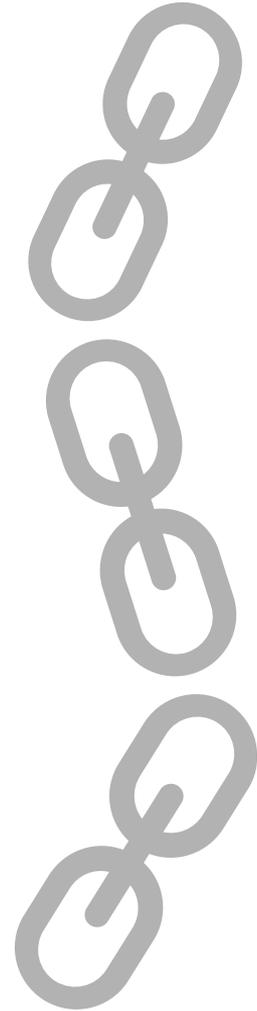
NCHS Data Linkage Program: Overview

- Create linked data files that support high quality research and program evaluation
- Utilize state of the art linkage methodologies and provide documentation and support for analyzing linked data files
- Explore innovative methods for maintaining researcher access to linked data

Why we link

Linking data is a powerful and efficient mechanism for producing policy-relevant information

- *Brings together additional information about one person or entity to create a new, richer dataset*
- *Allows for the construction of longitudinal events in previously cross-sectional datasets.*



What surveys we link

National Health Interview Survey (NHIS)

A nationally representative, cross-sectional sample of the US civilian noninstitutionalized population, which includes a household interview that serves as an important source of information on the nation's health

National Health and Nutrition Examination Survey (NHANES)

A nationally representative, cross-sectional sample of the US civilian noninstitutionalized population, which includes a household interview followed by an examination in a mobile examination center (MEC) that serves as an important source of information on the health and nutritional status of adults and children

National Hospital Care Survey (NHCS)

NHCS collects data on patient care in hospital-based settings (inpatient, emergency, and outpatient departments) to describe patterns of health care delivery and utilization in the US

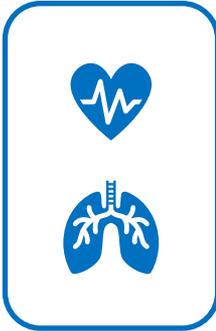
What we link (continued)

Survey Data

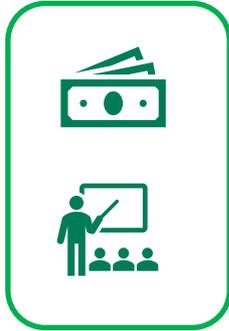
Sampling frame
Known inference



Health behaviors



Health conditions



Socioeconomic status



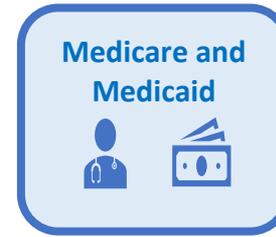
Health care access and utilization

Administrative Data

Program participation/vital status
Not designed for research purposes



Housing and Urban Development



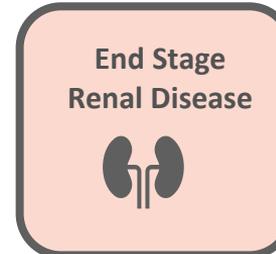
Medicare and Medicaid



Mortality



Geocoded Addresses



End Stage Renal Disease

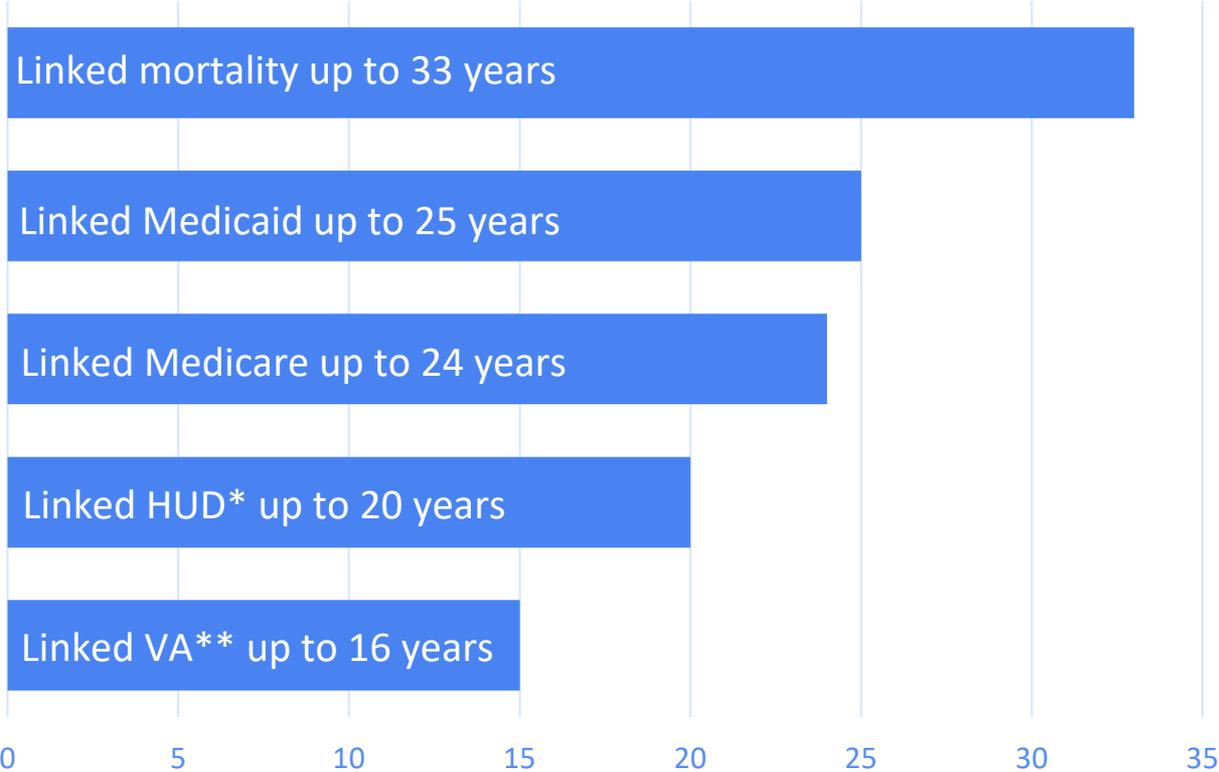


Department of Veterans Affairs

How we link

- Linkage conducted in two passes:
 1. Deterministic match using Social Security Number (SSN) collected in the survey
 - Identifier fields such as name, state of residence, and date of birth are compared to validate
 - This dataset becomes the “truth deck”
 2. Probabilistic matching techniques used to identify likely pairs using other identifiers (not SSN)
 - Pair scores are calculated on the agreement status of the identifiers such as name, state of residence, and date of birth
 - SSN is not used to score pairs instead, it is used to measure linkage accuracy (when available)

Years of follow-up with NCHS linked data



<https://www.cdc.gov/nchs/data/datalinkage/LinkageTable.pdf>

*HUD, Department of Housing and Urban Development; **VA, Department of Veteran Affairs

**First linkage of NCHS survey data with
Department of Veterans Affairs (VA)
administrative data**

Linking NCHS and VA data

- Conducted through a collaborative agreement between the NCHS Data Linkage Program and the VA
- VA administrative data
 - United States Veterans Eligibility Trends and Statistics (USVETS) database, prepared by the Office of Data Governance and Analytics within the Office of Enterprise Integration at the VA
- Linked NCHS data
 - 2005-2018 NHIS (sample adults) and 2005-2018 NHANES (adults 18 and over)
 - 2016 NHCS restricted data files available in fall 2022
- Supported in part with funding from the CDC Data Modernization Initiative

NCHS-VA Linked Data Files to support research objectives

Survey data

- Sex
- Race and ethnicity
- Urban/rural
- Health care access
- Preventative care
- Chronic conditions
- Health behaviors



VA administrative data

- Service-connected disability
- VA health care enrollment
- VA benefit program utilization



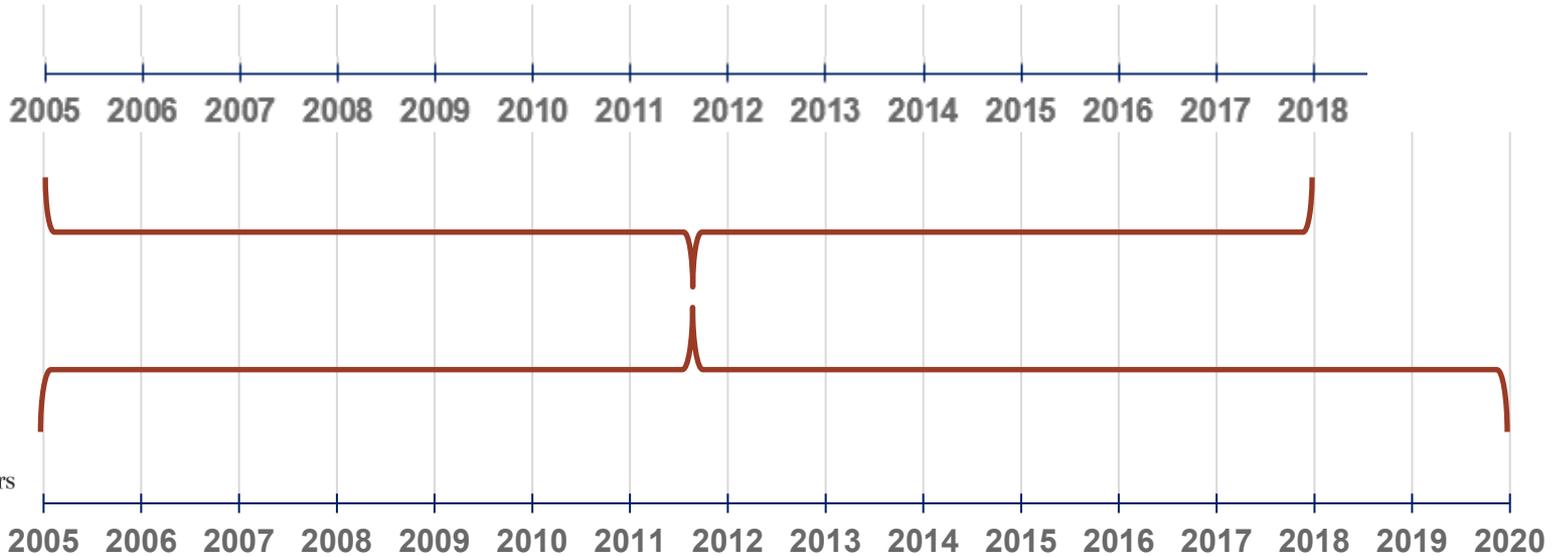
Research topic areas

- Veterans compared to nonveterans
- Health equity/social determinants of health
- Inform VA policy

NHIS and NHANES Linked VA Data Files: Survey and administrative years



VA U.S. Department of Veterans Affairs



Note: Linkage of the 2014 and 2016 NHCS to VA data is forthcoming

Sample sizes and match rates: NHIS

	Total sample	Eligible for linkage		Linked to VA	
		n	Percent of total	n	Percent of eligible
2005-2018 Sample Adult NHIS participants					
Total	413,417	334,701	81.0	29,357	8.8
18-64	320,151	258,641	80.8	17,756	6.9
65 and over	93,266	76,060	81.6	11,601	15.3
Male	184,243	149,898	81.4	26,159	17.5
Female	229,174	184,803	80.6	3,198	1.7

- Match rate tables: <https://www.cdc.gov/nchs/data-linkage/va-methods.htm>

Sample sizes and match rates: NHANES

	Total sample	Eligible for linkage		Linked to VA	
		n	Percent of total	n	Percent of eligible
2005-2018 NHANES participants aged 18 and over					
Total	42,143	37,595	89.2	3523	9.4
18-64	32,441	28,856	89.0	1,924	6.7
65 and over	9,702	8,739	90.1	1,599	18.3
Male	20,458	18,337	89.6	3,236	17.7
Female	21,685	19,258	88.8	287	1.5

- Match rate table: <https://www.cdc.gov/nchs/data-linkage/va-methods.htm>



NCHS-VA Linked Match Status File

- Identifies which survey participants were eligible for linkage and linked to VA administrative records (Match status variable)
- Contains one record for each 2005-2018 NHIS sample adult and each 2005-2018 NHANES participant aged 18 and older

Other variables include:

- Linkage eligibility-adjusted sample weights
- Match probability



NCHS-VA Linked Service Record File

- Includes information detailing military service in the US uniformed services
- Contains a record for each linked participant (Veteran)

Variables include:

- Branch of service
- Character of service at discharge
- Era of service
- Race and ethnicity information from VA administrative records

NCHS-VA Linked VA Utilization File



- Includes VA administrative information
- One record for each linked participant (Veteran)
- Variables provided for each Fiscal Year (FY) 2005-2020

Variables include:

- Healthcare enrollment priority status
- Number of service-connected disabilities
- VA health care enrollment
- Benefit program utilization: VA health care, service-connected disability compensation, pension, VA guaranteed home loan, life insurance, education, training, veteran readiness (vocational rehabilitation) and employment

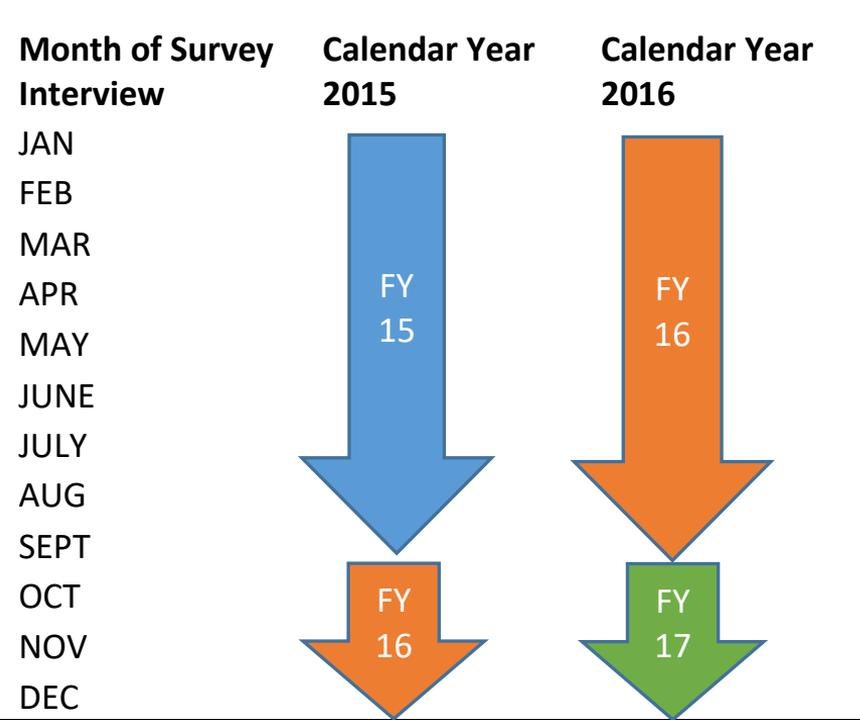
NCHS-VA Linked VA Utilization File (continued)

Timing of benefit program utilization in relation to NCHS survey:

- NCHS survey data linked to multiple years of VA administrative data through fiscal year 2020
- VA benefit program utilization data may be available for survey participants concurrent to, before, or after the survey interview (or exam for NHANES) date

NCHS-VA Linked VA Utilization File (continued)

Figure 1. Relationship of Calendar Year and Fiscal Year (FY)

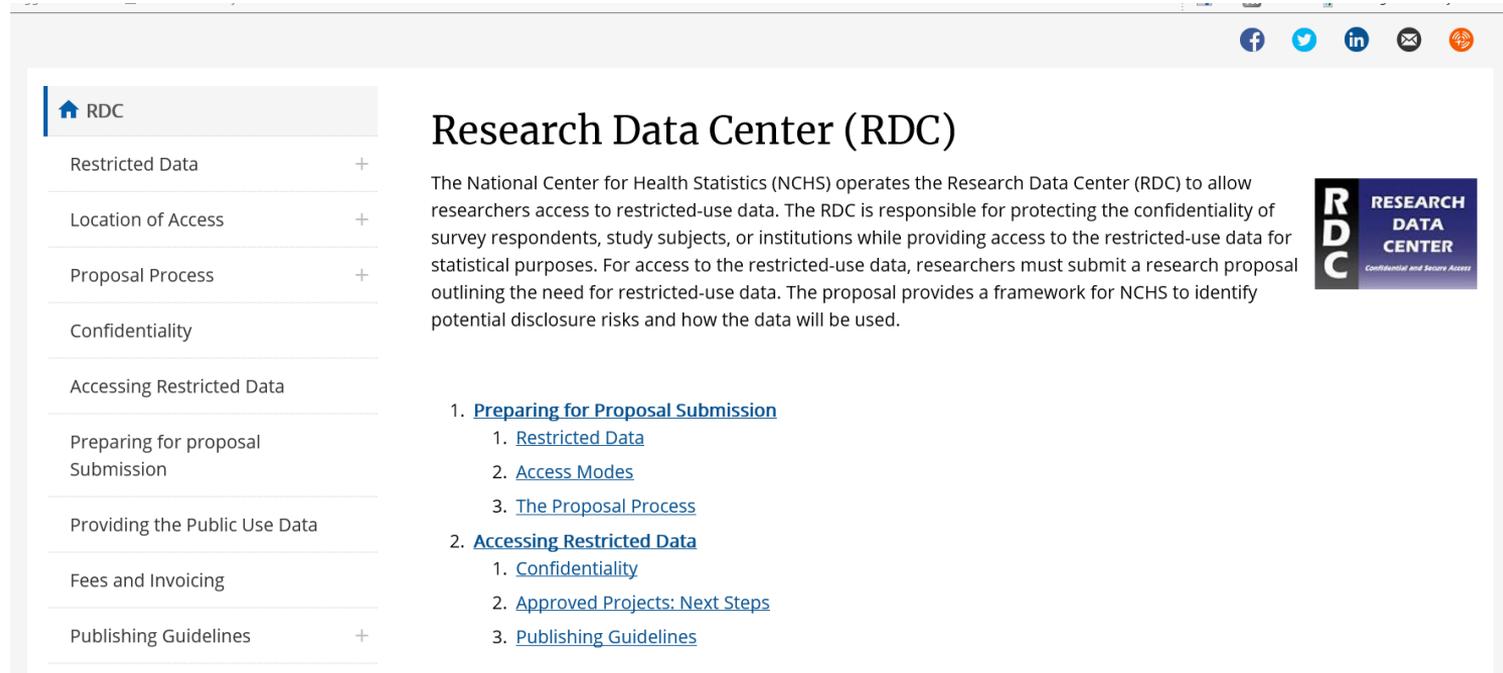


Alignment of FY and Survey Year(s):

- Surveys are conducted on calendar year basis.
- VA administrative data are organized by FY (Oct 1 to Sep 30)
- Example of NHIS alignment (yearly)
 - To analyze concurrent VA utilization information for NHIS 2015 – use FY 2015 and FY 2016 variables
 - If interview occurred in Q4 2015, then the corresponding VA data FY would be 2016

Restricted-use NCHS-VA Linked Data Files

- NCHS-VA Linked Data Files will be available through the NCHS Research Data Center (RDC) and Federal Statistical RDCs managed by Census



Research Data Center (RDC)

The National Center for Health Statistics (NCHS) operates the Research Data Center (RDC) to allow researchers access to restricted-use data. The RDC is responsible for protecting the confidentiality of survey respondents, study subjects, or institutions while providing access to the restricted-use data for statistical purposes. For access to the restricted-use data, researchers must submit a research proposal outlining the need for restricted-use data. The proposal provides a framework for NCHS to identify potential disclosure risks and how the data will be used.

1. [Preparing for Proposal Submission](#)
 1. [Restricted Data](#)
 2. [Access Modes](#)
 3. [The Proposal Process](#)
2. [Accessing Restricted Data](#)
 1. [Confidentiality](#)
 2. [Approved Projects: Next Steps](#)
 3. [Publishing Guidelines](#)

<https://www.cdc.gov/rdc/index.htm>

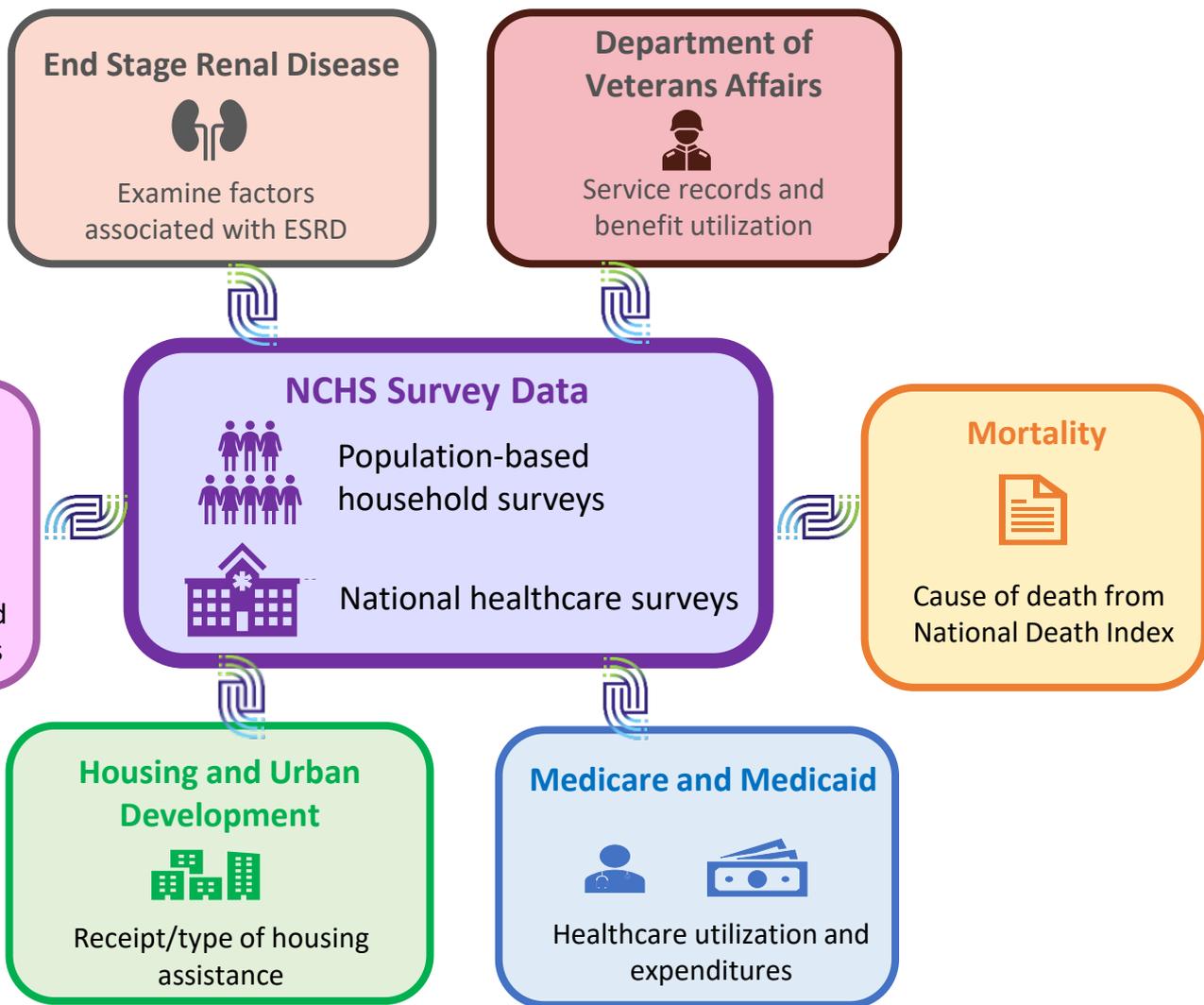
Email: rdca@cdc.gov

Key survey variables to consider for research proposals

The following NCHS variables are suggested for RDC proposals:

- Sample weights and design variables are needed to account for the complex design of the NCHS surveys
- Demographic information such as sex, age, race and ethnicity
- Socio-economic information such as income and education
- Date of survey:
 - NHANES month and year of examination and interview (restricted-use variables)
 - NHIS calendar quarter and year of interview

Additional linked data



Summary

Key summary points

- The NCHS Data Linkage Program continually updates and maintains linkages between NCHS survey data and administrative data
- 2005-2018 NHIS and 2005-2018 NHANES have been linked to VA administrative data through fiscal year 2020 (<https://www.cdc.gov/nchs/data-linkage/va.htm>), and data files are now available through the RDC
- NCHS-VA Linked Data Files provide a new policy relevant resource that combines information on health characteristics and access to care with veteran status, service history, and VA benefit program utilization



NCHS Data Linkage Program

Contact: datalinkage@cdc.gov; www.cdc.gov/nchs/data-linkage

Subscribe to the NCHS Data Linkage Program LISTSERV to receive updates! Email a message to list@cdc.gov Leave the subject line blank. In the body of the message, type or paste:

SUBSCRIBE NCHSDATALINKAGEPROGRAM lastname,firstname

where 'lastname,firstname' is your last and first name.

For more information, contact CDC
1-800-CDC-INFO (232-4636)
TTY: 1-888-232-6348 www.cdc.gov

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.



Questions

- Please submit your questions via the Q & A feature in the Zoom application
- The facilitator will address questions as time allows
- Questions not answered may be forwarded to paoquery@cdc.gov

For more information, contact CDC
1-800-CDC-INFO (232-4636)
TTY: 1-888-232-6348 www.cdc.gov

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

