

Rocky Mountain Research Data Center (RMRDC):  
Advantages of Using Restricted Data Collected by Federal Statistical Agencies

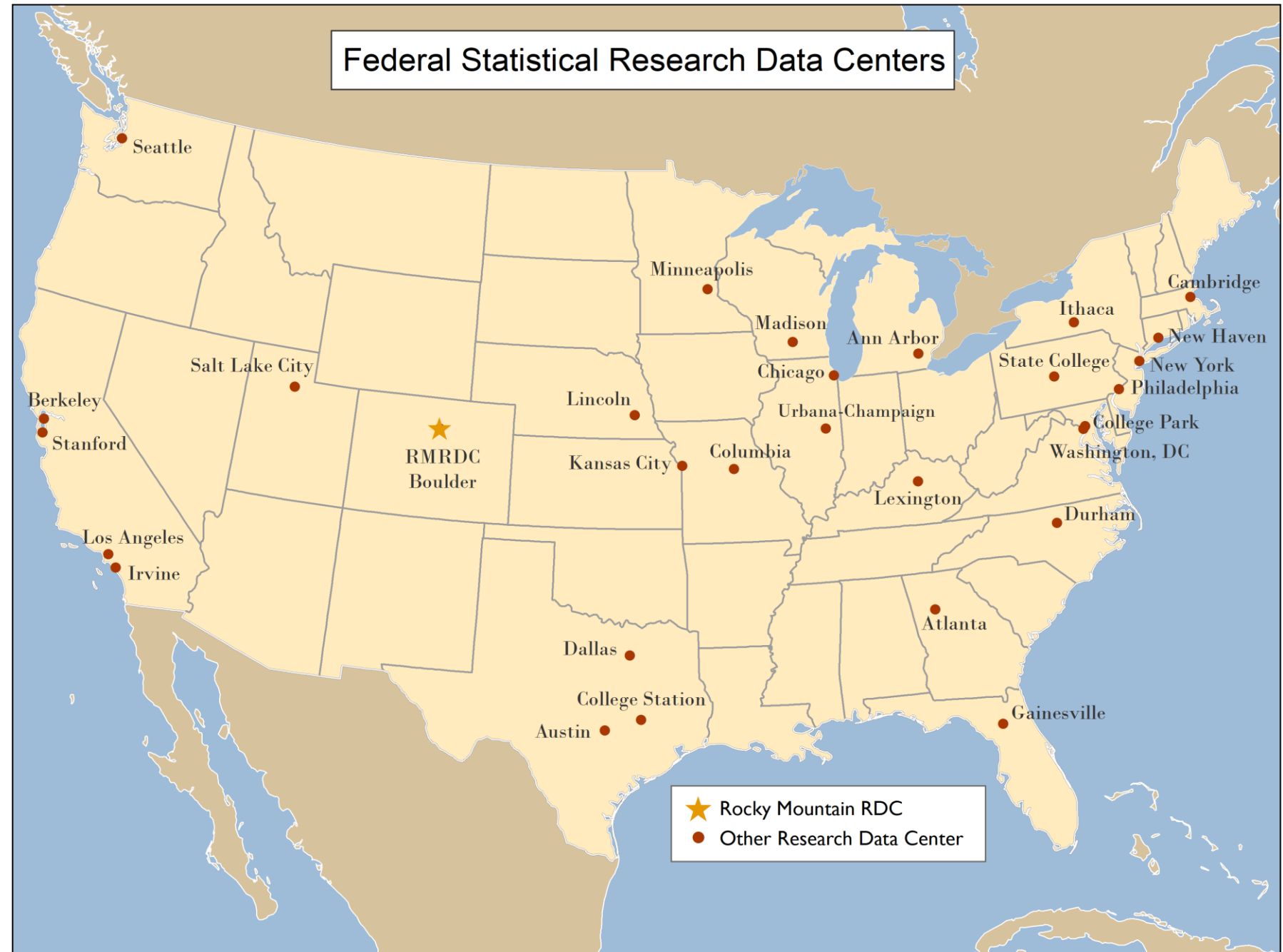
<https://www.colorado.edu/rocky-mountain-research-data-center/>

Jani Little, Executive Director

[jani.little@colorado.edu](mailto:jani.little@colorado.edu)



**The RMRDC is  
part of a  
National  
Network of  
FSRDCs  
Providing  
Access to  
Restricted-Use  
Data Collected  
by Federal  
Statistical  
Agencies**



# RMRDC Membership Network

\*All faculty, staff, and graduate students associated with member universities have free access to the lab and services



University of Colorado  
Boulder



Denver



Anschutz



Colorado  
State  
University



UNIVERSITY of  
DENVER

— **ROCKY MOUNTAIN** —  
**POISON & DRUG SAFETY**



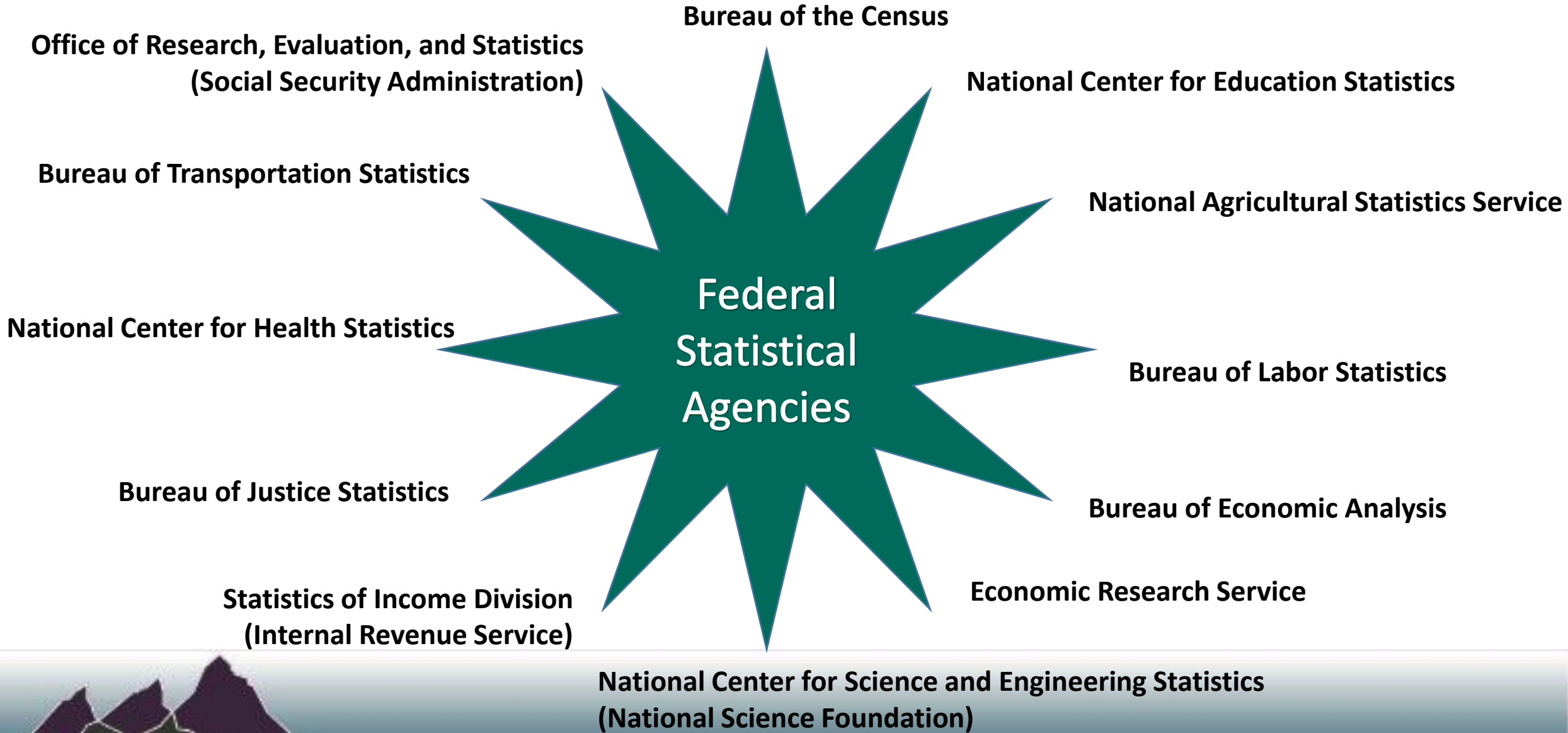
# What RMRDC offers:

- Access to federal restricted data and staff assistance in a secure facility

## Staff assist with:

- Successful proposal development and submission
  - Researchers must develop an agency-approved proposal to use the data.
- “Special Sworn Status (SSS)” applications
  - Researchers undergo a background check which gets them lab access
- Technical support
- Disclosure review and statistical output approval
  - Results must be formally reviewed for disclosure violation before they leave the secure facility.





# Foundation of Evidence-Based Policymaking Act of 2018

- This bill requires agency data to be accessible and requires agencies to plan to develop statistical evidence to support policymaking.
- The bill establishes in the OMB a Chief Data Officer Council for establishing government-wide best practices for the use, protection, dissemination, and generation of data and for promoting data sharing agreements among agencies.



# Restricted-Use Data Addressing Society's Most Pressing Challenges

Local Labor  
Markets

Immigration  
Impacts on People  
and Business

Natural  
Hazards  
Impacts on People  
and Business

International  
Trade Policies

Health  
Disparities

Crime  
Victims and  
Consequences

Rural vs Urban  
Disparities

Drug Use and  
Mental Health

Business  
Sector  
Growth and Decline



# Advantages of Restricted-Use Data

- **For Researchers:**

- Greatly expands the policy and basic questions that can be addressed
- Builds on past research findings with richer data
- Improves competitive edge for grants, publications, recruitment
- Improves graduate education (big data/statistical techniques) and placement

- **For Research:**

- Microdata not available in public-use data, e.g., business data
- Longitudinal data collections, e.g., NCES and NLS
- Variables not available in public-use versions of data sets, such as low-level geography
- Full population counts or larger samples (Decennial Census, ACS, CPS) not available in public-use data
- Linked restricted-use Census data sets
- Full range of response items (e.g., ICD codes, industry codes, occupational codes, detailed race answers, income is not top-coded, etc.)





**Public-Use Microdata:** created so that individual records cannot be linked to PII or to other databases

1. No obvious identifiers or PII (personally identifiable information)
2. Includes only a portion of the surveyed sample
3. Includes only a selection of variables. Sensitive variables removed.
4. Limits geographic detail
5. Limits the number and detailed breakdown of categories within variables on the file, e.g., race and ethnicity, exact dates of events, ICD-10 codes
6. Truncate extreme codes for certain variables (top or bottom coding), e.g., income

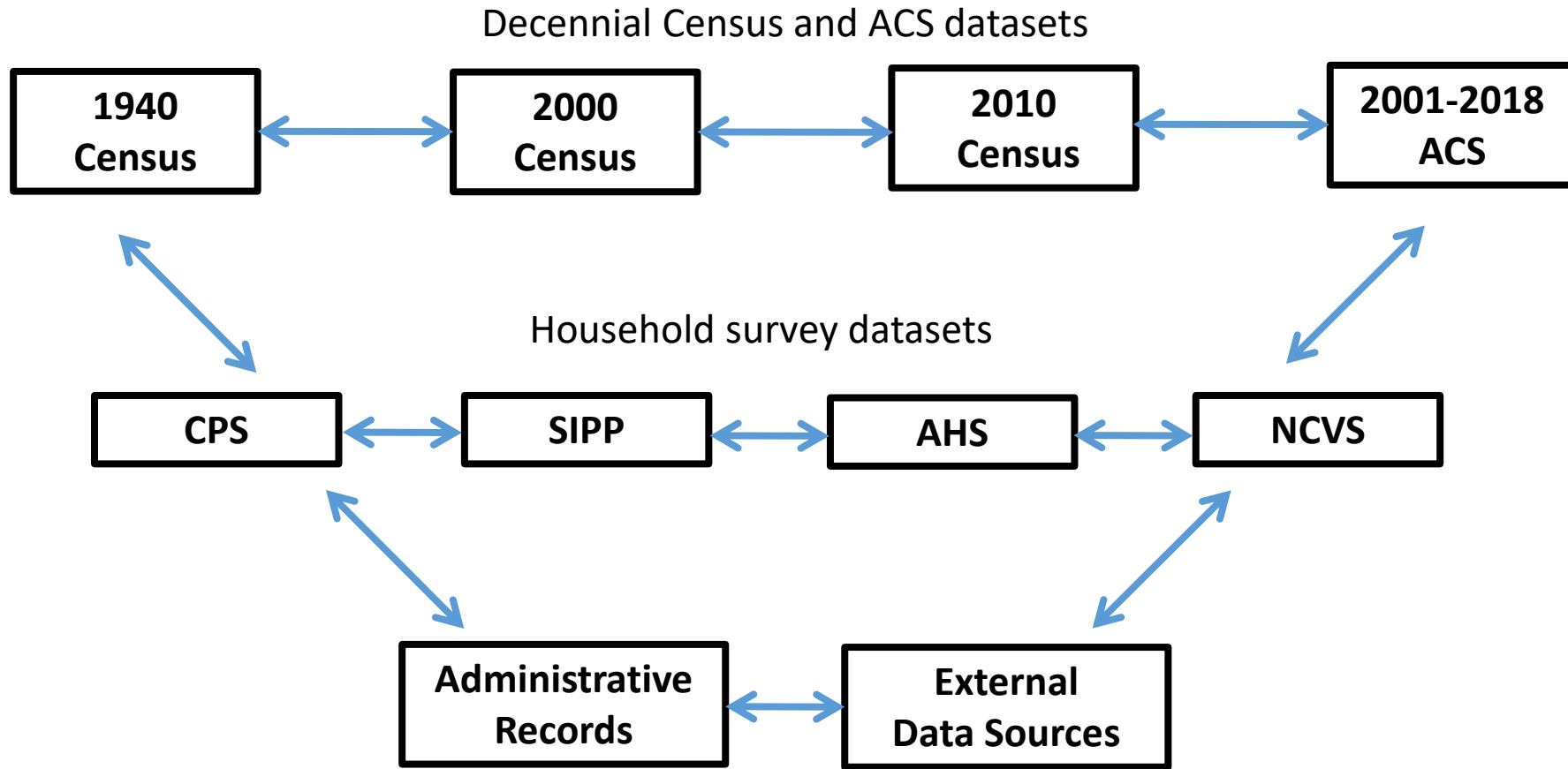
## Restricted-Access Microdata

1. No obvious identifiers or PII (personally identifiable information) replaced by PIK (Protected Identification Key) to allow linkages
2. Includes all microdata records
3. Includes unlimited number of variables including sensitive variables and LINKED data records
4. Geographic detail as precise as original data
5. Unlimited categories within variables on the file, e.g., race and ethnicity, exact dates of events, ICD-10 codes
6. No truncated extreme codes for certain variables (top or bottom coding), e.g., income

# Microdata on Individuals and Households Available from Census Bureau:

- Full-count **Decennial Censuses**
- Administrative records
  - **Master Address File (MAF)** includes Addresses of every HH in US
  - Births and Deaths
- Sample surveys
  - **American Community Survey (ACS)**
  - **Current Population Survey (CPS)**
  - **American Housing Survey (AHS)**
  - **Survey of Income Program Participation (SIPP)**
  - **National Crime Victimization Survey (NCVS)**
- Most data are linkable at the person level

# Census Longitudinal Infrastructure



# American Community Survey (ACS)

- Annual, since 2005. The most comprehensive survey of U.S. households and individual residents
- Helps local officials, community leaders, and businesses understand the changes taking place in their communities. It is the premier source for detailed population and housing data
- Coverage provides reliable estimates for the nation, regions, states, counties, cities, tracts, block groups
- Information from the survey generates data that help determine how more than \$675 billion in federal and state funds are distributed each year.

ACS Initial Addresses and Sample Selected and Final Interviews				
	Housing Units		Group Quarters People	
Year	Initial	Final	Initial	Final
	Addresses	Interviews	Sample	Actual
	Selected		Selected	Interviews
<u>2018</u>	3,544,000	2,143,000	166,000	150,000
<u>2017</u>	3,526,808	2,145,639	168,590	157,721
<u>2016</u>	3,527,047	2,229,872	206,415	160,572

## ACS Questionnaires:

<https://www.census.gov/programs-surveys/acs/methodology/questionnaire-archive.2019.html>

Household composition

Each person's relationship to Person 1 (INCLUDING SAME-SEX PARTNER)

Age and date of birth

Hispanic origin

Race

Work

Occupation

Education

Detailed housing characteristics, e.g., receiving SNAP benefits

Other person characteristics (citizenship, language skill)



**ACS Questionnaires:** <https://www.census.gov/programs-surveys/acs/methodology/questionnaire-archive.2019.html>

Health Insurance: (Yes or No, what type, premiums, subsidy for premium)

Disability (deaf, blind, cognitive, walking, dressing, bathing, doing errands)

Veteran (when served, VA disability, VA disability rating)

Fertility (birth in the last year)

# Current Population Survey (CPS)

<https://www.census.gov/programs-surveys/cps/technical-documentation/questionnaires.html>

- Conducted monthly for over 50 years. Data from approximately 54,000 households of 70,000 selected to represent U.S. and each state are interviewed every month.
- Each household interviews for 4 consecutive months and again a year later for the same corresponding 4 consecutive months. **This is the panel component of CPS** and these data are restricted use only.
- CPS provides monthly estimates of employment (farm and nonfarm); self-employed, domestics, unpaid workers in nonfarm enterprises, wage and salary employees and estimates of total employment.
- **Annual Social and Economic Supplements (ASEC)**, are annual data collected with the March interviews since 1962. Provides data on family structure, household composition, education, health insurance coverage, income sources, poverty, geographic mobility.

# Current Population Survey (CPS): Supplement Topics:

<https://www.census.gov/programs-surveys/cps/about/supplemental-surveys.html>

- School Enrollment: Collected in October, every year
- Fertility and Marriage: Collected in June, every even year
- Voting and Registration: Collected in November, every even year
- Food Security: Collected in December, every year
- Tobacco Use: Collected more sporadically, last was July 2018

# Some Research Topics Using Restricted Census Data:

- Carolyn Liebler (UM): Race Response Decisions and Response Changes Due to Context
- 
- Joe Chestnut (DU) Race/class differences in accessibility of public transit for workplace commuting
- Lori Hunter (CU Boulder) Determinants of aging and migration in/out of rural, small-town USA

# NCVS Sample Design

- Longitudinal household data from a rotating sampling frame
- Housing units remain in sample for 3.5 years, comprising 7 total interviews occurring every 6 months
- Initial interview is performed in-person with all household members age 12+, proxy interviews with incapacitated persons

# Unique Design Enables Innovative Research

- Assess differences in victimization reports by method of data collection: in-person with household vs. telephone
- Examine trajectories of victimization experiences
- Investigate causal impacts of victimization on economic characteristics
- Observe associations between crime/household/victim characteristics and differential attrition from the sample
- Evaluate state/local crime control and public health policies related to public safety, victim services, doctor/hospital procedures, etc.



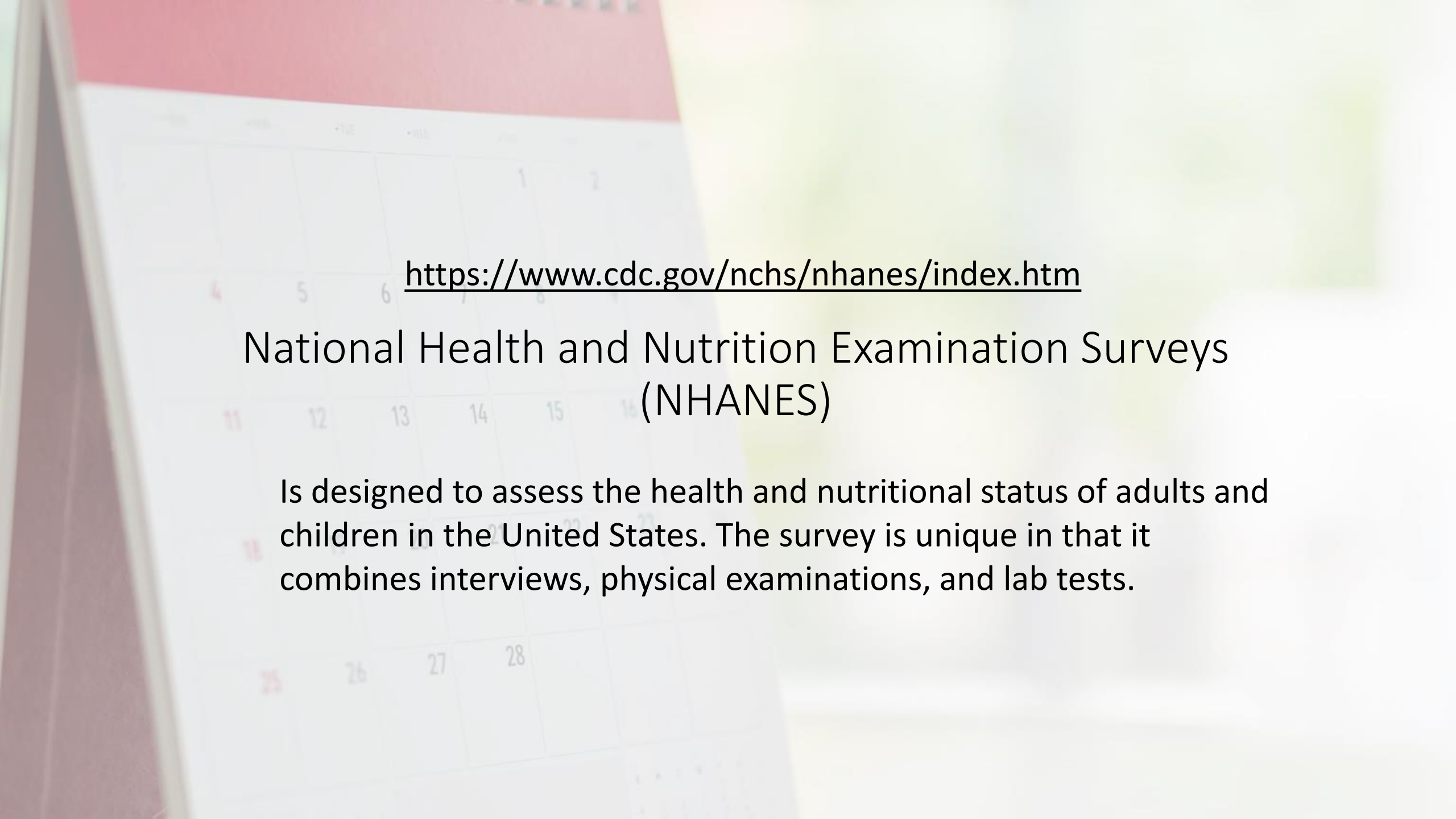
# Local Examples of NCVS Research

- *State Variation in Intimate Partner Violence Rates, Reporting and Healthcare Utilization* – Sarah Small, Anita Pena (CSU)
- *The Social and Economic Consequences of Violent Victimization in the NCVS* – Joshua Clapp, Ben Gilbert, Phil Pendergast (UWyo, CSM, CU-B)

# Microdata on Individuals and Households Available from NCHS/CDC:

## NCHS/CDC:

- National Vital Statistics System (NVSS)
- National Health and Nutrition Examination Surveys (NHANES)
- National Health Interview Survey (NHIS) linked with National Death Index (NDI)
- NHIS linked with Medical Expenditure Panel Survey (MEPS)
- National Survey of Family Growth (NSFG)
- Behavioral Risk Factor Surveillance System (BRFSS)
- National Survey of Drug Use and Health (NSDUH)

The background of the slide features a blurred image of a calendar with a red header and yellow sticky notes. The calendar shows dates from 1 to 28, with some dates highlighted in red. The text is overlaid on this background.

<https://www.cdc.gov/nchs/nhanes/index.htm>

## National Health and Nutrition Examination Surveys (NHANES)

Is designed to assess the health and nutritional status of adults and children in the United States. The survey is unique in that it combines interviews, physical examinations, and lab tests.

The background of the slide features a blurred image of a calendar and a document. The calendar is on the left, showing dates from 1 to 28. The document is on the right, with some text and a green sticky note visible. The overall tone is light and professional.

# NHANES

Findings from NHANES are used to determine the prevalence of major diseases and risk factors for diseases.

NHANES findings are also the basis for national standards for such measurements as height, weight, and blood pressure.

Data from this survey will be used in epidemiological studies and health sciences research, which help develop sound public health policy, direct and design health programs and services, and expand the health knowledge for the Nation.



# National Health and Nutrition Examination Survey: 1999–2020 Survey Content Brochure



The background of the slide is a blurred image. On the left, a portion of a calendar is visible, showing dates from 1 to 28. The calendar has a red header and a white grid. On the right, a person wearing a white lab coat is partially visible, standing in a clinical or laboratory setting. The overall lighting is bright and soft.

## NHANES

The interview includes demographic, socioeconomic, dietary, and health-related questions. The physical examination component consists of medical, dental, and physiological measurements, as well as laboratory tests administered by highly trained medical personnel.



The background of the slide features a blurred image of a calendar on the left and a map of the United States on the right. The calendar shows dates from 1 to 28, with some dates in red. The map is a light green color.

## NHANES

It began in the early 1960s and has been conducted as a series of surveys focusing on different population groups or health topics. In 1999, the survey became a continuous program that has a changing focus on a variety of health and nutrition measurements to meet emerging needs. The survey examines a nationally representative sample of about 5,000 persons each year. These persons are located in approximately 15 counties across the country each year.



# NHANES

Findings from NHANES are used to determine the prevalence of major diseases and risk factors for diseases.

NHANES findings are also the basis for national standards for such measurements as height, weight, and blood pressure.

Data from this survey will be used in epidemiological studies and health sciences research, which help develop sound public health policy, direct and design health programs and services, and expand the health knowledge for the Nation.

The sample for the survey is selected to represent the U.S. population of all ages. To produce reliable statistics, NHANES over-samples persons 60 and older, African Americans, and Hispanics.

The diseases, medical conditions, and health indicators studied include:

- Anemia
- Cardiovascular disease
- Diabetes
- Environmental exposures
- Eye diseases
- Hearing loss
- Infectious diseases
- Kidney disease
- Nutrition
- Depression
- Obesity
- Oral health
- Osteoporosis
- Physical fitness and physical functioning
- Reproductive history and sexual behavior
- Respiratory disease (asthma, chronic bronchitis, emphysema)
- Sexually transmitted diseases
- ADHD in youth



[https://www.cdc.gov/nchs/nhis/about\\_nhis.htm](https://www.cdc.gov/nchs/nhis/about_nhis.htm)

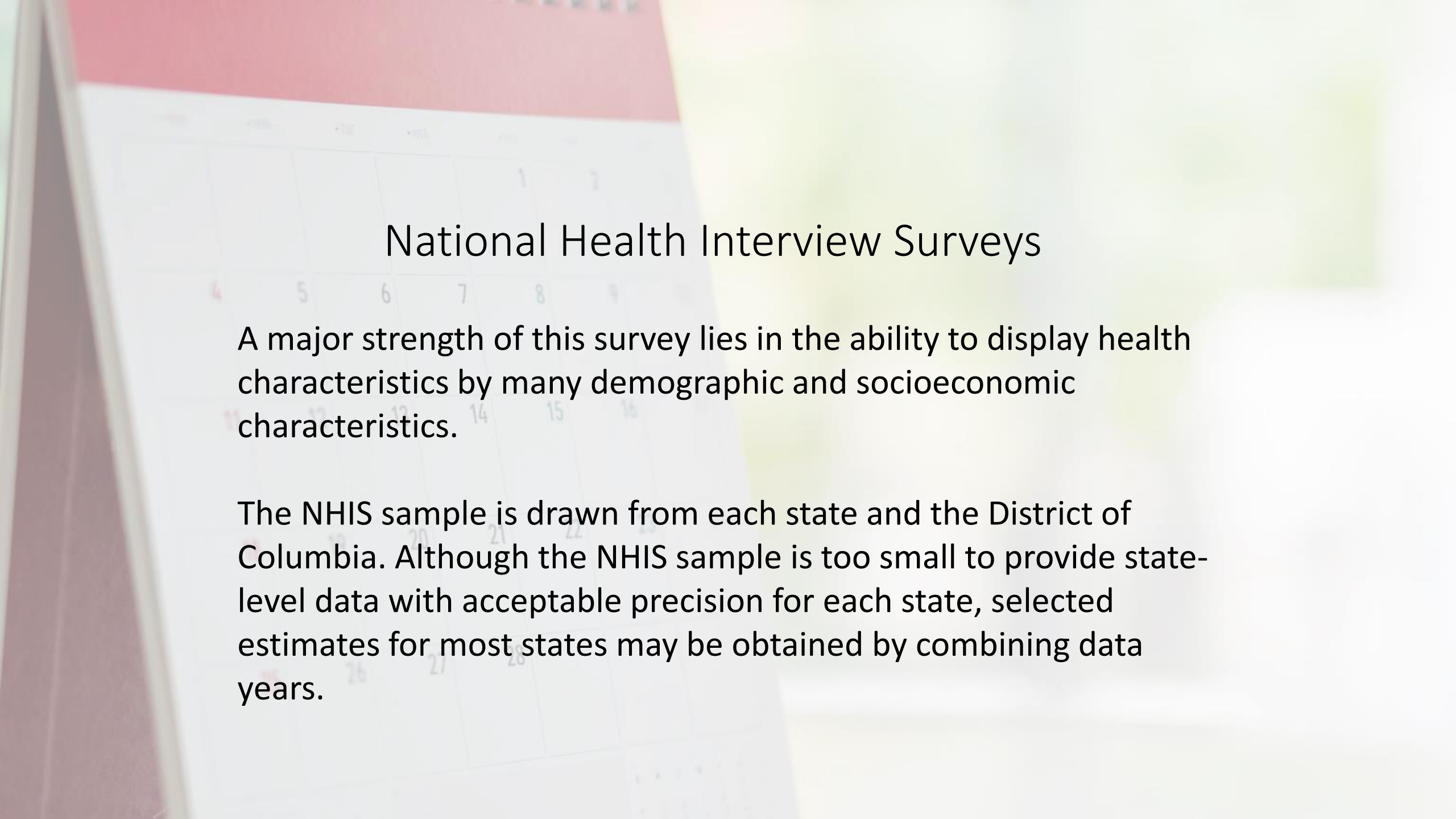
## National Health Interview Surveys (NHIS)

Began as a result of National Health Survey Act of 1956. It provided for a continuing survey and special studies to secure accurate and current statistical information on the amount, distribution, and effects of illness and disability in the United States and the services rendered for or because of such conditions.

The background of the slide features a blurred image of a calendar with a red header and yellow sticky notes. The calendar grid shows dates from 1 to 28, with some dates highlighted in red. The text is overlaid on this background.

## National Health Interview Surveys

NHIS data are used widely to monitor trends in illness and disability. The data are also used by the public health research community for epidemiologic and policy analysis of health problems, determining barriers to accessing and using appropriate health care, and evaluating Federal health programs.

The background of the slide features a blurred image of a calendar with a red header and several yellow sticky notes attached to it. The calendar shows dates from 1 to 27. The text is overlaid on this background.

## National Health Interview Surveys

A major strength of this survey lies in the ability to display health characteristics by many demographic and socioeconomic characteristics.

The NHIS sample is drawn from each state and the District of Columbia. Although the NHIS sample is too small to provide state-level data with acceptable precision for each state, selected estimates for most states may be obtained by combining data years.



## Major Health Topics in the NHIS:

Physical and mental health status

Chronic conditions, including asthma and diabetes

Measures of functioning and disability

Access to and use of health care services

Health insurance coverage and type of coverage

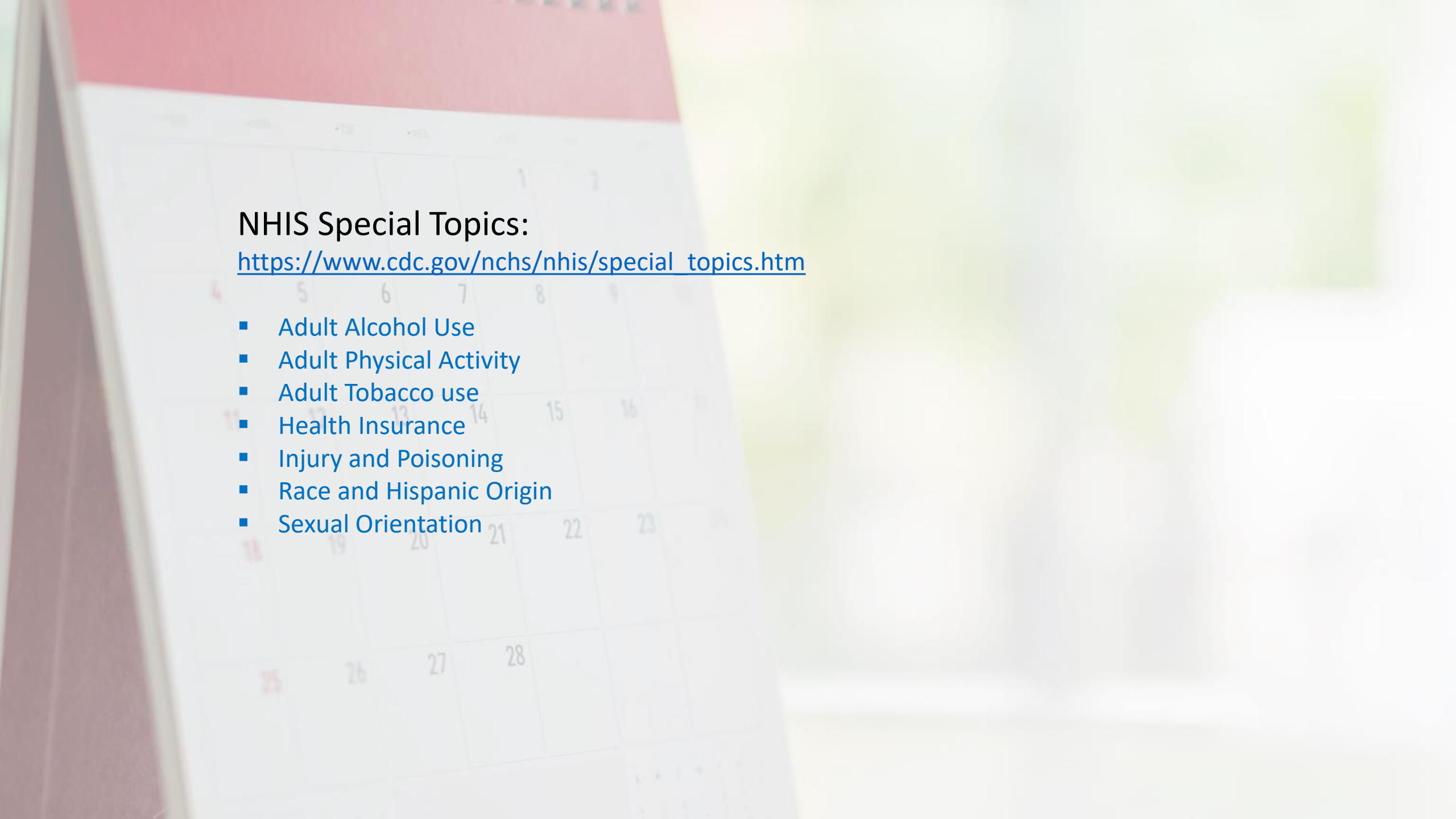
Chronic pain, pain management, and prescription opioid use

Demographics and social determinants of health

# NHIS

Supplements by Year and Topic:

[https://www.cdc.gov/nchs/data/nhis/NHIS Supplements and Sponsors.pdf](https://www.cdc.gov/nchs/data/nhis/NHIS_Supplements_and_Sponsors.pdf)

The background of the slide features a blurred image of a calendar on the left and a map of the United States on the right. The calendar is a standard monthly grid with dates visible. The map shows the continental United States with state boundaries. The text is overlaid on the calendar portion.

## NHIS Special Topics:

[https://www.cdc.gov/nchs/nhis/special\\_topics.htm](https://www.cdc.gov/nchs/nhis/special_topics.htm)

- Adult Alcohol Use
- Adult Physical Activity
- Adult Tobacco use
- Health Insurance
- Injury and Poisoning
- Race and Hispanic Origin
- Sexual Orientation

# NHIS vs NHANES

- **NHIS**

- Conditions AND Behaviors Self-reported
- Utilization and Insurance
- 35,000 Households Annually
- Nationally Representative Annually
- State Estimates reliable if years combined
- Extensive demographic detail with representation

- **NHANES**

- Conditions from Physical Exam and Lab results
- Less Utilization and Insurance
- 5000 Households Annually
- Nationally Representative Annually
- State Estimates unreliable
- Demographic detail but less representation



## NHIS Linked with National Death INDEX (NDI) Mortality DATA

### **Restricted-Use Data:**

<https://www.cdc.gov/nchs/data-linkage/mortality-restricted.htm>

Linked to 1999-2014 NHIS. More complete set of mortality variables. Linked for adult participants and children.

## Medical Expenditure Panel Survey (MEPS):

<https://www.ahrq.gov/data/meps.html>

- MEPS is the only national data source measuring how Americans use and pay for medical care, health insurance, and out-of-pocket spending.
- Annual surveys of individuals and families, as well as their health care providers, provide data on health status, the use of medical services, charges, insurance coverage, and satisfaction with care.
- Longitudinal--captures changes in health status, medical conditions, healthcare utilization, and healthcare expenditures

## The MEPS Household Component (MEPS-HC)

A nationally representative survey of the U.S. civilian noninstitutionalized population. The sampling frame is drawn from respondents to the National Health Interview Survey (NHIS)

Uses an overlapping panel design. A new panel of sample households is selected each year, and data for each panel are collected for two calendar years. The two years of data for each panel are collected in five rounds of interviews that take place over a two and a half year period.

# Health Data Course: Spring 2022

## “Federal Statistical Data for Health Research and Policy”

1. Students develop the knowledge and skills required to effectively use a variety of federal statistical data sets for health research and policy analysis.
2. Students get “hands on” exercises using SAS to analyze public versions of the federal data sets.
3. The semester project is a restricted data proposal.





# Primary Health Data Sets Covered in the Course

## NCHS/CDC:

- National Vital Statistics System (NVSS)
- National Health and Nutrition Examination Surveys (NHANES)
- National Health Interview Survey (NHIS) linked with National Death Index (NDI)
- NHIS linked with Medical Expenditure Panel Survey (MEPS)
- National Survey of Family Growth (NSFG)
- Behavioral Risk Factor Surveillance System (BRFSS)
- National Survey of Drug Use and Health (NSDUH)

## Census:

- American Community Survey (ACS)
- Current Population Survey (CPS)\*

## Bureau of Justice Stats (BJS):

- National Crime Victimization Survey (NCVS)

## Bureau of Labor Stats (BLS):

- Survey of Occupational Injuries and Illnesses (SOII)
- Census of Fatal Occupational Injuries (CFOI)
- American Time Use Survey (ATUS)

# The Rocky Mountain Research Data Center (RMRDC)

<https://www.colorado.edu/rocky-mountain-research-data-center/>

Thank You!!

Jani Little, Executive Director

[jani.little@colorado.edu](mailto:jani.little@colorado.edu)

Phil Pendergast, Administrator (Census)

[philip.m.pendergast@census.gov](mailto:philip.m.pendergast@census.gov)

Kas McLean, GRA (Econ PhD Candidate)

[kas.mclean@colorado.edu](mailto:kas.mclean@colorado.edu)

A stylized graphic of a mountain range with three peaks. The peaks are filled with a dark purple color and outlined in white. The mountains are set against a background that transitions from a dark blue at the top to a dark green at the bottom. The text 'ADVANCING THE FRONTIERS OF SOCIAL SCIENCE' is written in a white, bold, sans-serif font across the middle of the image, and 'ECONOMICS - POPULATION - HEALTH' is written in a similar font at the bottom.

ADVANCING THE FRONTIERS OF SOCIAL SCIENCE

ECONOMICS - POPULATION - HEALTH