National Center for Health Statistics
Data Linkage

The NCHS Data Linkage Program: Leveraging the nation's health data for evidence-based decision making

Lisa B. Mirel

Director, Data Linkage Program

NCHS Webinar

October 20, 2020

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

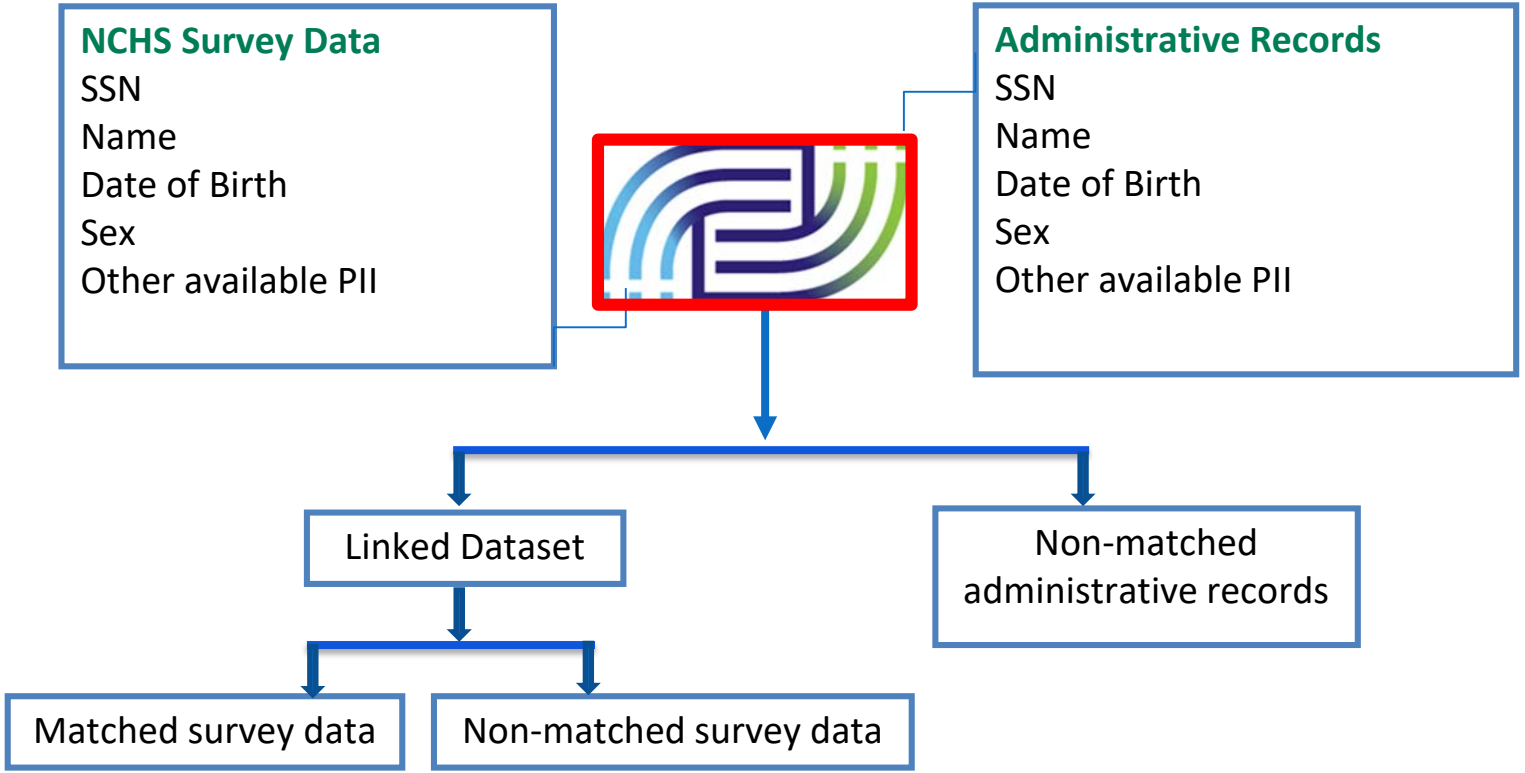
Policy Questions Require Evidence-Based Answers

- Many of the most pressing policy questions require complex, detailed data
  - *Do Social Security Disability Insurance beneficiaries have access to care during the waiting period before Medicare entitlement?*
  - *Are there adverse health effects associated with the mandatory folic acid fortification policy for grain products in older adults?*
  - *How effective are health policies in reducing lead exposure?*
- Collecting data to answer such questions is time-consuming and expensive

Data Linkage as a Solution

- Linking data is a powerful mechanism to provide policy relevant information in an efficient way
  - Health survey data are collected to monitor health status, well-being, and health care access
  - Administrative data are collected for programmatic purposes
- Combining these types of existing data efficiently creates opportunities to answer key health and policy relevant questions

NCHS Data Linkage Program



NCHS Surveys Included in Linkages

National Health Interview Survey (NHIS)

A nationally representative, cross-sectional household interview survey that serves as an important source of information on the health of the civilian noninstitutionalized population of the United States

National Health and Nutrition Examination Survey (NHANES)

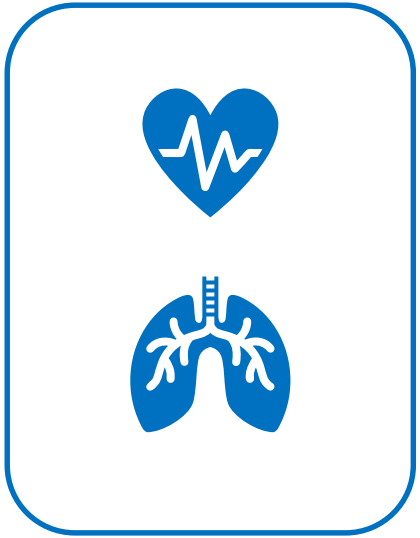
A nationally representative, cross-sectional sample of the U.S. civilian noninstitutionalized population, which includes an interview in the household followed by an examination in a mobile examination center

National Health Care Surveys

A family of data collection efforts that gather information about providers of health care services and the patients they serve across the spectrum of health care settings from ambulatory and hospital care to long term care settings

NCHS Surveys Collect Information on

Health behaviors



Health conditions

Socioeconomic status

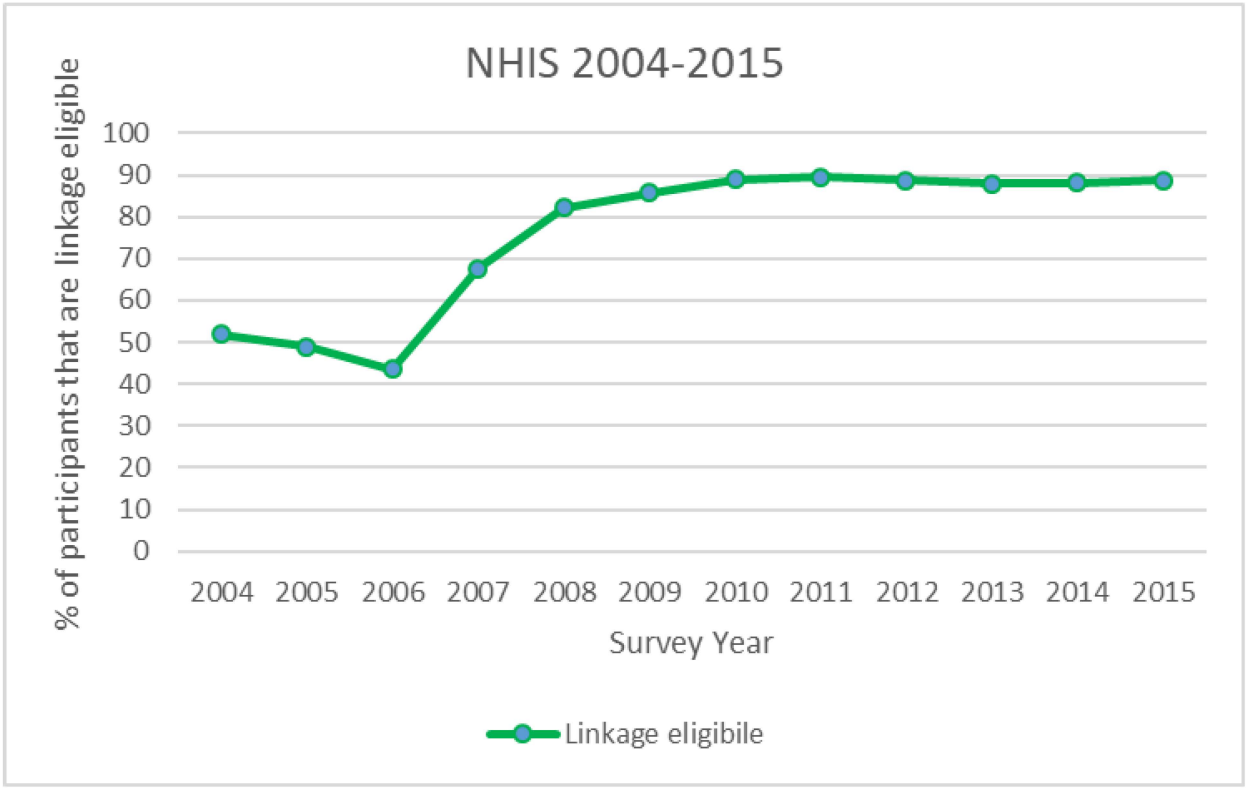
Healthcare access and utilization

Linkage Eligibility

- Only NCHS survey participants who have provided consent as well as the necessary personally identifiable information (PII) are considered ***linkage-eligible***

- NCHS survey participants are informed of NCHS' intent to conduct data linkage activities through a variety of informed consent procedures:
  - Advance letters
  - Participant brochures
  - Signed consent forms
  - Questionnaires

Percent of NHIS Participants that are Linkage Eligible



Note: NHIS is the National Health Interview Survey; this graph only includes sample adults/ sample children

Vital and Administrative Data Sources Used in Linkages

National Death Index (NDI)

- centralized database of U.S. death records gathered from states' vital statistics offices

Centers for Medicare & Medicaid Services (CMS)

- provides health insurance and health care coverage for age 65+ and some vulnerable populations in the U.S., including low-income children and the aged or disabled poor

Housing and Urban Development (HUD)

- responsible for overseeing and managing domestic housing programs and policies, including specialized programs for high-needs populations (e.g., the elderly, homeless, and disabled) in the U.S.

NCHS Survey Data

Population-based household surveys

National healthcare surveys

Housing and Urban Development

Timing of housing assistance

Receipt/type of housing assistance

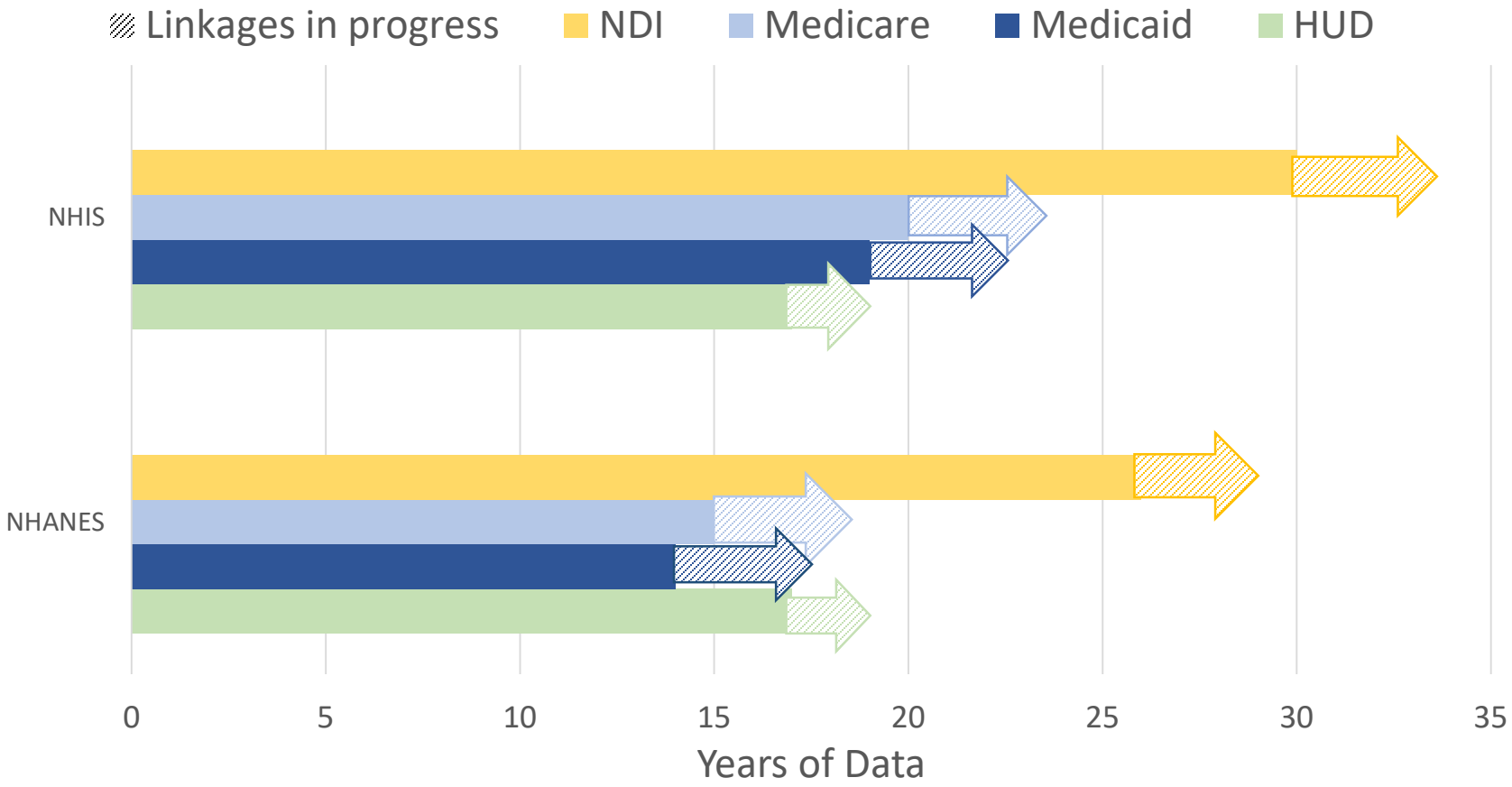
Mortality

Cause of death from National Death Index

Medicare and Medicaid

Healthcare utilization and expenditures

Availability of Linked NHIS and NHANES Data



New Data Releases:
Linked National Hospital Care Survey data

- Linked electronic health records and claims from the National Hospital Care Survey with Medicare data to provide new resources for patient outcomes research, including evaluating treatment effectiveness and post-acute care utilization

- Linked Inpatient and Emergency Department records with death certificate data to assess 30, 60, 90-day post hospitalization mortality

National Health Statistics Reports

National Hospital Care Survey Demonstration Projects: Opioid-involved Emergency Department Visits, Hospitalizations, and Deaths

https://www.cdc.gov/nchs/data-linkage/nhcs_linkage.htm

Examples of Epidemiologic Research with Linked Data

*Over **950** publications based on NCHS linked data*

Linked Mortality Data

- Deaths Associated with Underweight, Overweight, and Obesity
- Air Pollution Exposure and Heart Disease Mortality
- Educational Differentials in US Adult Mortality

Linked NCHS-CMS Data

- Characteristics of those who chose Medicare Advantage upon Medicare enrollment
- Health service use among the previously uninsured
- Concordance between survey reported childhood asthma and linked Medicaid

Linked NCHS-HUD Data

- Housing assistance and blood lead levels
- Cigarette smoking and adverse health outcomes among adults receiving federal housing assistance
- Housing assistance associated with insurance rates and unmet medical need

Creating Resources for Evidence-based Policy Research

Link NCHS survey data with the National Death Index and other health related administrative data

Provide documentation on linkage methodology, linkage quality and analytic guidelines

Release the curated data files for multiple research projects and to replicate the findings of other researchers

Example 1: Linked Data Support Policy Evaluation

Policy:

There is a mandatory 24-month waiting period for Social Security Disability Insurance (SSDI) beneficiaries prior to their eligibility for Medicare benefits.

Question:

Do SSDI beneficiaries have health insurance coverage and adequate access to care during the waiting period before Medicare entitlement?

Example 1: Linked Data Support Policy Evaluation Results

Health insurance and access to care among Social Security Disability Insurance beneficiaries during the Medicare waiting period *(Riley, Inquiry 2006)*

Data Sources:

1994-1996 National Health Interview Survey (NHIS) participants aged 18-64 that linked with SSA Medicare administrative data

Key Findings:

About 1 in 4 of SSDI beneficiaries reported having no health insurance, with the uninsured reporting more problems with access to care than insured individuals.

Example 2: Linked Data Support Policy Evaluation

Policy:

In 1998, the U.S. Food and Drug Administration required that folic acid be added to enriched grain products (such as bread, pasta, rice, and cereal) in effort to prevent certain birth defects.

Question:

Are there adverse health effects associated with the mandatory folic acid fortification policy for grain products in older adults?

Example 2: Linked Data Support Policy Evaluation Results

Intake and Biomarkers of Folate and Risk of Cancer Morbidity in Older Adults, NHANES 1999-2002 with Medicare Linkage *(Hu et al. PLOS One 2016)*

Data Sources:

1999-2002 National Health and Nutrition Examination Survey (NHANES) participants aged 57 years and older at baseline that linked with 1999-2007 Medicare administrative data and NDI data through 2006

Key Findings:

In post-folic acid fortification years, high folate intake and biomarkers in older adults appear to be protective against cancer compared to lower intake. Study does not show a negative impact of folic acid fortification on cancer risk.

Example 3: Linked Data Support Policy Evaluation

Policy:

US Department of Housing and Urban Development (HUD) has initiated lead-based paint regulations in federal housing to reduce hazards associated with lead-based paints in homes.

Question:

How effective are HUD's policies at reducing lead exposure for children living in federal housing, as measured by blood lead levels?

Example 3: Linked Data Support Policy Evaluation Results

Housing Assistance and Blood Lead Level: Children in the United States, 2005-2012

(Ahrens et al. AJPH, 2016)

Data Sources:

2005-2012 NHANES data linked with 1999-2014 HUD administrative data for children aged 1-5 years

Key Findings:

Children receiving federal housing assistance had a statistically significant lower geometric mean blood lead levels than comparable children who did not receive housing assistance

Public-use NCHS Linked Data

- Two types of public-use files released by the NCHS Data Linkage Program:

2015 Public-Use Linked Mortality Files

Public-use Linked Mortality Files (LMF) are available for 1986-2014 NHIS, 1999-2014 NHANES, and NHANES III. The files include a limited set of mortality variables for adult participants only. The public-use versions of the NCHS Linked Mortality Files were subjected to data perturbation techniques to reduce the risk of participant re-identification. For select records, synthetic data were substituted for follow-up time or underlying cause of death. Information regarding vital status was not perturbed.

Data Files and Data Dictionaries

- File Description: Public-Use 2015 LMF [PDF – 175 KB]
- Public-Use Data Dictionary: Public-Use 2015 LMF [PDF – 101 KB]
- Public-Use Linked Mortality Files: FTP Site
- Software Programs:
  - Sample SAS program [SAS – 12 KB]
  - Sample Stata program [DO – 6 KB]
  - Sample R program

Public-Use NCHS-CMS Medicare Feasibility Files

To maximize the use of the restricted-use linked NCHS-CMS Medicare files, NCHS has created publicly available NCHS-CMS Medicare Feasibility Study data files that can be downloaded directly from this website.

The feasibility files provide a limited set of variables that can be used to determine the maximum available sample size for each linked file. These files are especially useful to researchers considering whether to initiate a RDC proposal to analyze the restricted-use linked NCHS-CMS Medicare files.

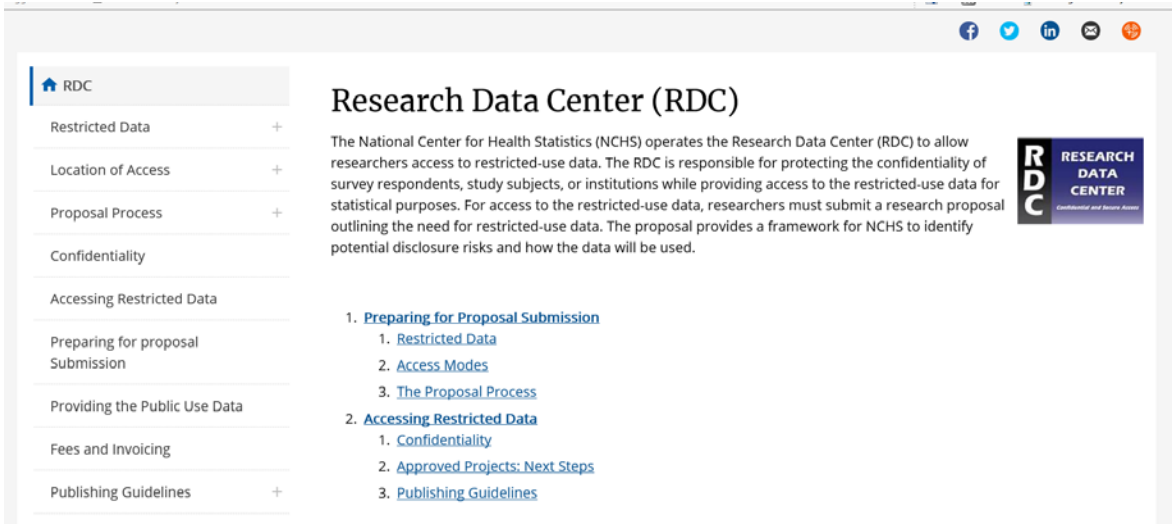
Each feasibility file is NCHS survey and survey-year specific. The following information is included on each feasibility study file:

- NCHS survey public-use identifier
- Survey participant eligibility status for CMS Medicare linkage
- Final match status
- Other variables indicating which specific CMS Medicare data files are available for each successfully linked survey participant

It is important to note that the feasibility study files do not contain any specific information about CMS Medicare benefits.

Restricted-use NCHS Linked Data

- Most linked data products are available through the NCHS and Federal Statistical Research Data Centers



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

Upcoming NCHS Data Linkages

Veterans Affairs

What are the health characteristics, health outcomes, and health care utilization for Veteran's within and outside the VA health system?

Medicaid T-MSIS

How do changes in health care policy affect health status for Medicaid recipients?

End Stage Renal Disease

What is the association of dietary intake and patients diagnosed with ESRD?

Social Security Administration

What are the population health factors that influence disability and unmet needs for SSA beneficiaries?

NCHS Data Linkage Program Future Directions

- **Machine learning**: Incorporating machine learning techniques to improve linkage accuracy and efficiency

- **Encrypted linkage**: Assessing privacy preserving record linkage algorithms to increase linkage activities across health care spectrum

- **Synthetic data:** Piloting innovative methods for public-use data files to improve accessibility and reduce disclosure risk

Commitment to Data Access

- Public-use files
  - Assess disclosure risk
  - Explore development of synthetic public-use micro data files that are analytically useful and valid
- Resources for users
  - E-learning tools (e.g., web tutorial)
  - Compiling examples of use, conducting analyses
- Outreach and increasing user base
  - Federal partners
  - Policy groups
  - Academia

More Information

NCHS Data Linkage Program:

datalinkage@cdc.gov

www.cdc.gov/nchs/data-linkage

Lisa Mirel: LMirel@cdc.gov

Data Access:

NCHS Research Data Center

rdca@cdc.gov

www.cdc.gov/rdc

National Center for Health Statistics
Data Linkage

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

http://www.cdc.gov/nchs

http://www.cdc.gov/nchs/data-linkage/index.htm

National Center for Health Statistics
Data Linkage