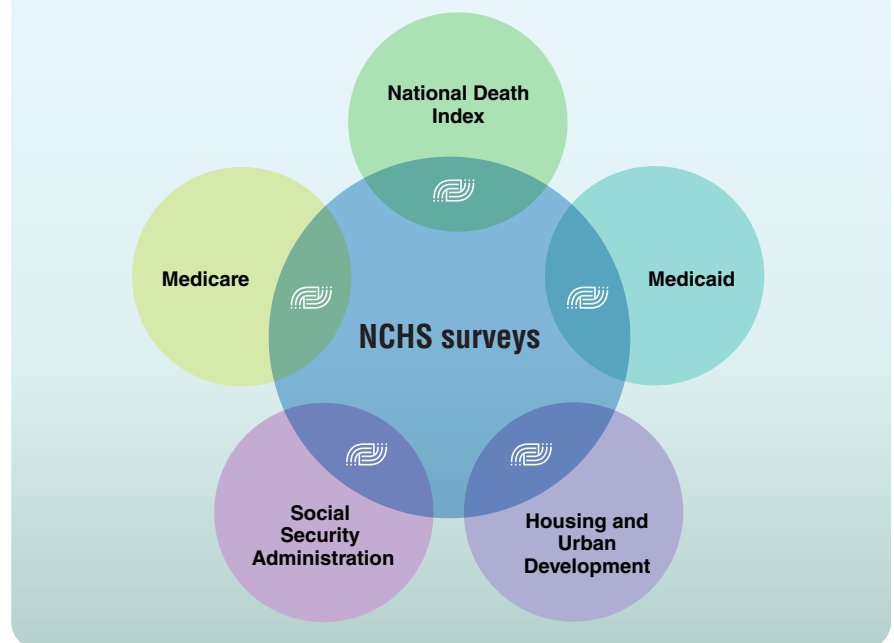# NCHS Data Linkage Program

About NCHS

The National Center for Health Statistics (NCHS) is the nation's principal health statistics agency, providing data to identify and address health issues and to evaluate health policy.

Collaborating with other public and private health partners, NCHS uses a variety of data collection mechanisms to obtain accurate information from multiple sources. This process provides a broad perspective to help understand the population's health, influences on health, and health outcomes.

NCHS data linkage

The NCHS Data Linkage Program is a cross-cutting program designed to maximize the scientific value of the Center's population-based surveys by linking data from health surveys with data from vital and other administrative records. Linked data files enable researchers to examine factors that influence disability, chronic disease, health care utilization, and mortality. Linked data files efficiently broaden the utility of survey and administrative data sources to support development of data-driven evidence for health program and policy evaluation.

Current National Center for Health Statistics Data Linkages

National Death Index

Medicare

NCHS surveys

Medicaid

Social Security Administration

Housing and Urban Development

NCHS survey data used in linkages include:

- National Health Interview Survey (NHIS)
  - NHIS monitors a broad range of health topics with data collected through personal household interviews.
- National Health and Nutrition Examination Survey (NHANES)
  - 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 with physical examinations.
- National Health Care Surveys
  - The National Health Care Surveys are designed to answer key questions on the use and quality of health care services, including safety and disparities in care provided to population subgroups. Patient data collected from Electronic Health Records and health encounters are linked to other sources of health-related data.

Current administrative data sources used in linkages include:

- National Death Index
  - A centralized database of death certificate information provided by State vital statistics offices for deaths in the United States.
- Centers for Medicare and Medicaid Services (CMS)
  - CMS administers the Medicare and Medicaid health insurance programs, which provide health insurance coverage for people aged 65 and over, younger people with disabilities, and other vulnerable populations, including low-income children and pregnant women.
- U.S. Department of Housing and Urban Development (HUD)
  - HUD is responsible for overseeing and managing domestic housing programs and policies, including specialized programs for high-needs U.S. populations (e.g., elderly, homeless, and disabled).
- Social Security Administration (SSA)
  - SSA administers retirement and disability benefits for elderly and disabled populations in the United States.

Administrative data sources to be used in future linkages:

- Veterans Affairs (VA)
  - The Veterans Health Administration at the VA is America's largest integrated health care system, providing care at 1,255 health care facilities, including 170 medical centers and 1,074 outpatient sites of varying complexity. It serves 9 million enrolled Veterans each year.

Goals of the Data Linkage Program

The NCHS Data Linkage Program aims to:

- Create data to support high-quality statistical research and federal evidence-building activities
- Expand the use of other data sources to provide opportunities for outcomes-based analyses
- Apply innovative approaches to expand linked data acquisition and increase data accessibility

Examples of research topics using linked NCHS data

- Public Health
  - Deaths associated with body mass categorized as underweight, overweight, and obese
  - Air pollution exposure and heart disease mortality

- Policy Program Impact
  - Assessment of the association of folate fortification and cancer morbidity
  - Utilization of smoking cessation medication benefits among Medicaid fee-for-service enrollees
  - Blood lead levels among recipients of federal housing assistance
- Health Care Financing and Coverage
  - Health insurance and access to care among Social Security Disability Insurance beneficiaries during the Medicare waiting period
  - Medical costs of chronic kidney disease in the Medicare population

Data access

To protect the confidentiality of survey participants, the linked data files are available for analysis through the NCHS Research Data Centers or Federal Statistical Research Data Centers. Public-use feasibility files containing a limited set of variables have been published to help researchers determine the maximum available sample size for intended analysis.

NCHS releases a confidentiality-protected public-use version of its linked mortality files for NHIS and NHANES (https://www.cdc.gov/nchs/data-linkage/mortality-public. htm).

Analyzing linked data

Linkage methodology and analytic guidance reports are publicly available for all NCHS linked data resources. These reports provide data analysts with important information on data sources, linkage methodology, and key analytic considerations including variable descriptions and summary linkage statistics. NCHS has also created an online tutorial for the NHANES-CMS-linked data to help researchers gain experience using linked survey and health care utilization and cost data (https://www.cdc.gov/nchs/tutorials/NHANES-CMS/index.htm).

Future challenges and opportunities

The NCHS Data Linkage Program will continue to develop these resources by exploring opportunities to:

- Expand data accessibility while maintaining participant confidentiality
- Use linked data resources to inform survey redesign efforts
- Create data that support policy and program evaluation