



National Center for Health Statistics

Data Linkage

National Center for Health Statistics' Data Linkage Program

Lisa B. Mirel and Irma Arispe

Office of Analysis and Epidemiology

National Center for Health Statistics

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

Presentation to the NCHS Board of Scientific Counselors

May 9, 2019

Outline

- NCHS Data Linkage Program
- History of NCHS Data Linkages
- Current Linkages
- Recent Work: Dissemination of Linked Products
- Current Areas of Focus
- Challenges and Opportunities Moving Forward

NCHS Data Linkage Program

- **Cross-cutting program housed in OAE**
 - Supports the Center and its mission as a statistical agency
 - Standardizes linkage algorithms to link the Center's population based health survey data with vital and administrative records
 - Creates datasets that are curated and can be used for repeat analyses to confirm findings
 - Ensures consistent quality of linked datasets
 - Documentation
 - Analytic guidelines
 - Applications of methodology

NCHS Data Linkage Program (cont.)

- **Augments available information for major diseases, risk factors, and health service utilization**
 - Links exposures to outcomes
 - Provides longitudinal component to survey data
- **Expands detail of data difficult to collect in surveys**
 - Benefit periods
 - Health care utilization
- **Reduces cost burden: follow-up surveys can be expensive**

Types of Linkages: NCHS Data Linkage Program

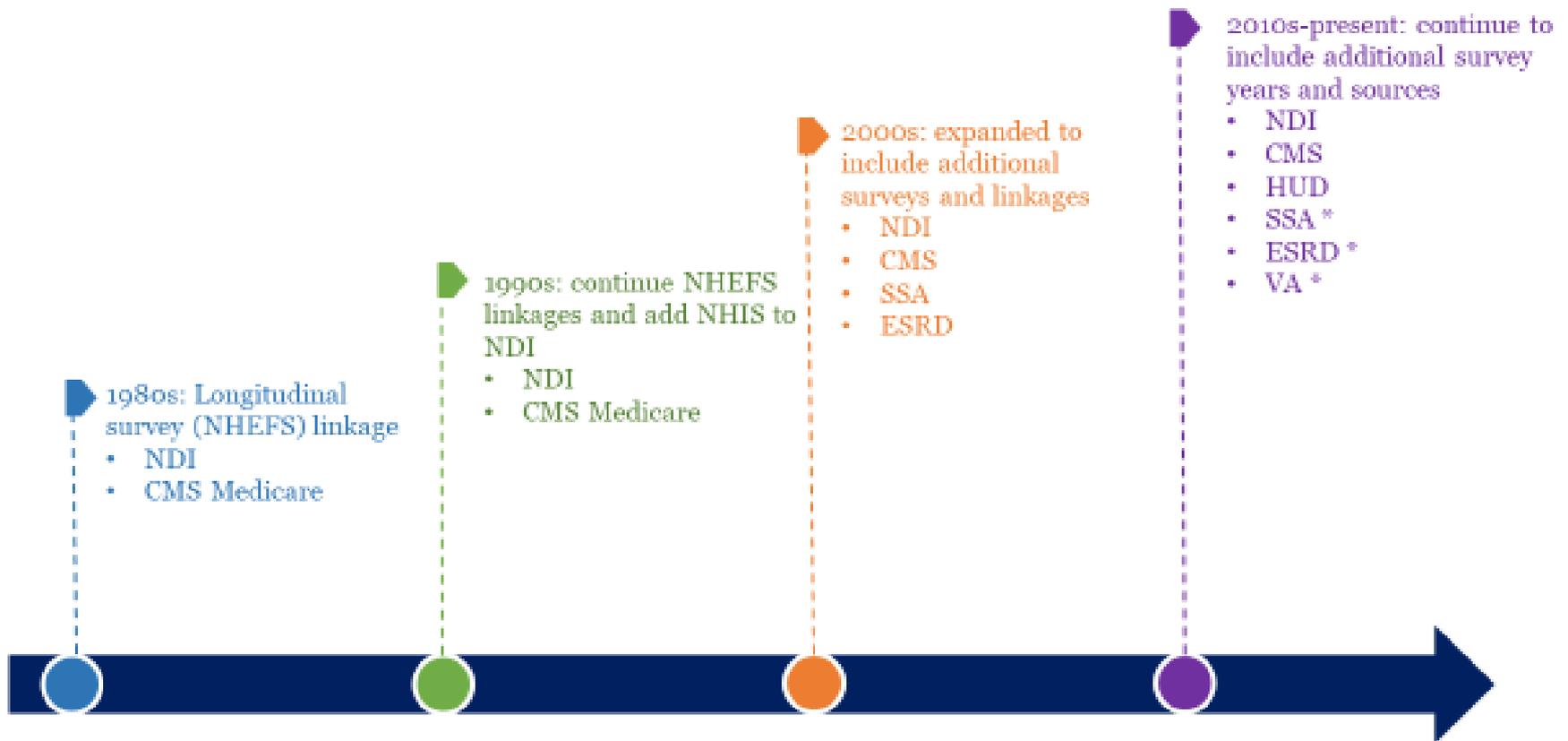
- **Contextual data**
 - Addresses are geocoded to standard Census geo-areas

- **Person or patient-level**
 - Person/patient level survey data linked with administrative data (e.g., Medicare)

Examples of Research with NCHS Linked Data

- **Aging**
 - Risk factors for poor health outcomes (e.g., hip fractures, stroke)
- **Health services**
 - Health care utilization and costs for previously uninsured
- **Disability**
 - Effects of chronic illness and obesity on disability and mortality
- **Disparities**
 - Mortality patterns by race/ethnicity
- **Methodology Studies**
 - Validation of self-report vs. administrative records

History of NCHS Data Linkage



* In discussion

Vital and Administrative Data Sources



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)

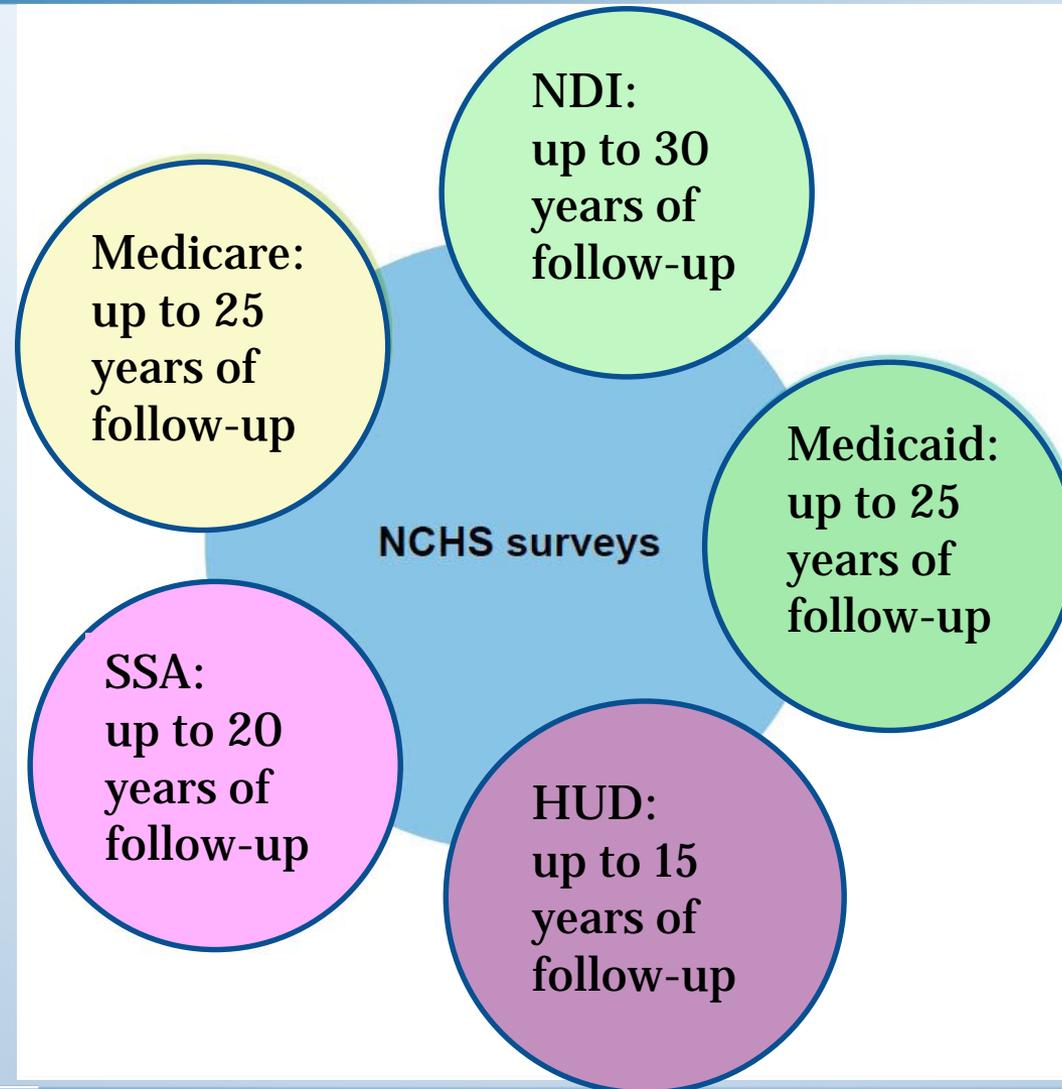
- responsible for overseeing and managing domestic housing programs and policies, including specialized programs for high-needs populations (e.g., the elderly, homeless, and disabled) in the U.S.



Social Security Administration (SSA)

- administers retirement and disability benefits for elderly and disabled populations in the U.S.

Current NCHS Data Linkages



Recent Work: Dissemination of Linked Products

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 MA Upon Medicare Enrollment at Age 65
- 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
- Linkage of 1999-2012 NHIS and NHANES data to HUD administrative records

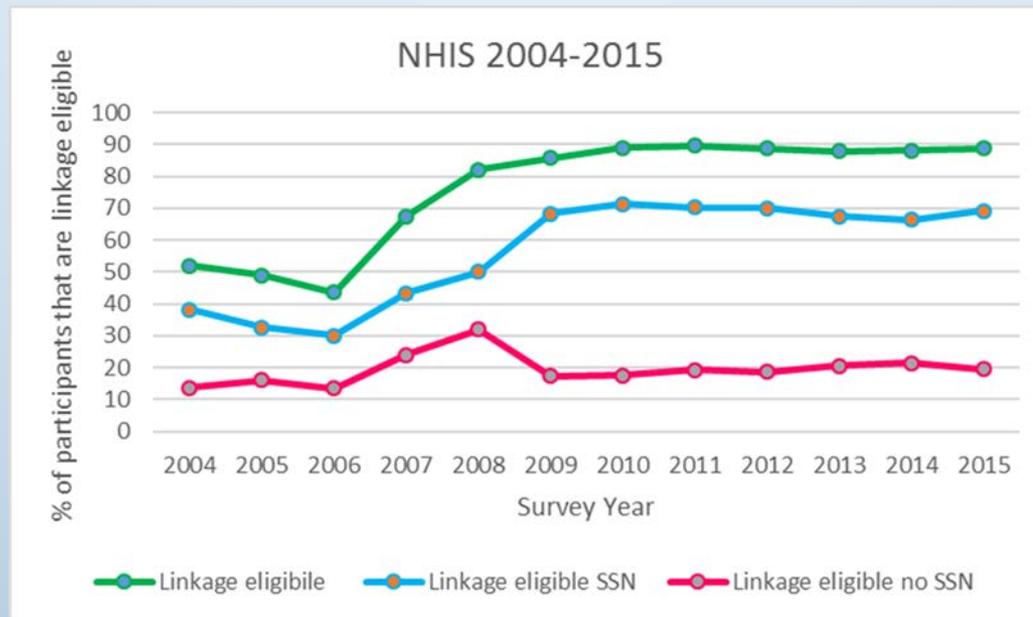


Current Areas of Focus

- Improved linkage algorithms and quality
- Public-use files
- Periodicity of linkages
- New linkage opportunities

Improved Linkage Algorithms I

- NHIS linkage eligibility changed in 2007
 - PII collection changed with NHIS only collecting last 4 digits of Social Security Number (SSN) instead of 9
 - If participants refused SSN, asked a specific question about permission to link without SSN



Improved Linkage Algorithms II

- Potential linkage bias
- Algorithms designed to match with SSN4
 - Medicaid linkage
- Algorithms designed to match no SSN
 - National Hospital Care Survey (NHCS) data linkage

Improved Linkage Algorithms III

- Linkage of NHCS data to NDI and CMS' Medicare files
 - Claims and electronic health records
 - New and improved probabilistic based algorithms
 - Use of machine learning to define blocking
 - Linkage error estimates

Public-use Files

- There are two types of public-use files released by the NCHS Data Linkage Program:

Public-use feasibility data files

- Developed to help interested researchers determine the maximum available sample sizes to assess the feasibility of conducting analyses utilizing the restricted-use linked files in the NCHS RDC.
- Available online for linked CMS and SSA data

Public-use microdata files

- Available online for linked NDI data only

Public-use Files

- **New public-use Linked Mortality Files (LMF) in February, 2019**
 - Most utilized linked file
 - Literature review assessment: at least two thirds of papers rely on the public-use file
- **Public-use LMF**
 - Only adults, limited number of causes of death, some records perturbed
 - Comparison analysis: public-use vs. restricted use file survival analysis estimates

Periodicity

- Develop a schedule for core set of linkages on a recurrent basis
 - Help research community plan for projects and obtain funding
- Produce updated linked files every two to three years:
 - National Death Index
 - Centers for Medicare & Medicaid Services
 - Housing and Urban Development

New Linkages

- *In progress:*
 - PCORTF 2017: link NHCS data to NDI and CMS records
 - PCORTF 2019: link NHCS data to HUD and CMS
- *In discussion:* link NCHS survey data to
 - Veteran's Administration Health Care
 - End Stage Renal Disease data
 - Social Security Administration data

The Future of NCHS' Data Linkage Program

Opportunities	Challenges
Increased interest in linkage	Negotiation of agreements
Interagency collaboration	Ownership and use
Linking new data sources	Releasing updates on a timely predictable schedule
Diversifying linkages to augment survey content, context	Protecting privacy, minimizing disclosure risk
Protecting privacy	Promoting use
Developing new data	Creating tools and conducting analyses to facilitate use



Foundations for Evidence-Based Policymaking Act of 2018 (PL115-435)

Official Title: To amend titles 5 and 44, United States Code, to require Federal evaluation activities, improve Federal data management, and for other purposes.

Title I – Federal Evidence-Building Activities

Title II – Open Government Data Act

Title III – Confidential Information Protection and Statistical Efficiency

Title IV – General Provisions

Some key sections, potential NCHS contributions

- Agency development of evidence building and evaluation plans (§312); Advisory Committee on Data for Evidence Building recommendations (§315)
 - Monitor policy-relevant and evaluation questions, planned data to be used or collected, stakeholder engagements, recommended linkages
 - Conduct outreach about data availability, data quality, methodologies, or analogous policy-relevant analyses
- Strategic plans include assessment of the coverage, quality, methods, effectiveness, and independence of the statistics, evaluation, research, and analysis efforts of the agency
 - Criteria for assessing coverage, quality, effectiveness, independence
- Data inventory (**§3511**)
 - Approaches for storing information about data sets

Increasing communication, managing expectations

PL 115-435 will	Potential implications
Increase number of statistical officials; Information about the need for data; Evidence plans	Increase linkage opportunities; Demand for expertise (statistical, statistical agency, and technical expertise); Use linked data for policy
Increased access to CIPSEA data	Increased need for education about CIPSEA
Expanded provision of data to qualified users	Increased interest in customization, technical consultation



Negotiating Interagency Agreements

- There are agency differences in legislative mandates and requirements to protect data and survey participants' confidentiality
- NCHS is working within federal statistical system to work on standard approaches
- Managing expectations means thinking about...
 - Can a public use file be created?
 - Where will linked data reside?
 - Who will control access?
 - Will analysts from outside NCHS be required to become designated agents?
 - How long should the data file be kept?
 - Who will conduct disclosure review?

Increasing Data Access

- **Public use files: possible but challenging**
 - Assess disclosure risk
 - Develop 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

What's Involved in Using Linked Data for Policy?

- Identify and collect information on policy interests and use of policy levers
 - Regulation, deregulation, taxation, spending, tort liability
- Understand how the data we have can be enhanced to address questions (through new or existing linkages)

Ten Greatest Public Health Achievements, 2000-2010

- Mandating that schools test water for lead
- Requiring landlords to disclose if housing units have lead-based paint

- Inspecting worksites to ensure they follow safety guidelines
- Creating data agreements to share workers' compensation data with public health surveillance systems

- Raising the minimum legal age of tobacco use to prevent youth initiation
- Enforcing smoke-free air laws and extending them to more places

- Implementing Complete Streets programs to encourage exercise
- Addressing liability issues to encourage schools to make playgrounds available after hours

- Joining mutual aid compacts, like the Emergency Medical Assistance Compact, to share resources across jurisdictions
- Issuing emergency declarations

Vaccine-Preventable Diseases

- Requiring children who enter schools to have certain vaccinations
- Funding the Vaccines for Children Program

Cancer Prevention

- Creating the Breast and Cervical Cancer Early Detection Program
- Requiring schools to provide information about HPV vaccines

Prevention and Control of Infectious Diseases

- Requiring hospitals to report central line infection rates
- Creating MOUs among state agencies to respond to foodborne disease outbreaks

Motor Vehicle Safety

- Enacting child restraint laws
- Establishing ignition interlock programs to address DUI offenses in the court system

Maternal and Infant Health

- Adding new conditions to newborn screening programs
- Requiring that certain foods be fortified with folate

Source: <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6019a5.htm>

Using Linked Data for Policy

- How can the linked data be used for policy analyses?
 - Lead exposure and mortality
 - Bipartisan Policy Center (BPC) chapter in *Evidence Works*
- BPC efforts and others to raise awareness of the potential of the linked data to support Evidence Based Policymaking