

# DAE Update

## NCHS Data Linkage Program

### Building evidence for policymaking

Irma Arispe, PhD

NCHS Division of Analysis and Epidemiology

March 6, 2024

# Evidence building for policymaking

- Many pressing policymaking and evaluation questions require complex data that are not available within a single federal agency
- **The Foundations for Evidence-Based Policymaking Act of 2018 (Evidence Act) requires federal agencies to develop evidence to support policymaking**
  - Title III of the Act presumes accessibility for statistical agencies and units to obtain data from federal agencies upon request to support evidence-building.
- **Congress has enacted a National Secure Data Service (NSDS) Demonstration Project at NCSES (within NSF) to facilitate data integration and data access infrastructure for evidence building**

# Data linkage is critical to evidence building



- **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
  
- **Executive branch agencies are looking for data on health and health outcomes**
  - Primary data collection is time-consuming and expensive
  - Combining existing data efficiently creates opportunities to answer health, policy and evaluation questions

# NCHS' Data Linkage Program is a leader in evidence building

- **Leadership and specialized expertise on innovative linkage methods and products**
  - Early evaluator of Privacy Protecting Record Linkage (PPRL) technology
  - Applied machine learning techniques to improve efficiency and accuracy of linkage algorithms
  - Developed first-ever NCHS partially synthetic public use linked mortality files
- **Innovator of systematic approaches for linkage, operational efficiencies that become government-wide examples for innovation**

**Contributing to the development of the  
National Secure Data Service**

# NCHS is partnering to support the NSDS Demonstration Project

- **National Secure Data Service (NSDS) Demonstration Project at NCSES/NSF - enacted by Congress to facilitate data integration and data access infrastructure for evidence building**
- **NCSES-NCHS project develops foundational processes to launch the NSDS**
  - Accessing data in secure ways has been a challenge for evidence building
  - Privacy enhancing technologies can help mitigate privacy risks, facilitate linkage, analysis and dissemination
  - PPR-based linkage model as proof of concept for establishing NSDS linking services
  - Standard NSDS data sharing agreement

# PPRL: Linking Federal Data from Two Statistical Agencies

Data sources	Goal	Current status
<ul style="list-style-type: none"><li>• National Center for Health Statistics: National Health Interview Survey</li><li>• NCSES: Survey of Earned Doctorates (SED)</li></ul>	<ul style="list-style-type: none"><li>• Link data across federal statistical agencies</li><li>• Inform initiatives such as congressionally mandated reports like the NCSES Women, Minorities, and Persons with Disabilities in Science and Engineering</li></ul>	<ul style="list-style-type: none"><li>• Award: HealthVerity/Mathematica</li><li>• PPRL tool: HealthVerity (commercial)</li><li>• Software License Agreement signed</li><li>• Data Sharing Agreement in process<ul style="list-style-type: none"><li>• CIPSEA</li><li>• Privacy Act</li></ul></li></ul>

# Data Linkage Innovations



# Data Linkage Innovations

- **Privacy Preserving Technologies/Privacy Preserving Record Linkage**
  - Testing and evaluation of open source and commercial products
  - Impact: Informs broader implementation of PPRL across HHS OP-DIVs (particularly CMS), Federal Agencies and the entire Health Data Ecosystem
- **Synthetic Data**
  - NCHS's first fully synthetic linked data file with verification service to verify synthetic results against restricted data
  - Impact:
    - Applies innovation to meeting Evidence Act tiered access requirements
    - Expands the accessibility and utility of NCHS's linked data portfolio
    - Basis of standards for synthetic data assessment that can be used across the federal data ecosystem

# Data Linkage Innovations (Continued)

## ■ Machine Learning and AI

- ML techniques incorporated into state-of-the-art linkage algorithms for very large data sets.
- AI techniques used to improve data linking quality and efficiency
- Impact:
  - Informs other entities attempting to link large volumes of data accurately and efficiently
  - Improves the ability to link large data within the health data ecosystem.

## ■ Identifying new sources of data

- One of the first federal linkages of T-MSIS data
- Impact:
  - New data resource that responds to Evidence Act interest in evaluating effectiveness of HHs programs
  - Helps policymakers and evaluators understand the unmet health needs and disparities in health outcomes of Medicaid population
  - NCHS published the first of its kind assessment of TMSIS PII quality and analytic guidance regarding analysis of linked TMSIS data

**Data Linkage Program: In progress and on  
the horizon**

# NCHS Internal Strategic Planning Meeting on Data Linkage



- How can NCHS best use data linkage to support the Evidence Act?
  - What linkages might NCHS request under the Evidence Act to inform our own evidence building needs?
- How should the NCHS Data Linkage Program evolve to:
  - Support NCHS's Evidence Act responses and;
  - Strategically prioritize current and new data linkages?

# Examples of discussion topics

- Identify **new partnership opportunities** with other federal agencies in the rapidly changing areas of privacy enhancing technologies (PETs) and incorporation of artificial intelligence (AI) in data integration.
- **Expand contextual linkage** activities (use of geocodes) at NCHS.
- Seek **new opportunities** with non-typical partners **to expand use of NCHS linked data**
- **New data sources**, systematic process for external data exploration and assessment
- Develop **new operational efficiencies to initiate and conduct linkages** in an environment where calls for linkage are increasing



# Upcoming Data Linkage Releases

## Linkages in progress or planned for 2024

- Most projects supported or requested by ASPE
- All involve internal or external partnerships
- All are structured toward efficiency (multiple NCHS data systems linked where possible)

## Other data releases planned for 2024

- Synthetic linked NHIS-CMS-HUD file (2018) with (internal) verification process – *expected release late Spring 2024*
- Synthetic linked NHCS-NDI file (2016) with (internal) verification process – *expected release Summer 2024*

## Linkages in progress or planned for 2024

Survey and years	Administrative data source and years	Expected release date
2021 NAMCS FQHC	HUD 2020 - 2022	August 2024
2020 NHCS	HUD 2019 - 2021	Fall 2024
2019 NHCS	HUD 2018 - 2020	Fall 2024
1999 – 2021 NHIS	HUD 1999 -2022	Fall 2024
1999 – 2020 NHANES	HUD 1999 -2022	Fall 2024
2021 NAMCS FQHC	NDI through 2022	August 2024
2020 NHCS	NDI through 2022	Fall 2024
2019 NHCS	NDI through 2022	Fall 2024
1986 – 2021 NHIS	NDI through 2022	Fall 2024
1999 – 2020 NHANES	NDI through 2022	Fall 2024
NHANES III	NDI through 2022	Fall 2024
NHEFS	NDI through 2022	Fall 2024
2017 - 2020 NHANES	Medicare 2019 - 2021	Fall 2024
2019 - 2020 NHIS	Medicare 2019 - 2021	Fall 2024
2019 NHCS	Medicare 2019 - 2020	Fall 2024
2020 NHCS	Medicare 2020 - 2021	Fall 2024

# Thank you!

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

