

# Linkage of the National Ambulatory Medical Care Survey Health Center Component to U.S. Housing and Urban Development Administrative Data

Jessie L. Parker, Health Statistician  
Data Linkage Methodology and Analysis Branch  
National Center for Health Statistics (NCHS)

FCSM Research and Policy Conference  
October 22-24, 2024



# NCHS Data Linkage Program: Overview

- Create linked data files that support high quality research and program evaluation
- Utilize state of the art linkage methodologies and provide documentation and support for analyzing linked data files
- Explore innovative methods for improving researcher access to linked data
- New linked files are always in development



# Outline

- **Introduce** the National Ambulatory Medical Care Survey (NAMCS) Health Center (HC) Component and the U.S. Housing and Urban Development (HUD) administrative data
- **Assess** the quality of the NAMCS HC personally identifiable information (PII) data
- **Describe** the methodology used to link NAMCS HC survey data with Housing and Urban Development (HUD) administrative data
- **Detail** plans to link NAMCS HC to additional data sources and describe how these linked resources can be accessed



# NAMCS Health Center Component & U.S. Housing and Urban Development Administrative Data

# NAMCS Health Center Component: 2021 Sample Design & Methods



- **Sample:** Federally qualified health centers (FQHCs) and FQHC look-alikes as classified by the Health Resources and Services Administration (HRSA)
- **Mode:** Electronic health record (EHR) data
- **Data collection:** January-December annually
- **Sample size:** 29 health centers submitted data on 3,543,927 patient visits; weighted response rate is 26.8%
- **Data:** Patient PII as well as visit information such as date of visit; diagnoses and services provided or ordered during the visit; reason for visit; and clinical notes

# U.S. Housing and Urban Development Administrative Data

- HUD administrative data systems contain PII and program participation data for recipients of housing assistance programs in the U.S.
- NAMCS HC-HUD linkage includes administrative data from 2020-2022
- Linkage focuses on HUD's three largest housing assistance programs:
  - **Housing Choice Vouchers**
    - Participants choose any housing unit that meets requirements; includes Homeownership Vouchers, Project-Based Vouchers, Section 8
  - **Public Housing**
    - Tenants are assigned housing
  - **Multifamily Programs**
    - Subsidies are paid to private property owners who provide a percentage of their housing units at affordable rates for low-income persons

# Linkage Motivation

This linkage was funded by OS-PCORTF with a goal of enhancing the surveillance of maternal health and clinical outcomes with FQHC EHR visit data.

NAMCS HC – HUD linked data creates a new data resource that can support a wide array of public health surveillance and policy evaluation studies focused on the role of housing assistance program participation as a key social determinant of health.

# Assessing the Quality of NAMCS Health Center Component PII Data



# Assessment of 2021 NAMCS HC PII Variables Used for Linkage

- NAMCS PII were subject to data cleaning and standardization protocols, for example:
  - Invalid or out of range dates set to null
  - Removal of titles, suffixes, or descriptive words from name field
    - Mister/Miss, Junior, Daughter, Baby Girl, etc.
- Standardized PII fields were evaluated to assess linkage feasibility
- Assessments examined:
  - Variable completeness (i.e., amount of non-missingness)
  - Distribution of values
  - Patient-level linkage eligibility (i.e., sufficient PII to link)
  - Discrepancies in PII across records for patients with multiple visits

## Completeness of Linkage Variables

Linkage variable name	% complete	
	Total Sample	Linkage Eligible
Social security number (SSN)	49.1	49.2
First name	99.9	100.0
Last name	99.9	100.0
Day of birth	99.9	100.0
Month of birth	99.9	100.0
Year of birth	99.9	100.0
Sex	99.6	99.7
ZIP code	99.7	99.8
State of residence	99.7	99.9



# **NAMCS Health Center Component – HUD Data Linkage Methodology**

# Linkage Eligibility

Linkage was attempted only for NAMCS HC Component patient records that had at least two of the following three identifiers present:

- Valid SSN
- Valid date of birth (month, day and year)
- Valid name (first, middle initial, last)

# How We Link

- Linkage conducted in two passes:
  1. Deterministic match using Social Security Number (SSN) collected in the survey
    - Patient PII fields such as name, state of residence, and date of birth are compared to validate
    - This dataset becomes the “truth source”
  2. Probabilistic matching techniques used to identify likely pairs using other PII fields (not SSN)
    - Pair scores are calculated based on the agreement status of the identifiers such as name, state of residence, and date of birth
    - SSN is not used to score pairs, it is used to measure linkage accuracy (when available)

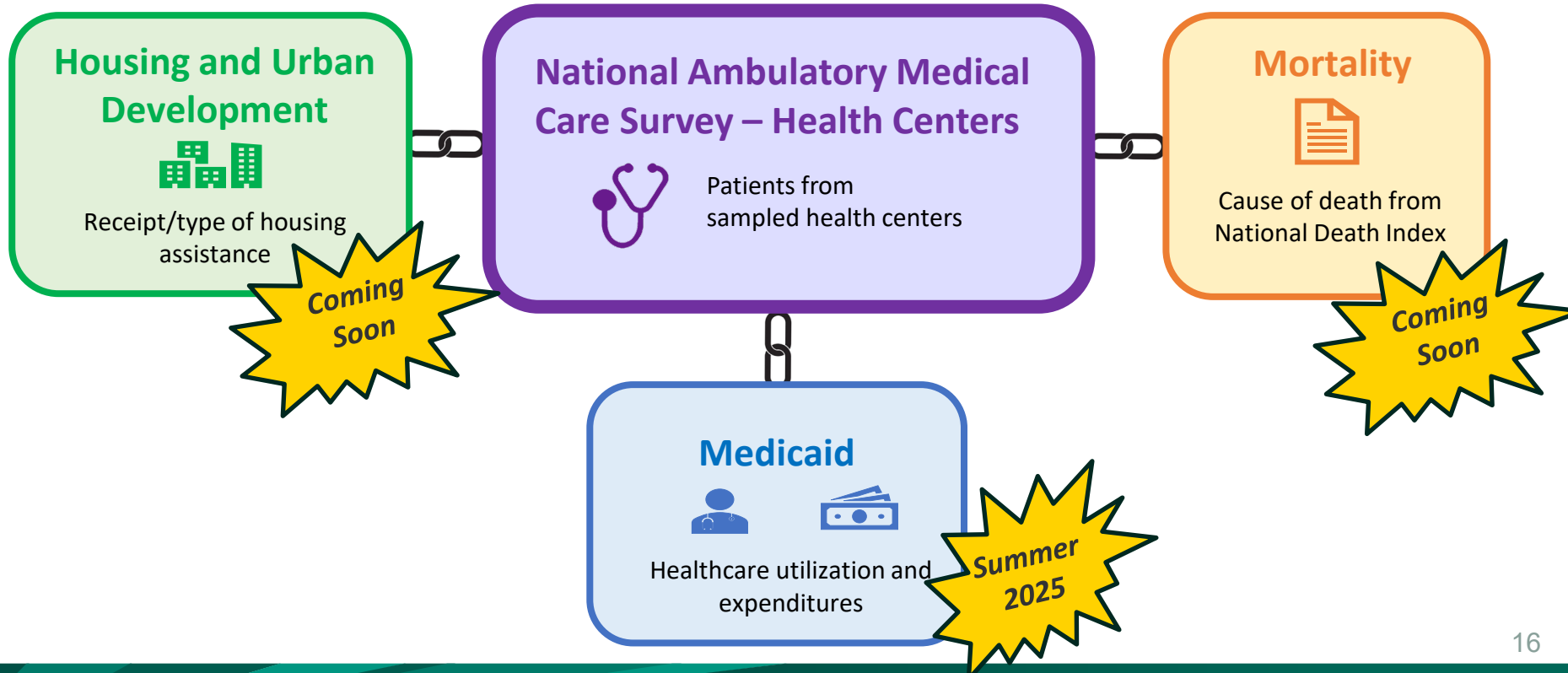
# Sample Sizes and Preliminary Linkage Rates, by Age and Sex

	Total Sample	Eligible for Linkage	Linked to HUD*	% Total Sample Linked*	% Eligible Sample Linked*
<b>Age</b>					
0-17	185,969	185,968	22,371	12.03	12.03
18-44	271,315	271,312	46,293	17.06	17.06
45-64	184,080	184,080	24,538	13.33	13.33
65 and over	85,020	85,019	10,072	11.85	11.85
Not calculated	848	5	0	0.00	0.00
<b>Total</b>	<b>727,232</b>	<b>726,384</b>	<b>103,274</b>	<b>14.20</b>	<b>14.22</b>
<b>Sex</b>					
Male	309,817	309,815	34,466	11.12	11.12
Female	414,112	414,109	68,525	16.55	16.55
Missing	3,303	2,460	283	8.57	11.50
<b>Total</b>	<b>727,232</b>	<b>726,384</b>	<b>103,274</b>	<b>14.20</b>	<b>14.22</b>

\* The number of patient records linked to HUD data may change during post-processing of the final files

# **Future NAMCS Health Center Component Linkages and Accessing Linked Data**

# NAMCS HC Component Data Linkages





# Accessing Restricted-Use NCHS Linked Data

- Linked data are available through the NCHS Research Data Center (RDC) and Federal Statistical RDCs managed by the Census Bureau

🏠 RDC	
Restricted Data	+
Location of Access	+
Proposal Process	+
Confidentiality	
Accessing Restricted Data	
Preparing for proposal Submission	
Providing the Public Use Data	
Output	
Fees and Invoicing	

## Research Data Center (RDC)

The National Center for Health Statistics (NCHS) operates the Research Data Center (RDC) to allow researchers access to restricted-use data. The RDC is responsible for protecting the confidentiality of survey respondents, study subjects, or institutions while providing access to the restricted-use data for statistical purposes. For access to the restricted-use data, researchers must submit a research proposal outlining the need for restricted-use data. The proposal provides a framework for NCHS to identify potential disclosure risks and how the data will be used.



### [Preparing for Proposal Submission](#)

1. [Restricted Data](#)
2. [Location of access](#)
3. [The Proposal Process](#)

### [Accessing Restricted Data](#)

1. [Confidentiality](#)
2. [Output](#)
3. [Publishing Guidelines](#)

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

Email: [rdca@cdc.gov](mailto:rdca@cdc.gov)

# RDC Proposals

- RDC Proposal Process info
  - <http://www.cdc.gov/rdc/b3prosal/pp300.htm>
- Website contains several helpful documents:
  - Standard Application Process
  - Completed sample proposal
  - Instructions to create a data dictionary
  - General guidelines
  - Data dictionary examples
  - Descriptions of common amendments

# NCHS Data Linkage Program

**Subscribe to the NCHS Data Linkage Program LISTSERV** to receive updates! Email a message to [list@cdc.gov](mailto:list@cdc.gov). Leave the subject line blank. In the body of the message, type:

– SUBSCRIBE NCHS-DATA-LINKAGE-PROGRAM lastname, firstname

Contact the Data Linkage Program: [datalinkage@cdc.gov](mailto:datalinkage@cdc.gov)

Contact the NAMCS Data Users Mailbox: [ambcare@cdc.gov](mailto:ambcare@cdc.gov)

Visit the Data Linkage Program website: [www.cdc.gov/nchs/data-linkage](http://www.cdc.gov/nchs/data-linkage)

Visit the NAMCS HC Component website: [https://www.cdc.gov/nchs/ahcd/namcs\\_hc.htm](https://www.cdc.gov/nchs/ahcd/namcs_hc.htm)

Thank you to Jill Ashman, Brian Ward, Sonja Williams, the NCHS Data Linkage team, and NORC at the University of Chicago for their contributions and support of this work!

# 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 presentation are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention or National Center for Health Statistics.

