Assessing Data Quality in Population-Based Surveys Linked to End-Stage Renal Disease Administrative Data

Jonathan Aram, Crescent B. Martin, and Lisa B. Mirel

NCHS Data Linkage Program

2022 FCSM Research and Policy Conference

October 26, 2022

Overview

This presentation will cover 5 topics:

1. Research question
2. Background, data sources, and data quality
3. Methods used for this data quality assessment
4. Results of comparison of estimates from linked data and national registry
5. Conclusions

Background: Research Question

- In November 2021, the NCHS Data Linkage Team released updated linked NCHS survey data and United States Renal Data System (USRDS) End Stage Renal Disease (ESRD) files with administrative data through 2018

- Understanding the quality of the linked data is important for interpreting analysis results

- Research question: how closely do estimates generated using the linked NCHS-USRDS ESRD data align with national prevalence estimates reported in the USRDS Annual Report?

Background

- National Center for Health Statistics (NCHS)
  - Nation’s principal health statistics agency
  - One of 13 federal statistical agencies
  - **Mission:** NCHS collects, analyzes, and disseminates timely, relevant, and accurate health data and statistics. NCHS products and services inform the public and guide program and policy decisions to improve our nation’s health.

Background: data linkage

- Linking survey and administrative records enhances both sources
  - Adds additional variables
  - Adds longitudinal outcome information
  - Reduces participant burden
  - Improves ascertainment of difficult-to-remember information

- Considerations when analyzing linked data
  - Potential reduction in sample size due to ineligibility for linkage
  - Potential for bias in estimates due to ineligibility for linkage

Background: survey data sources

National Health and Nutrition Examination Survey (NHANES)

A nationally representative, cross-sectional sample of the US civilian noninstitutionalized population, which includes a household interview followed by an examination in a mobile examination center that serves as an important source of information on the health and nutritional status of adults and children

National Health Interview Survey (NHIS)

A nationally representative, cross-sectional sample of the US civilian noninstitutionalized population, which includes a household interview that serves as an important source of information on the nation's health

Background: administrative data source

- ESRD
  - Permanent kidney failure requiring dialysis or a kidney transplant
  - Although <1% of Medicare beneficiaries have ESRD, the condition accounts for approximately 7% of Medicare spending

- USRDS
  - A national data system that collects information about ESRD in the United States – data are considered the “true values” in this analysis

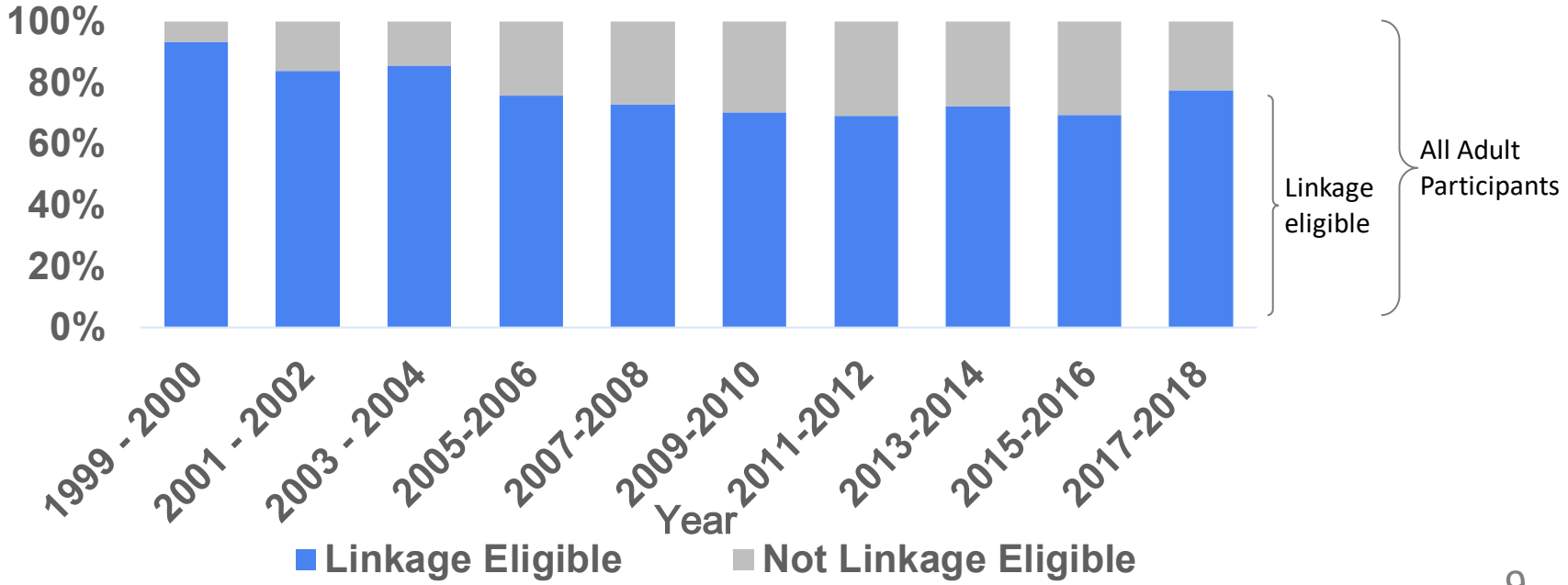
Background: linkage eligibility

- 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

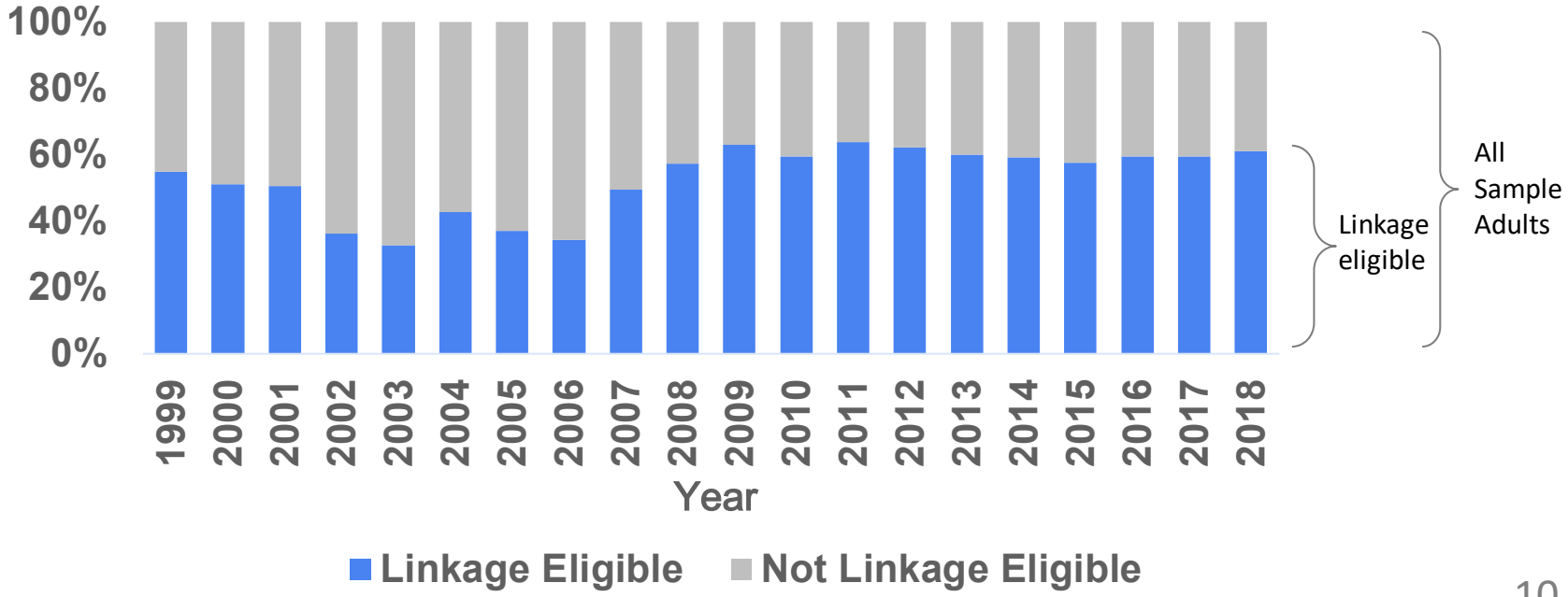
Background: ESRD linkage eligibility

Linkage Eligibility for ESRD, NHANES Participants Age ≥18

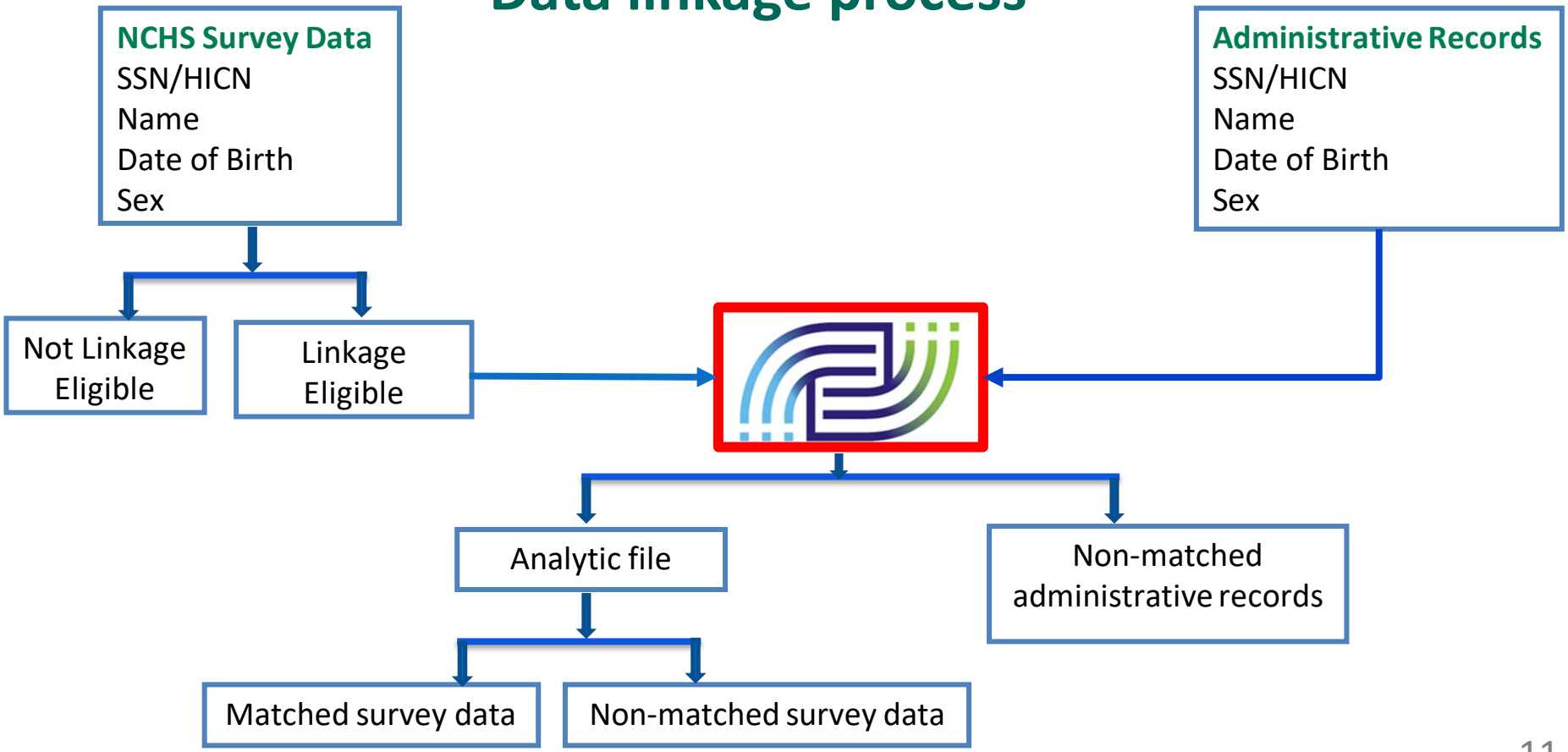


Background: ESRD linkage eligibility

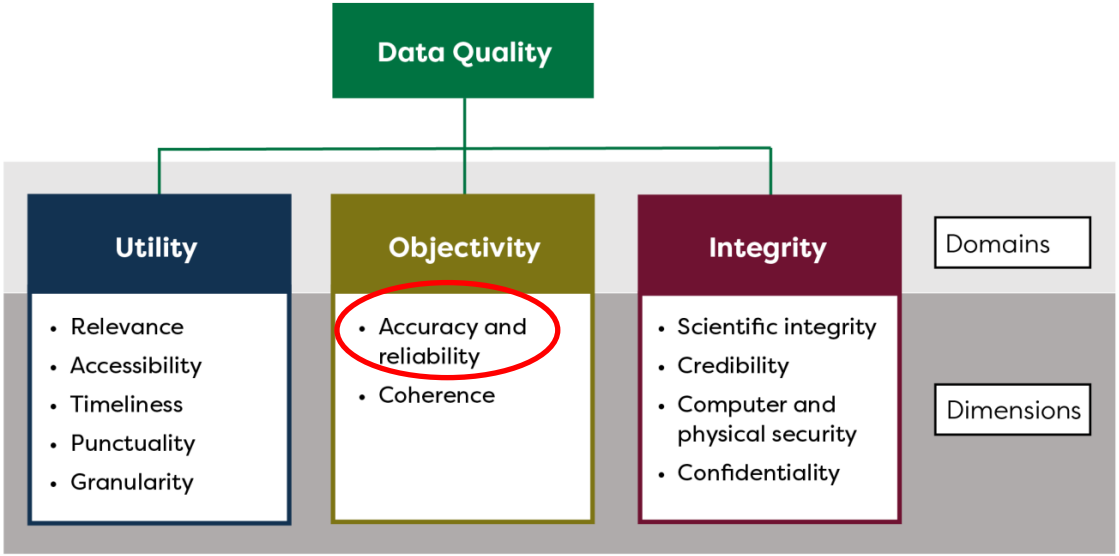
Linkage Eligibility for ESRD, NHIS Sample Adults



Data linkage process



Background: data quality



- The closeness of an estimate from a data product to its true value
- The consistency of results when the same phenomenon is estimated more than once under similar conditions

Federal Committee on Statistical Methodology. 2020. A Framework for Data Quality. FCSM 20-04, September 2020.

Methods: data quality assessment

- Prevalence of ESRD for linked survey participants
  - Date of ESRD diagnosis prior to NCHS survey interview date
  - From USRDS ESRD administrative data

- Grouping of survey years: 1999-2008 and 2009-2018
  - Survey years pooled to obtain ≥30 ESRD cases in each age category, based on age at interview
  - Same years were pooled for NHANES and NHIS

Methods: weight adjustment

- Sample weight adjustment for linkage eligibility
  - Some survey participants are not eligible for record linkage, creating the potential for bias similar to nonresponse bias

- Analyses use adjusted survey sample weights and Taylor series linearization to estimate standard errors, accounting for complex survey design

$$Model\ for\ adjusted\ weights:$$
$$Linkage\ Eligible = Sex * Age\ Group * Race/Ethnicity$$

Methods: Prevalence of ESRD

Linked Survey Data

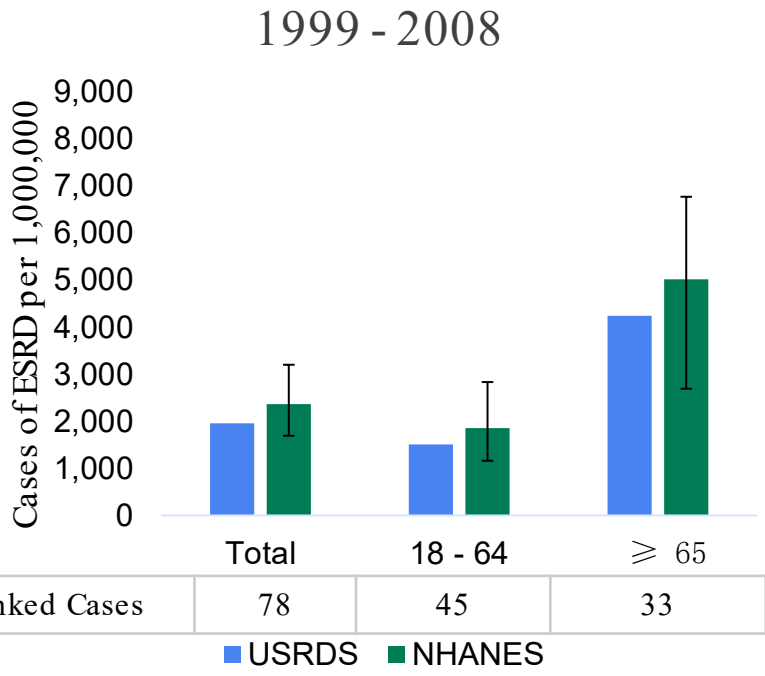
- Estimation of ESRD prevalence from linked survey data
  - Weighted percentage of US adults with ESRD and Korn and Graubard confidence intervals calculated for both NHANES and NHIS
  - Converted to cases per 1,000,000 for interpretability
- Estimates reflect US civilian, non-institutionalized, adult population

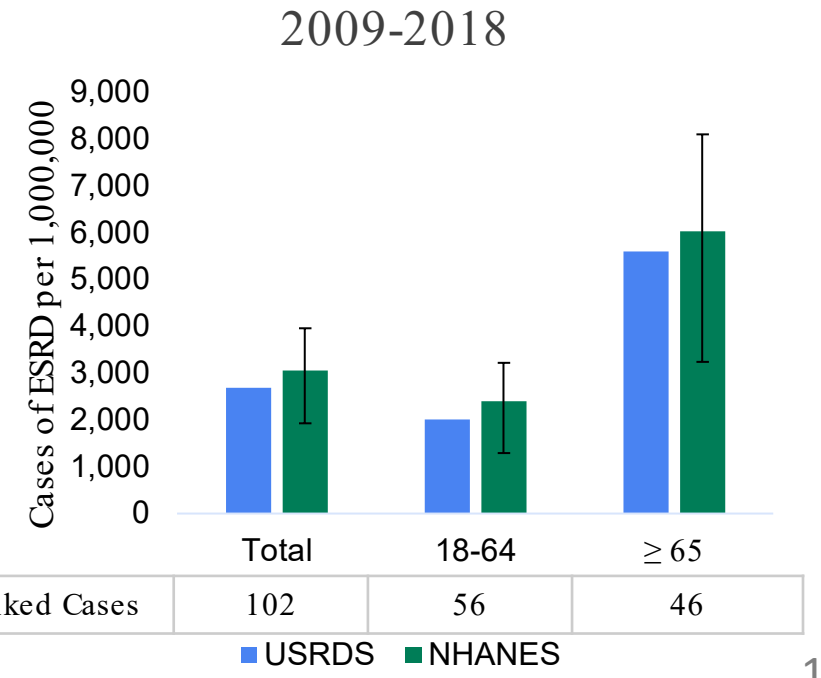
USRDS Administrative Data

- The 2020 USRDS Annual Report was used to determine the true reported prevalence of ESRD
  - ESRD prevalence estimates were pooled to align with the pooled survey data
  - Prevalence estimates excluded those missing certain demographic data
  - Proportions were converted to cases per 1,000,000 for interpretability

Results: NHANES and USRDS ESRD

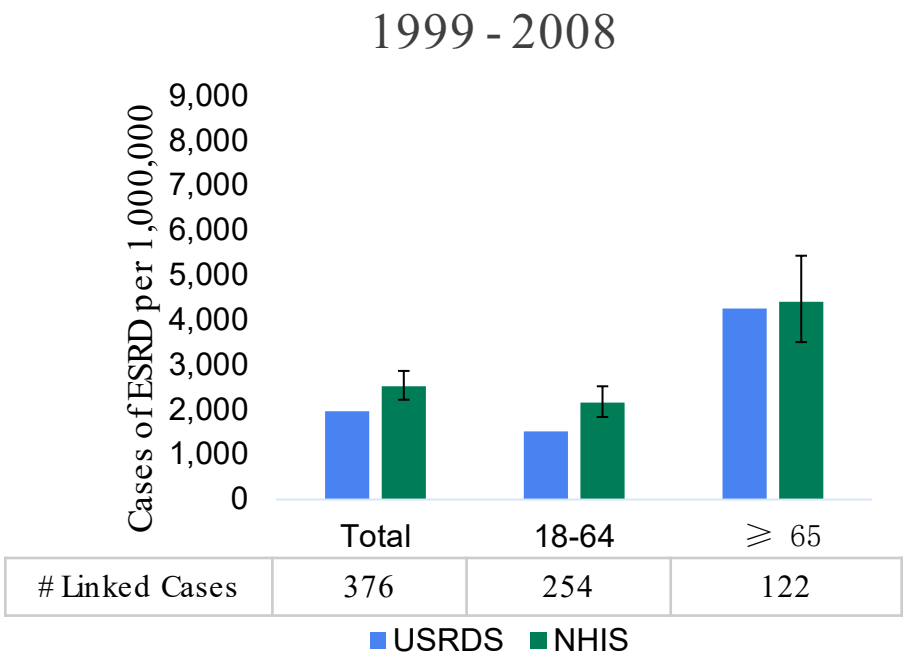
Comparison of Estimates from Registry and Linked NHANES Data

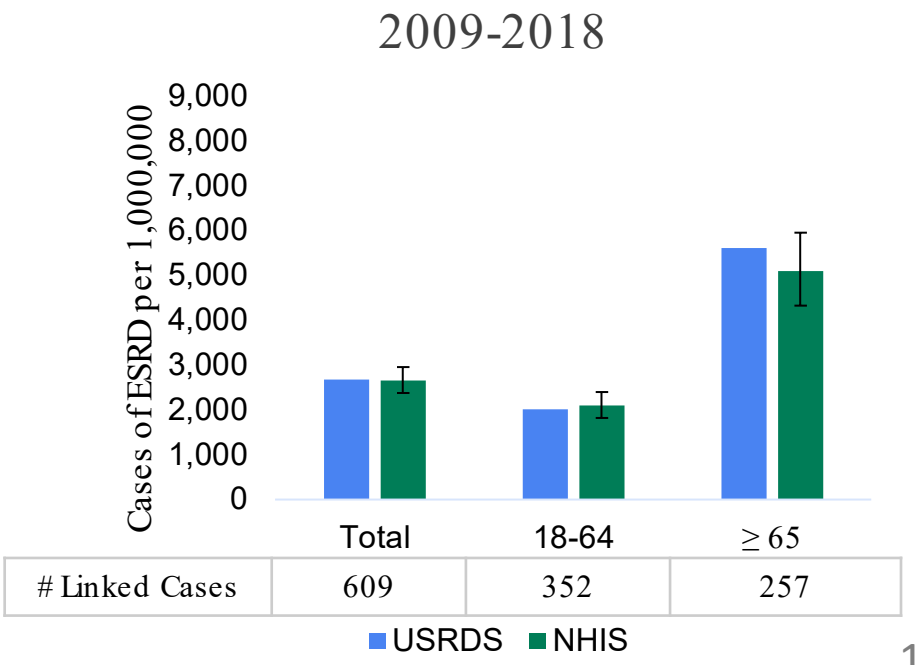




Results: NHIS and USRDS ESRD

Comparison of Estimates from Registry and Linked NHIS Data





Conclusions

- Estimates from the linked NHANES-USRDS ESRD data are similar to USRDS ESRD benchmarks
- There were some differences noted with the estimates from the linked NHIS-USRDS ESRD data
  - CIs overall and for 18-64 do not overlap with benchmarks in earlier years, when linkage eligibility was lowest
  - CIs do overlap with benchmarks in later years, when linkage eligibility was higher

Discussion

- Finding a benchmark that aligns with a linked data source is important to assess data quality
  - Benchmark comparisons are one dimension of assessing data quality based on FCSM framework
- Exclusion of cases with missing demographic information from the benchmark data sources may account for some differences in estimates
- Previous research has found that lower linkage eligibility rates coincide with higher “linkage eligibility bias”
  - Average linkage eligibility for the two time periods was
    - NHANES 79% and 71%
    - NHIS 44% and 61%

References

- Aram J et al. Assessing linkage eligibility bias in the National Health Interview Survey. National Center for Health Statistics. Vital Health Stat 2(186) 2021.
- Federal Committee on Statistical Methodology. 2020. *A framework for data quality*. FCSM 20-04. Federal Committee on Statistical Methodology. September 2020.
- National Center for Health Statistics, Division of Analysis and Epidemiology. The linkage of National Center for Health Statistics Survey Data to United States Renal Data System End-Stage Renal Disease patient data – methodology and analytic considerations.
- United States Rental Data System. 2020. *USRDS annual data report: epidemiology of kidney disease in the United States*. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2020.

More Information

NCHS Data Linkage Program:

datalinkage@cdc.gov

www.cdc.gov/nchs/data-linkage

Jonathan Aram: JAram@cdc.gov

Subscribe to our ListServ (updates on program including release dates): Send an email message to list@cdc.gov. Leave the subject line blank. In the body of the message, type or paste: SUBSCRIBE NCHS-DATALINKAGE-PROGRAM lastname, firstname where 'lastname, firstname' is your last and first name.

For more information, contact CDC
1-800-CDC-INFO (232-4636)
TTY: 1-888-232-6348 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.