National Center for Health Statistics
Data Linkage

Assessing Consent Bias in Linkage Studies

Lisa B. Mirel, Cordell Golden and Cindy Zhang
Special Projects Branch
Office of Analysis and Epidemiology
National Center for Health Statistics

Federal Committee on Statistical Methodology Research Conference
March 8, 2018

Outline

- Purpose
- Background: NCHS Data Linkage Program
- Changes in linkage eligibility over time
- Research question
- Methods to assess bias in estimates
- Results
- Conclusions

Purpose

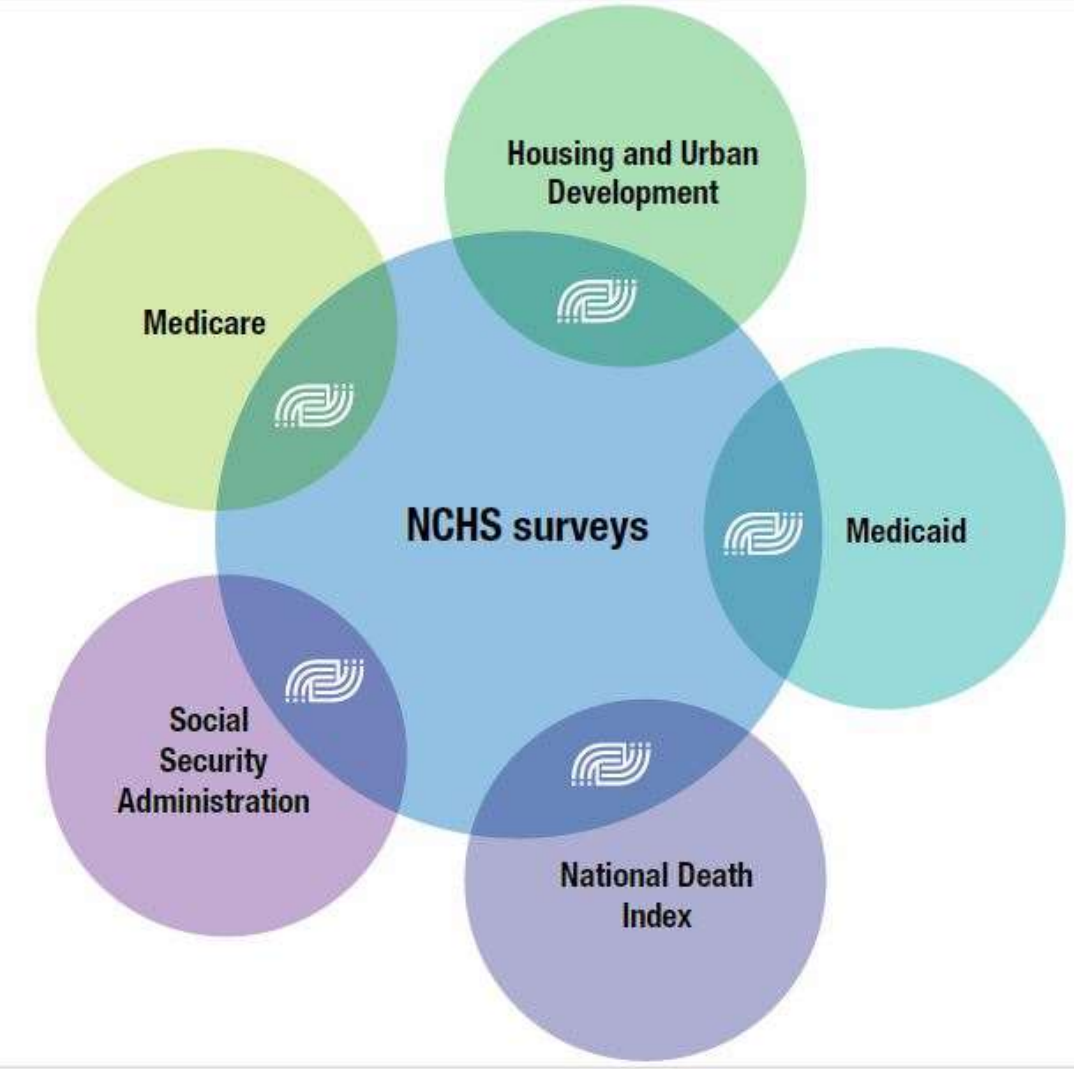
- NCHS only links survey participants who grant linkage consent
- Concerns if inference drawn from linked data represent the full sample
- Linkage consent rates may introduce consent bias with estimates
- In 2007 linkage eligibility criteria changed which increased linkage consent rates
- This talk addresses if changing the linkage eligibility had an impact on consent bias

Background: NCHS Data Linkage Program

- Designed to maximize the scientific value of the NCHS surveys
- Links survey data with vital and administrative records
  - Efficient means to add information
  - Allows analyses that would not be possible with each data source alone
- Includes the following NCHS Surveys :
  - National Health Interview Survey (NHIS)
  - National Health and Nutrition Examination Survey (NHANES)
  - National Health Care Surveys

More information about surveys: http://www.cdc.gov/nchs/data/factsheets/factsheet_summary.pdf

Current NCHS Data Linkages



A summary of the available linkages can be found at: https://www.cdc.gov/nchs/data/datalinkage/linkagetable.pdf

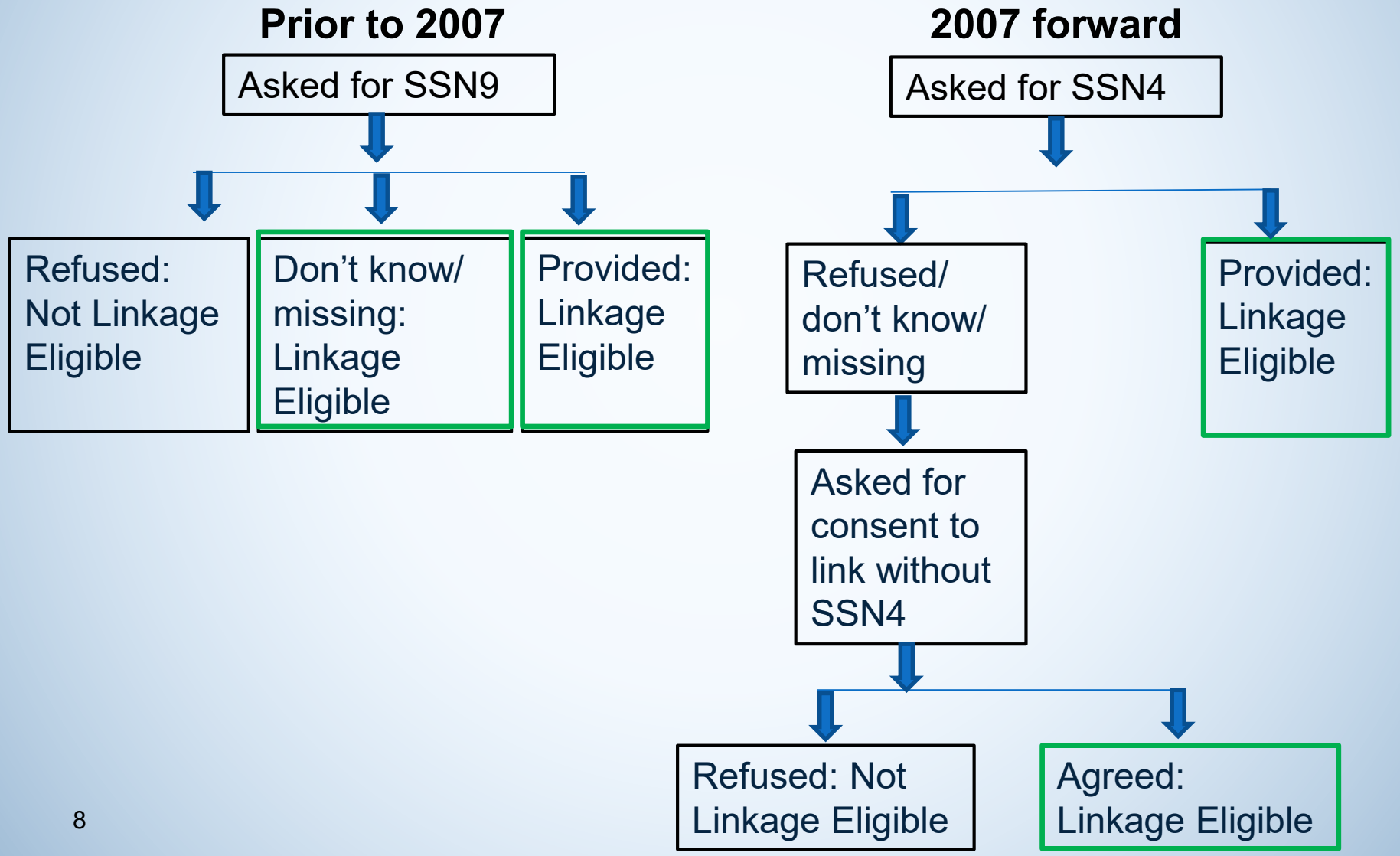
Linkage Eligibility

- Only linkage eligible (LE) survey participants are able to be linked
- Linkage eligibility varies based on survey and year
- Linkage eligibility depends on having sufficient personally identifiable information (PII) and not refusing data linkage
- The fact that not every one is linkage eligible could have implications for inference when using the linked data

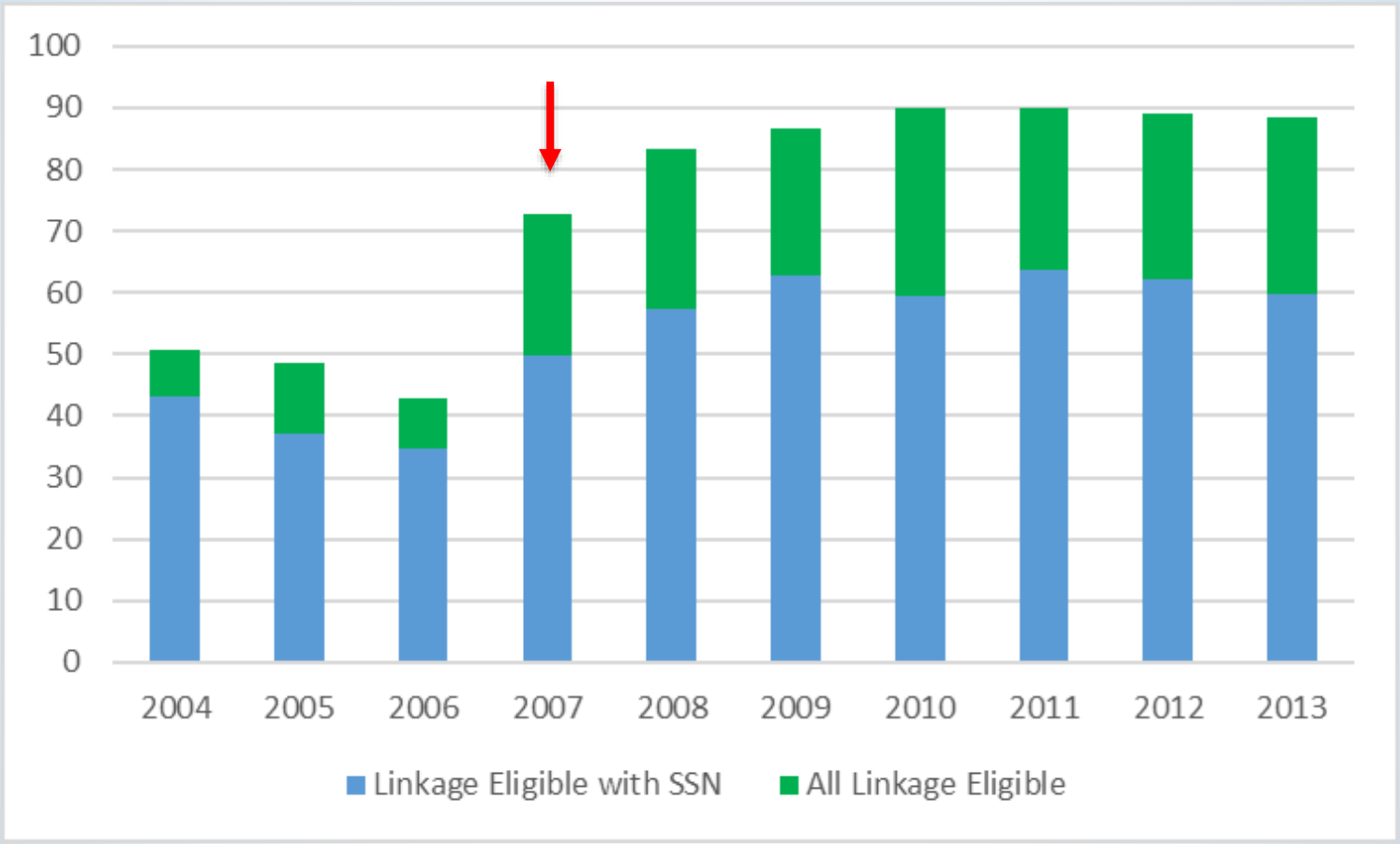
Changes in linkage eligibility over time

- NHIS linkage eligibility changed in 2007
- For PII collection:
  - NHIS only collects last 4 digits of Social Security Number (SSN) instead of 9
- For participants who refuse to provide last 4 digits of SSN:
  - NHIS asks a specific question about permission to link without SSN<*this was not previously asked*>

NHIS


Prior to 2007
Asked for SSN9
Refused: Not Linkage Eligible
Don't know/ missing: Linkage Eligible
Provided: Linkage Eligible
2007 forward
Asked for SSN4
Refused/ don't know/ missing
Provided: Linkage Eligible
Asked for consent to link without SSN4
Refused: Not Linkage Eligible
Agreed: Linkage Eligible

Percent of Participants that are Linkage Eligible



Note: this graph only includes sample adults

Research Question

Did the changes in linkage eligibility affect the bias in estimates for those who consented compared to the full sample?

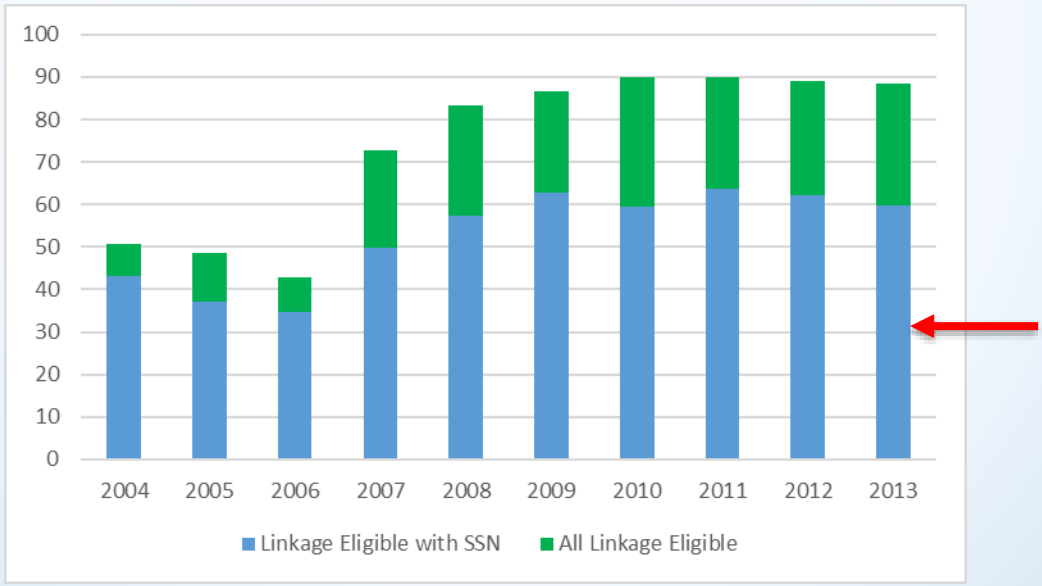
Methods

- Analyzed survey collected variables for sample adults age 18+
- Identified three groups:
  - All sample adults
  - Linkage eligible sample adults who provided SSN (9 or 4)
  - Linkage eligible sample adults
- Calculated estimates for demographic and health related variables
  - Verified questions were asked the same over time

Calculation of Relative Bias: Linkage Eligible with SSN

- Absolute Relative Bias=

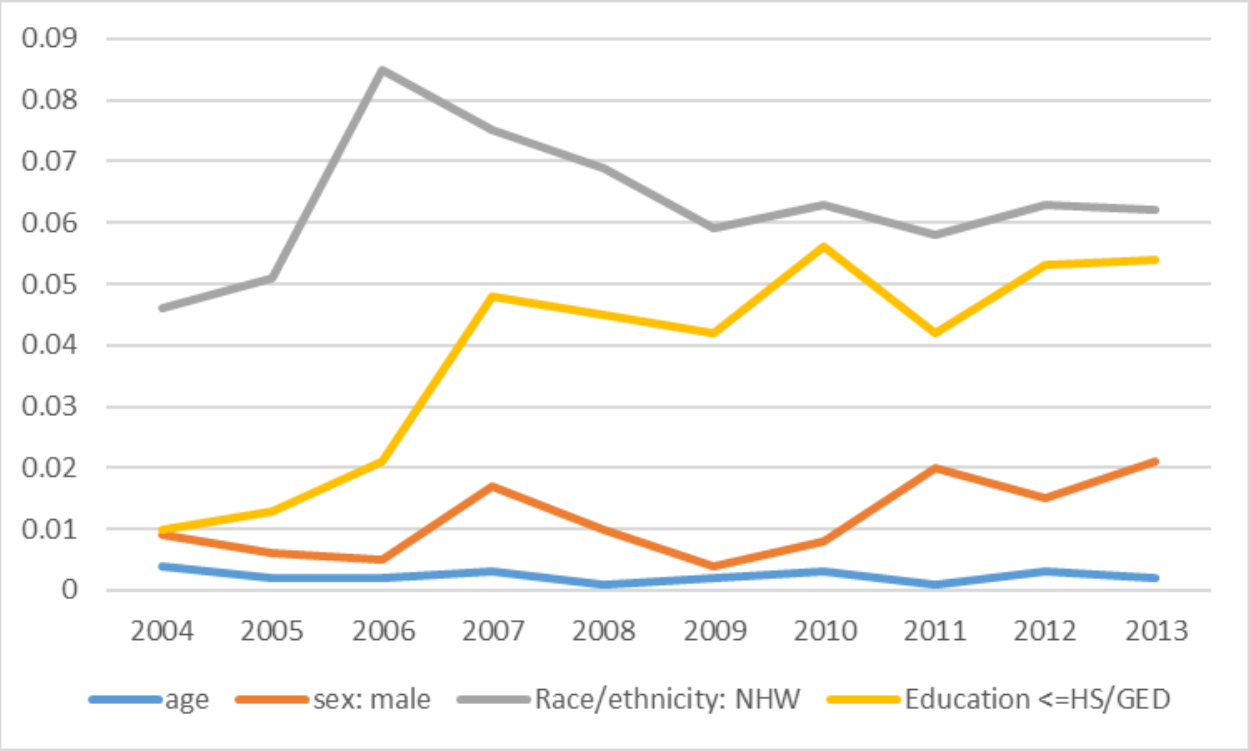
$$\frac{\text{Estimate of LE w/SSN- Estimate of all Sample Adults}}{\text{Estimate of all Sample Adults}}$$



Sakshaug, J. W., & Huber, M. (2016). An Evaluation of Panel Nonresponse and Linkage Consent Bias in a Survey of Employees in Germany. *Journal of Survey Statistics and Methodology, 4*(1), 71-93. doi:10.1093/jssam/smv034

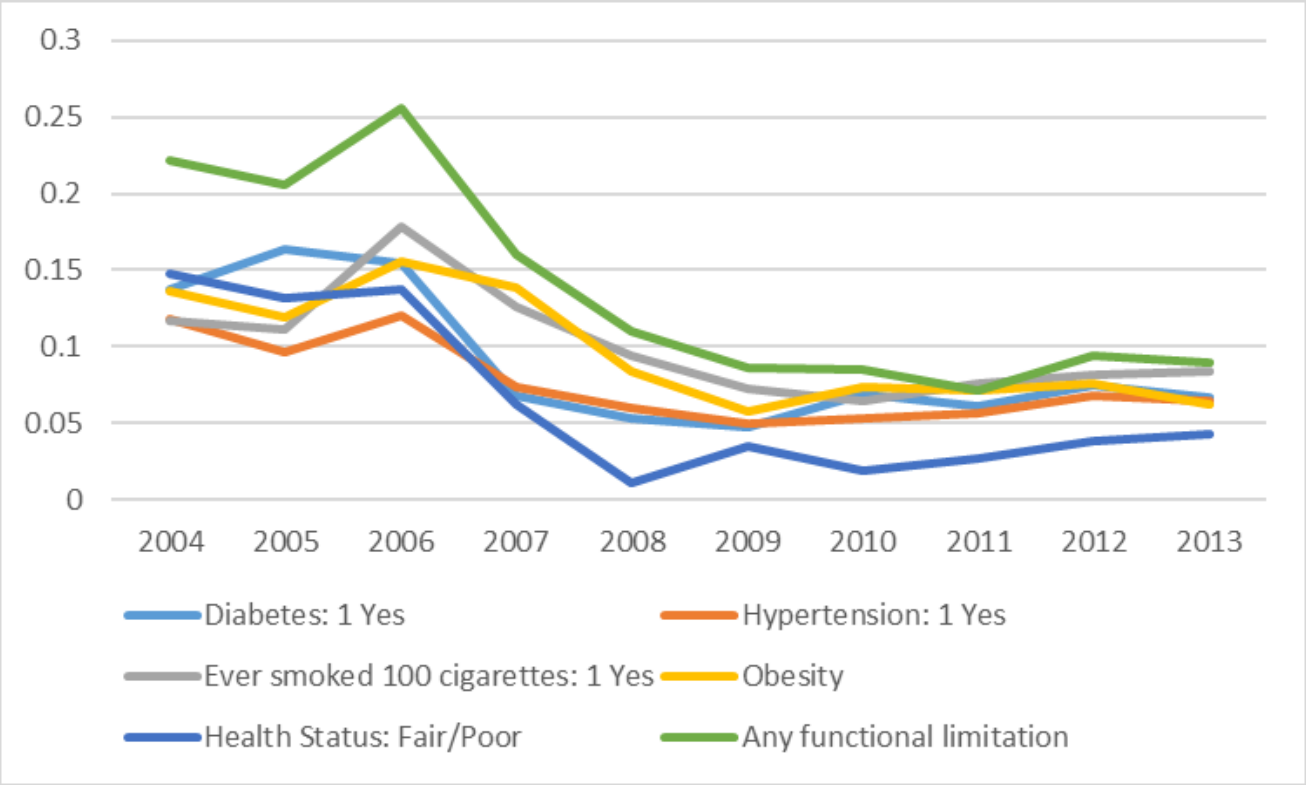
Relative Bias Demographics: Linkage Eligible with SSN

- Age, sex, race/ethnicity and education



Relative Bias Health Related: Linkage Eligible with SSN

- Diabetes, smoking status, overall health status, hypertension, obesity, and any functional limitation



Results Absolute Relative Bias: Linkage Eligible with SSN

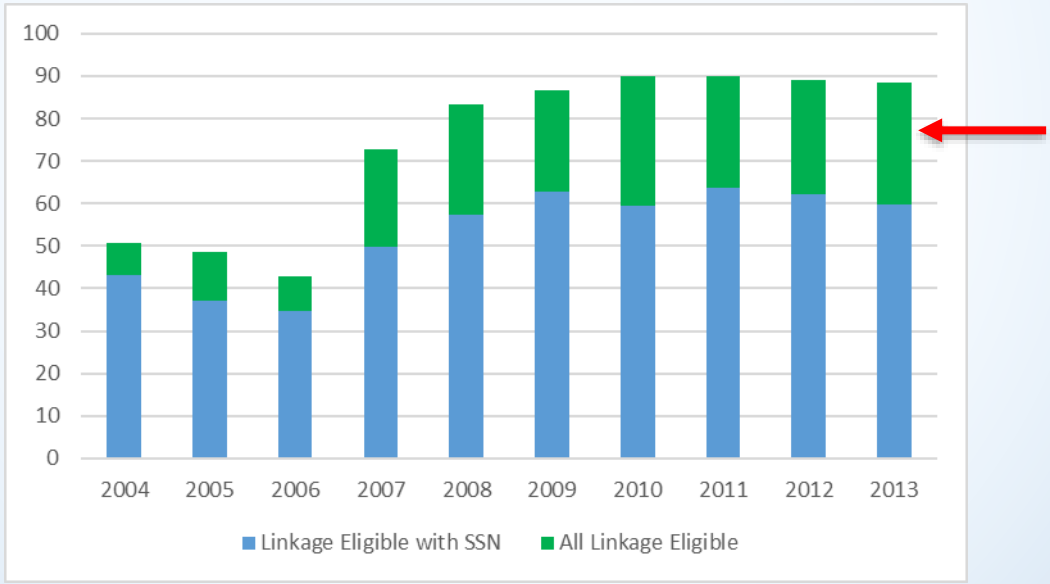
Variable Type	Maximum Relative Bias Before 2007	Maximum Relative Bias After 2007
Demographic	9%	6%
Health Related	25%	15%

- Education increased in relative bias after 2007
- The relative bias for the other demographic variables stayed relatively constant
- The relative bias for health related variables decreased after 2007

Calculation of Relative Bias: All Linkage Eligible

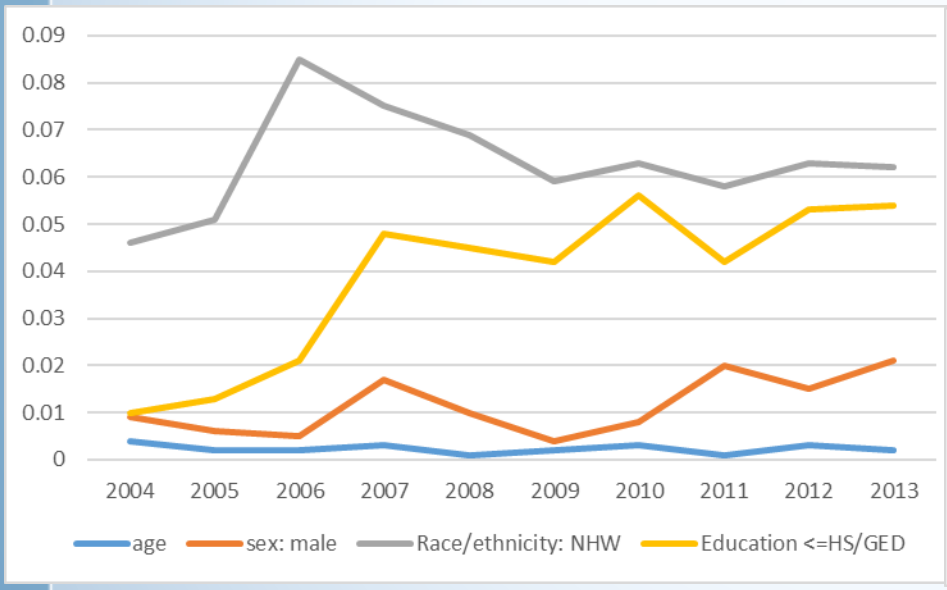
- Absolute Relative Bias=

$$\frac{\text{Estimate of Linkage Eligible- Estimate of all Sample Adults}}{\text{Estimate of all Sample Adults}}$$

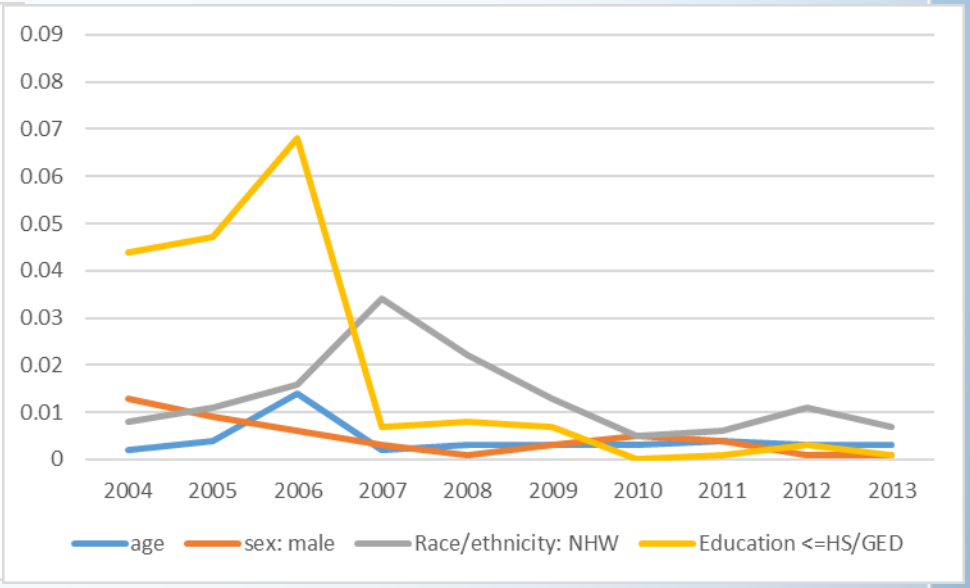


Comparing Relative Bias: Demographic Variables

Linkage Eligible with SSN

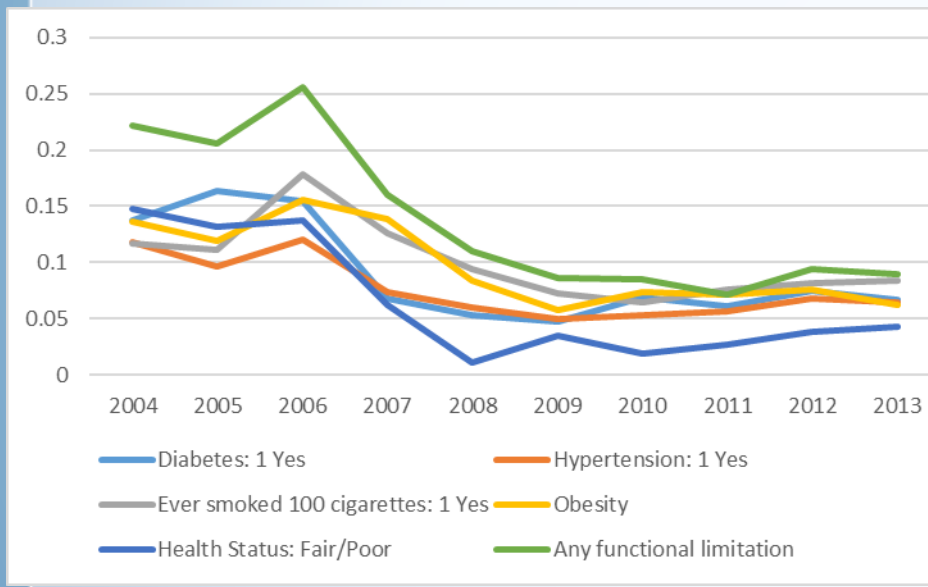


Linkage Eligible

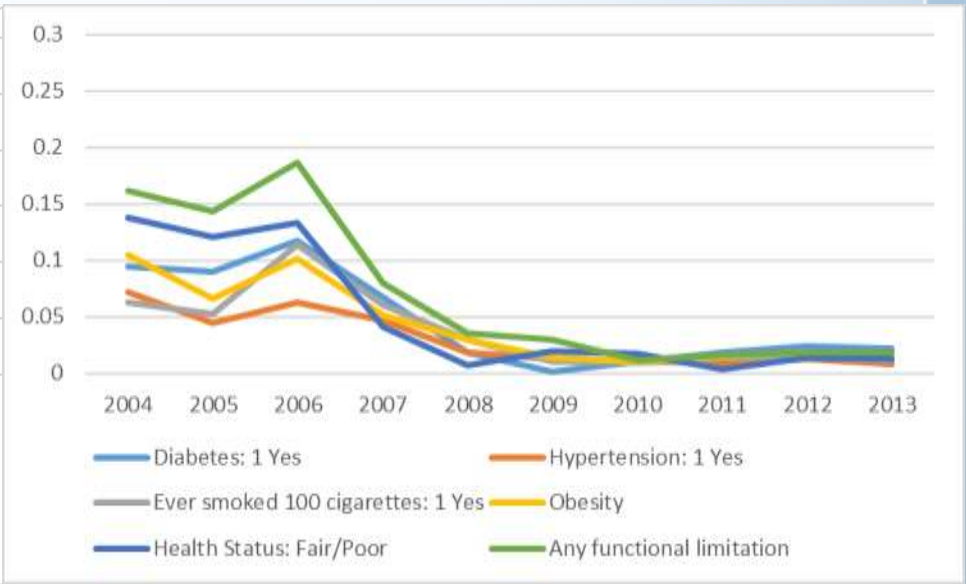


Comparing Relative Bias: Health Related Variables

Linkage Eligible with SSN



Linkage Eligible



Results Absolute Relative Bias: Linkage Eligible

Variable Type	Maximum Relative Bias Before 2007	Maximum Relative Bias After 2007
Demographic	7%	3%
Health Related	19%	7%

- Decrease in relative bias in 2007 when linkage eligibility criteria changed
- Since 2007 relative bias has remained relatively constant

Comparison of Relative Bias

- Estimates for all linkage eligible appear to have lower relative bias than when restrict to linkage eligible who also provided SSN
- Note: no statistical testing was done for comparisons

Conclusions

- Changing the linkage eligibility criteria in 2007 helped with both increasing sample sizes and decreasing linkage consent bias
- Some estimates remained unchanged but others were more affected by linkage consent
- Caveat: assessments of bias used survey variables and may not indicate bias for analyses using variables from the linkages (e.g. from the administrative data source)

Next steps

- Develop algorithms that incorporate all linkage eligible not just those who also provided SSN
  - More evidence of bias when restrict to linkage eligible who also provided SSN
- Assess bias with other non-health related variables from the survey
- Explore methods to mitigate potential bias for specific analyses with linked data
  - Weight adjustments
  - Imputation

Acknowledgements

- Thanks to Dean Resnick and the Data Linkage Team:

James Brittain | Eileen Call|
Adam Fedorowicz | Cordell Golden |
Dedun Ingram | Yu Sun | Jennifer Sayers |
Clinton Thompson | Yeats Ye |
Keith Zevallos | Cindy Zhang

Questions?

Thank you!

Email: LMirel@cdc.gov

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