Data Linkage for Income and Health Statistics

Discussion by

Hilary Hoynes

University of California, Berkeley and CNSTAT

May 18, 2022

Berkeley
UNIVERSITY OF CALIFORNIA

Stepping back: where does this panel fit in with CNSTAT's work?

Towards a Vision for a New Data Infrastructure for Federal Statistics and Social and Economic Research in the 21st Century

↓

Workshop 2: The Implications of Using Multiple Data Sources for Major Survey Programs

↓

Session 4: Data Linkage for Income and Health Statistics

Goals of the 3 CNSTAT Workshops

- Vision of the future of data collection for official statistics

- Data infrastructure (partnerships, linkages)

- Using multiple sources to produce statistics

- Data protection: Privacy & confidentiality

Goals of Workshop 2

Implications of Using Multiple Data Sources for Major Survey Programs

- Which programs might benefit?
  - Threatened by diminution of data quality
  - Administrative, private-sector data could improve timeliness, granularity, coverage
- Implications of using multiple sources (+ unforeseen consequences)
  - How will shift affect measurement of key indicators?
  - Measurement of subpopulations
  - Longitudinal data sources (linking)
- Data equity issues, hard-to-reach populations
- Data integration by design

Jonathan Rothbaum (U.S. Census Bureau) "The National Experimental Wellbeing Statistics Project"

Lisa Mirel (UC National Center for Health Statistics) "Realizing the Power of Health Data through Linkages"

Jessica Faul and Helen Levy (Michigan) "Administrative Data Linkages in the Health and Retirement Study"

The presentations:

- Span the most important agencies for income (Census) and health (NCHS) statistics

- Include longitudinal survey data (HRS) as well as cross sectional survey data

- Highlight groups engaging in leadership in this area

Goals of Workshop 2

- Which programs might benefit?
  - Threatened by diminution of data quality
  - Administrative, private-sector data could improve timeliness, granularity, coverage
- Implications of using multiple sources (+ unforeseen consequences)
  - How will shift affect measurement of key indicators?
  - Measurement of subpopulations
  - Longitudinal data sources (linking)
- Data equity issues, hard-to-reach populations
- Data integration by design

Threads to pull on from 3 presentations

Rothbaum, Census

SURVEYS	ADMINISTRATIVE DATA	National Statistics
CPS ASEC, ACS Decennial Census	IRS 1040 & Info returns, LEHD, Numident, safety net data, firm data, addresses	Income, poverty

1. Focus on continual improvement, with strong guide towards national statistics

2. Highlight what IS NOT in administrative data (race, education, etc)

3. Identify challenges in linkages & admin data (not perfect)

4. Highlight survey nonresponse and representativeness

Mirel, NCHS

SURVEYS	ADMINISTRATIVE DATA	National Statistics
NHANES, NHIS	Vital statistics (mortality), Medicaid/Medicare, safety net data, social security, addresses, VA	Healthcare access and utilization, health behaviors and conditions

1. Highlight how linked data have been used for evidence based policy making

2. Benefit of linkages: "allows for the construction of longitudinal events in previously cross-sectional datasets"

3. Where we need more linkages

4. Strong focus on health equity in linkage development and projects

Faul/Levy, HRS

SURVEYS	ADMINISTRATIVE DATA	National Statistics
HRS (longitudinal)	SSA, Medicare, Medicaid, VA	

1. Excellent attention to *how* a survey does this (consent, access)

2. Supports ongoing research to improve CMS data and statistics (population in institutions) and survey quality (attrition, measurement)

3. Ongoing balance of interest in making data available to outside researchers with AND data security

4. Continual improvement in data linkages