Increasing the use of state and local administrative data for evidence-building
Robert M. Goerge

Presentation to the

Commission on Evidence-Based Policymaking

January 13, 2016

ChapinHall at the University of Chicago

Policy research that benefits children, families, and their communities

Goals for presentation

- Chapin Hall background

- Partnerships with data providers (federal, state and local agencies) to facilitate access to data

- Importance and uniqueness of state and local data for federal evidence and policy purposes

- Discuss a pilot project to understand the demand for linkage to federal data sources, methods and use cases

ChapinHall

Chapin Hall at the University of Chicago

- Provide public and private decision-makers with rigorous data analysis and achievable solutions for improving the lives of children, youth and families

- Our audiences are policymakers and funders, government and private agency leaders, and researchers

- Our ongoing partnerships with public systems, institutions, organizations, and programs are a core strategy to achieve our mission

ChapinHall

Data that supports large scale evidence-building

- Focus on:
  - Linked administrative datasets or administrative data linked to survey data
  - Microdata on individual, families or providers (organizations or individuals) with personally identifiable information (PII)
  - "Universe" data or data on the entire population so that sub-state or sub-group analyses can be done
  - Historical data to do longitudinal analysis
  - Going to scale!

ChapinHall

Need a partnership with the data providers

- Most federal, state and local agency leaders and policymakers don't want to be "researched"

- **It's their data!! The vast majority (if not all) do not have to provide their data to researchers**

- Therefore, they need to see the benefit in providing access to their data to researchers OR to other government agencies

- They need to feel confident that they will be included and not be treated at arm's length in any specific research endeavor

- This is a different way of how research has been done in the past

ChapinHall

The partnership

- What is different in order to build a relationship
  - Need to include input from agency staff
  - Need participation from agency staff in the substance and design, but not the doing of it
  - Need them to review results before external audiences see them
  - Need them to have a chance to respond through actions or words to the research
- Only with such a partnership will they perhaps see a benefit and provide their data   **ChapinHall**

Why administrative data from states

- Many federal programs implemented by states report microdata to federal agencies.

- These data make up many of the federal datasets

- **This federal data, however, is very different in format and content** from the data that is maintained and analyzed by the state agencies themselves or by external parties that are provided the state's data

- The state data is what should be put into an administrative data clearinghouse

- Data may be transformed, de-identified, sampled or be restricted in its use when sent to the federal agency

Examples of federal datasets that are close, but richer when accessed from the state

- UI (Unemployment Insurance) quarterly wage data at the Census
  - All states, no sampling, up to date
  - Common format
  - However, not all states allow it to be used for non-LEHD purposes
- Adoption and Foster Care Reporting System
  - All states, no sampling, up to date
  - Common format
  - However, no identifiers, 6 month summaries

Transformed data example - TANF

- Temporary Assistance to Needy Families
- HHS Office of Family Assistance receives an <u>annual</u> summary record of characteristics, benefits, services and outcomes of individual and families on TANF from each state in a specific format so that the data is comparable across states and can be used for national reporting purposes
- Universe data from 30 states

ChapinHall

De-identified data – Child care subsidies

- Child Care Development Fund - CCDF
- Often the largest work support program in the state
- Recently, the reauthorization of the CCDF program removed the requirement of providing PII (Social Security Numbers) to HHS for parents and children participating in this program

ChapinHall

Sampled data – SNAP Quality Control

- Data for "conducting quality control (QC) reviews of Supplemental Nutrition Assistance Program (SNAP) cases"

- Statistical sample

- While other data could be linked to this sample, the size of the sample prohibits sub-state analysis

- Cannot look at SNAP receipt as an outcome

Restricted data - NDNH

- National Directory of New Hires contains
  - New Hires
  - Quarterly Wage (QW)
  - Unemployment Insurance (UI)
- Researchers can only use if it is de-identified, or
- "for research purposes found by the Secretary of Health and Human Services (HHS) to be likely to contribute to achieving the purposes of Part A or Part D of the Social Security Act."
- 2 years of data is maintained by HHS

ChapinHall

"Raw" state administrative data is richer

- Data pulled from state information systems for either their own analysis or analysis by external parties (Chapin Hall, CARRA …)

- Contains state-specific variables (fields) and identifiers necessary for the state to **implement** the program

- Richness of sub-year variation, non-summarized data and the ability to calculate and transform to fit research question

Multi-state studies with state data

- If we are not going to use federally help data which is comparable, what had to be done?

- Data has to be made comparable
  - Often little or no metadata
  - Requires researchers have significant subject matter and local service system expertise

- However, the richness of a particular state's data can enhance the analyses
  - A state may have more historical data

Using Linked Data to Advance Evidence-Based Policymaking

- Demonstrate an efficient way to link state and local data to Census-held data to answer important questions while protecting privacy

- Create compelling use cases for strengthening the Census linkage infrastructure to serve multiple levels of government

- Inform Federal, state, and local strategies for facilitating data linkage across programs

- Supported by the Laura and John Arnold Foundation

- CARRA is a collaborator with CH on this project

- Distributed an RFP for research projects linking 'PI-held' data to data held by the Census Bureau

ChapinHall

Response to RFP

- 17 responses to the call for full project proposals

- 25 responses to the call for letters of interests

- Individuals and organizations ranged from researchers in state and local government agencies, local and national advocacy organizations, research organizations, and universities

- Less than a handful were below par

ChapinHall

Primary interests for Census held data

- Employment

- Post-secondary education

- Public benefits: Medicaid, Medicare, SSI, SNAP, TANF, HUD

- Decennial Census, American Community Survey

- Topics included:
  - long-term follow-up of welfare reform experiments;
  - study of evicted households;
  - long-term follow-up of students K-12
  - public aid for post-secondary education
  - health outcomes over the life course

ChapinHall

Data being brought by investigators

- County-level integrated human services data
- County court records
- State birth certificates
- State post-secondary records
- State juvenile justice youth population
- State/county public benefit receipt
- K-12 student data
- Applications for state and federal financial aid for college

ChapinHall

Appendix A: Proposal Topics

- Education: preschool, K-12, post-secondary
- Employment
- Minimum wage
- Housing: homelessness, eviction, mobility
- Criminal justice
- Health
- Lead exposure
- Suicide
- Disaster preparedness

- Intergenerational poverty
- Multi-system families
- Child support
- Eligibility determination
- Food insecurity/SNAP
- Refugees
- Immigration
- Predictive analytics
- Taxation
- Federal regulation

ChapinHall

Appendix B: Proposal Methodologies

- Descriptive studies
- Needs assessment
- Eligibility/program take-up
- Long-term follow-up of RCTs
- Quasi-experimental studies
- Regression discontinuity
- Propensity score matching
- Difference-in-difference
- Longitudinal analysis

- Life course/trajectory models
- Policy analysis
- Predictive analytics
- Data linkage/warehouses
- Cluster analysis
- GIS/mapping

ChapinHall