Challenges in Linking State and Federal Data

Robert M. Goerge

The Promises and Challenges of Administrative Data in Social Policy Research
U.S. DHHS/ACF/OPRE
Washington, D.C.
October 1-2, 2015

Chapin Hall at the University of Chicago
Policy research that benefits children, families, and their communities
My plan for this session

- Describe specific applications of linking state and federal data
- Describe the challenges
- Describe strategies to address challenges
- Discuss both potential opportunities and obstacles
Integrated Database on Child and Family Programs in Illinois
Linking Census Survey data with Administrative Data – an example

• Using the Census 2001 ACS (Supplementary Survey SS01), we defined a base population of child care subsidy eligible families in 2000-01.

• We prepared state databases of child care subsidy, food stamp (SNAP) and TANF receipt

• We used quarterly UI wage data for the Census Bureau’s LED program.

• The state databases were “blended” with the SS01 and used to track quarterly child care subsidy eligibility, child care subsidy use, and employment outcomes for different groups of low-income families through 2003.

The Base Population

• All CCS-eligible families, with children under age 13, in Illinois, Maryland and Texas, in 2000-01, from the SS01. (The SS01 sampled households in each month of 2001 and asked about income in the prior 12 months.)
Results

- **Employment Outcomes: Employment Termination**

- CCS receipt was associated with longer employment spells in IL.
- TANF receipt was associated with shorter employment spells in IL.
- Low educational attainment was associated with shorter employment spells in IL.
- CCS recipients were less likely to lose CCS eligibility by exceeding the income ceiling in IL.
- Low educational attainment was associated with reduced odds of exceeding the income ceiling for CCS eligibility in all three states.
- Having at least three children under age 13 was associated with reduced odds of exceeding the income ceiling in all three states.
**Benefits:** Linking state data to federal data solves particular challenges

- Looking for outcomes in other states
- Adds outcomes not available in state data (e.g. employment)
- Take advantage of the SSN for easier linking of state data that has good SSNs
- Using a census survey dataset, one has a sample of the entire population, not just those in the administrative data
Benefits: Linking federal data to state data solves particular challenges

- Provides data on programs that is unavailable to federal agencies
- State data can enrich and improve survey data
  - e.g. National Survey of Early Care and Education
  - Meyer and Mittag
Data Challenges

- Correspondence between who is in a case or housing unit or family in each dataset—reconciling differences in “rosters”
- Correspondence between events
- Identifying data available to conduct linkage
IDENTIFYING MULTI-PROBLEM FAMILIES
Step 1: Select all cases meeting criteria

DHS
Criteria: Food Stamp cases with woman age 18-45

Alice
  - Bill
  - Clarice

Alice
  - Bill
  - Frank

DCFS
Criteria: All cases

Gillian
  - Henry

Alice
  - Opal
  - Bill

Alice
  - Nancy
  - Frank
IDENTIFYING MULTI-PROBLEM FAMILIES

Step 2: Link cases within agencies

DHS
Criteria: Food Stamp cases with woman age 18-45

Alice
  Bill
  Clarice
  Bill
  Frank

Alice
  Julie
  Kyle

---

DHS
Super cases (318,927)

Alice
  Julie
  Kyle

---

DCFS
Criteria: All cases

Gillian
  Henry

---

DCFS
Super cases (286,408)

Gillian
  Henry

---

DCFS
Criteria: All cases

Alice
  Bill
  Nancy
  Frank
  Opal

Alice
  Opal
  Bill
  Nancy
  Frank
IDENTIFYING MULTI-PROBLEM FAMILIES

Step 3: Link cases across agencies

DHS
Criteria: Food Stamp cases with woman age 18-45

<table>
<thead>
<tr>
<th>Julie</th>
<th>Kyle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clarice</td>
</tr>
<tr>
<td></td>
<td>Bill</td>
</tr>
<tr>
<td></td>
<td>Frank</td>
</tr>
</tbody>
</table>

DCFS
Criteria: All cases

<table>
<thead>
<tr>
<th>Gillian</th>
<th>Henry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DHS
Super cases (318,927)

<table>
<thead>
<tr>
<th>Julie</th>
<th>Kyle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clarice</td>
</tr>
<tr>
<td></td>
<td>Bill</td>
</tr>
<tr>
<td></td>
<td>Frank</td>
</tr>
</tbody>
</table>

DHS-DCFS
Super cases (502,165)

<table>
<thead>
<tr>
<th>Gillian</th>
<th>Henry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DCFS
Super cases (286,408)

<table>
<thead>
<tr>
<th>Gillian</th>
<th>Henry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
IDENTIFYING MULTI-PROBLEM FAMILIES

Step 4: Link agency service records to families

DHS
Criteria: Woman age 18-45 who received Food Stamps

Alice
Bill
Clarice
Frank

Alice

Bill
Clarice
Frank

Index
Member

DHS
Super cases (318,927)

Julie
Kyle

Gillian
Henry

DCFS
Criteria: All cases

Alice
Opal

Bill
Nancy
Frank
Opal

Clarice

DCFS
Super cases (286,408)

Julie
Kyle

Gillian
Henry

DHS-DCFS
Super cases (502,165)*

Alice
Nancy
Bill
Frank

Opal

Clarice

DHS
Criteria: Woman age 18-45 who received Food Stamps

Index
Member

Voice
    

Foster care
Substance abuse treatment
Mental health treatment
Incarceration
Juvenile detention
Substance abuse treatment

Department Records Obtained
- Department of Corrections
- Department of Juvenile Justice
- Department of Health Care and Family Services
- Department of Child and Family Services
- Department of Human Services

*DHS-DCFS “super cases were excluded if they had more than 50 members.
“Good luck getting the data sharing agreement through our lawyers....”
It’s nobody’s job to provide access

- There is no divine right to administrative data at either the state or federal level
- Requestor always has to actively engage an local, state or federal agency and apply to (often negotiate with) that agency to receive the data
- Welfare and benefit programs, human services, law enforcement, education, employment/wages, Medicaid, public health
Strategies to improve access

• Develop a relationship with an agency that can provide you with access
• Develop a relationship with someone who has access
• Use your political capital to have a policymaker (Governor or Director of agency) tell lawyers, program and IT staff to provide you with access
• Get a contract or do free work, but all that may require the above
Long way to go?

- Big increase in integrated, “research-ready” data happening to different degrees in different cities, counties, and states (AISP).
- There is a wide variation in who has access to the data that is being created and the quality of the data that is being built.
- It’s also taking many years to develop these efforts in states and cities.
- Best practices have not been disseminated to a sufficient extent.
- States often rely on large corporate vendors, who will only go so far, and government agencies don’t have the skilled staff necessary to take full advantage of the efforts.
- Much of this is because …
Silos of all kinds

- Across levels of gov’t – federal, state, county, city
- Within levels of government – agency silos
- Within agencies and across agencies – program silos
- Across domains – health, education, workforce/employment, law enforcement, anti-poverty
- Academic/professional silos – disciplines have their own interests
- Advocacy silos
- All work to the detriment of comprehensive data made available in a format conducive to research and analysis
Local public sector not in sync with feds

• 30 years ago, when there was less data, most public sector agencies had handfuls of analysts
• Now, we’re lucky if there is one per agency
• Increasing focus on compliance, but that’s not all that new
• At the same time, the federal government is requiring evidence-based practice and evaluation in many areas of social programs, which is a major challenge, given the lack of research expertise in these agencies
A new model for accessing data

• Data sits in government or private databases continuously being updated by the transactions completed by the government agency.

• When needed or periodically, data is transferred to an analytic engine that conducts a specified analysis – descriptive, multivariate, mapped …

• OR, it is posted on a data portal with API capability for anyone (?) to access

• **Challenge:** How to apply this to data about individuals who participate in programs in a political environment.
The end

• There are haves and have nots
• There are real barriers that lead to data not flowing to those that need it
• The nature of these barriers vary from sector to sector and place to place, but there are common themes
• These barriers can be addressed and those that don’t have must learn from those that have
• Incentives have to be put into place for all jurisdictions to use their data to get smarter about what they are doing –
• Reviewing all federal research projects so that they are effectively using administrative data
Robert Goerge rgoerge@chapinhall.org