Long-Term Outcomes (LTO) for Employment Evaluations
Using Administrative Data: Assessing Options

Rick Hendra, MDRC

July 18, 2018
Project team

• MDRC is the contractor for the project (Project Officer is Brett Brown)
• Key staff:
  • Richard Hendra – Project Director
  • Alex Pennington – Deputy Project Director
  • Kelsey Schaberg – Task Leader
Project motivation

- Many social programs are designed to have long-term benefits
- But many evaluations do not (or are not able to) track long-term outcomes
- Recent interest in making administrative data more accessible for research
  - Commission on Evidence-Based Policymaking
  - New federal and university-based data centers
  - ReImagine HHS Goal Two: Leveraging the Power of Data
Project goals

• Identify promising evaluations for long-term follow-up based on assessment of:
  • Practical and legal feasibility
  • Potential costs
  • History of linkages and past findings
  • Availability of relevant data

• Investigate relevant administrative data sources
  • Availability and content
  • Identifiers needed for linking
  • Restrictions
Considerations for Matching A Research Sample to Administrative Data

- Existence and type of PII available – destroyed PII makes future linkages impossible; certain admin sources require specific PII for matching
- Contract requirements – sometimes speaks to how PII can be used and by whom
- Consent – sometimes speaks to what data can be collected and for how long; lack of consent introduces additional IRB considerations
- Data sharing agreements – sometimes contain language around who can access data
- Administrative data sources – some impose restrictions on the contents of the returned file (e.g. exclude PII) or regulate how data can be worked with, which IRBs consider
Four-phase approach

1. SCAN
   Compile initial list of evaluations that meet basic criteria (complete)

2. CURATE
   Pare down evaluation list using additional criteria and collect information on related administrative data sources (complete)

3. COLLECT
   Develop data collection template and collect in-depth information on evaluations (ongoing)

4. ANALYZE
   Analyze collected information and assess the feasibility of long-term follow-up in synthesis memo (Summer 2018)
Phase 1: Scan

- Identified employment-related evaluations meeting basic criteria
  - Funded by OPRE, ASPE, or DOL
  - Completed in or after 1995 (including ongoing evaluations)
- Used syntheses, evidence reviews, clearinghouses, and meta-analysis studies
Phase 2: Curate

- Rated each evaluation using four key criteria
  - Each criterion worth 1-3 points
  - Overall rating was the average of the individual ratings
- Narrowed list down to 16 “major” evaluations
- Identified and collected information on relevant administrative data sets and data centers
## Rating criteria

<table>
<thead>
<tr>
<th>Study design</th>
<th>Study quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Non-experimental/implementation only</td>
<td>(1) Low causal evidence</td>
</tr>
<tr>
<td>(2) Quasi-experimental</td>
<td>(2) Moderate causal evidence</td>
</tr>
<tr>
<td>(3) Randomized experiment</td>
<td>(3) High causal evidence</td>
</tr>
</tbody>
</table>

*Study quality ratings came from DOL’s Clearinghouse for Labor and Evaluation and Research and/or ACF’s Employment Strategies Evidence Review*
Rating criteria, continued

<table>
<thead>
<tr>
<th>Strength of previous economic findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Few or no effects on key outcome(s)</td>
</tr>
<tr>
<td>(2) Unstable/mixed effects on key outcome(s)</td>
</tr>
<tr>
<td>(3) Consistent effects on key outcome(s)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Treatment contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) No participation impacts</td>
</tr>
<tr>
<td>(2) Participation impacts on some key program components</td>
</tr>
<tr>
<td>(3) Participation impacts on all key program components</td>
</tr>
</tbody>
</table>
Administrative data sources compendium

National Death Index (NDI)

**Content** Death and causes of death records

**Ownership/location** Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS), Division of Vital Statistics

**Funding agency** U.S. Department of Health and Human Services

**In data center?** Yes, Census Bureau’s Federal Statistical Research Data Center

<table>
<thead>
<tr>
<th>Time</th>
<th>1979-2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>Periodicity</td>
<td>Annually</td>
</tr>
<tr>
<td>Lag</td>
<td>12 months – a pilot program may shorten this, but state records may not be complete</td>
</tr>
</tbody>
</table>

**Geographic coverage** National (reported at state-level)

**Data availability**

**Measures** Partial list of variables: date of death, state of death, death certificate number, cause of death codes in ICD-9 or ICD-10 format

**Personally identifiable information (PII) returned to researchers?** Yes

**Matching**

**PII needed for matching?** "NCHS requires that the researcher’s finder file has at least one of these three combinations of variables in order to search for matches in the NDI data set:

1. First name and last name and month and year of birth
2. First name and last name and Social Security number (SSN)
3. SSN and date of birth and sex

NCHS counts records as “matches” if they meet one of the following seven criteria:

1. Exact match of SSN
2. Exact month and +/- 1 year of birth, first and last name
3. Exact month and +/- 1 year of birth, first and middle initials, last name
4. Exact month and day of birth, first and last name
5. Exact month and day of birth, first and middle initials, last name
6. Exact month and year of birth, first name, father’s surname
   - If the subject is female: Exact month and year of birth, first name, last name (on user’s record), and father’s surname (on NDI record).“

**Matching process** Detailed in Chapter 2 of the National Death Index User’s Guide

**Relevant restrictions** Data must be used only for “the purpose of health statistical reporting and analysis.” More information on restrictions is available on page vii of the National Death Index User’s Guide.

**Documentation available**

- National Death Index User’s Guide
- Poverty Action Lab, National Death Index
Phase 3: Collect

• Complete data collection template for each major evaluation
  • Pre-populate with publicly available information
  • Reach out to original research teams to fill in remaining fields
# Data collection template

<table>
<thead>
<tr>
<th>Evaluation:</th>
<th>Evaluation dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Lead evaluator</td>
<td>5. Does the evaluation dataset still exist?</td>
</tr>
<tr>
<td>2. Funder(s)</td>
<td>6. Location</td>
</tr>
<tr>
<td>3. Federal project officer</td>
<td>7. Was a public access or restricted use file created?</td>
</tr>
<tr>
<td>4. Contract status</td>
<td>8. What personally identifying information is available in the evaluation dataset and/or a restricted use file for a potential match to an administrative data source?</td>
</tr>
</tbody>
</table>
Data collection template, continued

9. Administrative datasets linked to in original study

10. Measures and data sources available in the evaluation dataset and/or the restricted use file

11. Data dictionaries/codebooks/instruments available for the evaluation dataset and/or the restricted use file

12. Data sharing agreements for the evaluation dataset

Feasibility

13. Have any extended follow-up data collection or analyses been done by your or another organization?

14. Did the study collect informed consent?

15. Is there any information we should know regarding permissions to use the data?

16. Linking to administrative data sources
### Data collection template, continued

<table>
<thead>
<tr>
<th>Study and findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>17. Study design</td>
</tr>
<tr>
<td>18. Study enrollment period</td>
</tr>
<tr>
<td>19. Number of sites</td>
</tr>
<tr>
<td>20. Site location(s)</td>
</tr>
<tr>
<td>21. Provider name(s)</td>
</tr>
<tr>
<td>22. Sample size</td>
</tr>
<tr>
<td>23. Target population</td>
</tr>
</tbody>
</table>

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>24. Intervention type and key services tested</td>
</tr>
<tr>
<td>25. Key outcome in original study (e.g. primary outcomes)</td>
</tr>
<tr>
<td>26. Strength of findings</td>
</tr>
<tr>
<td>27. Treatment contrast</td>
</tr>
<tr>
<td>28. Control group embargo</td>
</tr>
<tr>
<td>29. Follow-up in original study</td>
</tr>
</tbody>
</table>
Phase 4: Analyze

- Analyze full range of feasibility considerations
- Identify and prioritize available opportunities for linkage to specific evaluations
- Identify benefits and challenges of long-term data collection
  - Best practices for future evaluations
  - Concrete next steps
- Share findings in an internal memo (*August 2018*) and a presentation at HHS (*September 2018*)
Follow-up Work

- Identify and collect information on 9 additional evaluations
  - Mix of employment-related and child/youth development evaluations
- Publish administrative data sources compendium
- Start linking the data and doing the research
Early lessons

• Most research studies do not include critical details which speak to the feasibility of long run matches (such as whether match files exist)
  • Therefore, interviews are necessary
• Take advantage of existing clearinghouses to gather data and have clear criteria to help sift through the dozens of studies out there
Early findings

• Half of the evaluations assessed have potential for a long-term match.

• Most economic admin sources maintain data to measure key outcomes, but the amount of data available, costs, and timeline for acquiring the data varies.

• These evaluations may still have issues with consent, data destruction/access rules, etc.