The Family Self-Sufficiency Data Center

Robert Goerge, Emily Wiegand | June 18, 2018

APDU conference, Arlington, VA
Background

• Family Self-Sufficiency Data Center under a cooperative agreement with OPRE - Robert Goerge and Scott Allard are the Co-Directors

• Acknowledge the help of Michael Wiseman, Susan Golonka and Yun Song from OFA and Nicole Deterding and Mark Fucello from OPRE
Center projects to be discussed today

- State leadership engagement
- Federal -- Analysis of national TANF data
- State-level -- Multi-state TANF data analysis
The purpose of the Data Center

• Provide TA to states around the analysis of their data from programs that support the self-sufficiency of families
• We have worked intensively with a few states, but would like to offer help to more, if they would find it useful
• One way to do that is to build on what states have in common and have us address those commonalities
• Another way is to address special topics—which we hope to do in future workshops
Overview of engagement experience

• States are important potential partners in evidence-building initiatives.

• Without increased engagement by state and local agencies, the evidence-building agenda of the federal government, as outlined in the Commission on Evidence-Based Policy report, will fail.

• Proper engagement requires a well thought-out approach to ensure success
Factors that affect engaging state agencies

• management changes
• temporary crises
• budget issues
• personnel and capacity available for data activities
• perception of the value of data and evidence, and
• risk aversion around the creation of evaluative information
What it takes

• Initial approach and connecting (1-2 months)
  • Cold call or intro from 3rd party or prior relationship

• Specifying the project (1-2 months)
  • Getting on the same page in terms of mutual interests

• Data sharing agreements/MOU (1-12 months)
  • Always different, some agencies more aware that others

• Data transfer/Data Prep/Linkage/Metadata (3-12 months)
  • Depends on whether data is currently in use

• The work can start! (Maybe)
Activities to improve engagement

• Due diligence

• Working with political, policy, and technical government leaders and staff to understand the factors that hold back data use

• Providing leaders with a vision that connects to their experience in an authentic way

• Providing the agency with short-term, tangible benefits in terms of training and products

• Providing the agency with a range of solutions to persisting challenges

• Being flexible as a rigid approach would not fit all circumstances
Benefit to states

• Improving program administration can be the basis for overcoming all other barriers—it changes the premise under which state agencies and researchers are discussing the reasons to provide access to researchers.

• The requests for data should always include how states would benefit from the research. If researchers cannot describe the benefit for a state providing the data, they should perhaps not get access.

• Training is another concrete benefit. The Coleridge Initiatives Applied Data Analytics classes are an intensive multi-session, hands-on that provided training to government agency staff. Chapin Hall’s Advanced Analytics classes provide training on analyzing child protective services data.
About the TANF Longitudinal Data

• Used state-submitted caseload data from 29 states that submitted universe data consistently over our study period
• Study spanned fiscal year 2009 to fiscal year 2016
• Able to identify cases longitudinally using combination of case ID and state (though this was not without challenges)
• Derived spells of benefit receipt from monthly snapshots of active cases, allowing us to identify not only monthly caseloads but also entry and exit counts
Research Questions

• How did caseloads change in size during and after the recession?
• How did these trends vary by state and case type?
• What kind of analyses are feasible using state quarterly data submissions to OFA?
National Caseload Trends Reflect the Recession...

- National caseload trends over the period from fiscal year 2009 through fiscal year 2016 match what we would expect given the macroeconomic climate during the period.
- Entries were highest at the beginning of the period and steadily declined.
- Caseloads peaked in 2010 and likewise declined.
...Especially for One-Parent Cases

January 2011:
55% one-parent cases
4% two-parent cases
41% child-only cases

September 2016:
44% one-parent cases
2% two-parent cases
54% child-only cases
Some Evidence of Seasonality in Caseloads

Entries peak in late summer and ebb in winter.
National Trend in Exits

Figure 3: National caseload exits
National Net Change

Figure 4: Net change of national caseloads
Setting the foundation for multi-state collaboration

• Understanding state data needs

• Building out data models

• Experimenting with code sharing
Cross-state collaborations

• Get states working on similar questions in different environments:
  • Are there common roadblocks?
  • If so, can we source (or develop) common solutions?

• Rather than ask what states need, allow them to show us (and themselves and each other).
Other benefits

• Examples are motivational and interesting!

• There are a lot more similarities between agencies than between agencies and non-governmental research organizations.

• Collaboration creates opportunities for peer benchmarking.
What does it look like?
Moving from Work Participation to Self-Sufficiency
A Research Proposal from the Family Self-Sufficiency Data Center

The Family Self-Sufficiency Data Center (FSSDC) is funded by the US Department of Health and Human Services Office of Planning, Research, and Evaluation to facilitate use of administrative data by researchers and administrators to improve understanding of and identify methods for increasing family well-being. The FSSDC is seeking state or local research partners for a study piloting the use of administrative data to explore self-sufficiency outcomes of TANF and/or SNAP recipients. This project’s goals are:

1) To demonstrate a use case for family self-sufficiency data to inform and improve program administration;
2) To define a set of metrics for benchmarking exercises;
3) To clearly and transparently document how these analyses are performed (including providing code and detailed definitions) so that other agencies can duplicate these metrics on their own data.

Participating agencies will receive:

- Extensive technical assistance from the FSSDC, including assistance with data preparation, data sharing, and/or data analysis, as appropriate;
- The opportunity to collaborate and develop a common understanding of best practices together with other partner agencies;
- State-specific results addressing the research questions outlined in this proposal. Results will not be published or disseminated in any way, except at the discretion of the partnering agency.

Potential partners should have access to the administrative data sources outlined in the “Required Data” section of this document, and either 1) the ability to share data with the FSSDC or 2) capacity to conduct the analysis in-house, in close communication with the FSSDC team and using consistent logic and definitions.

Anyone wishing to learn more about this opportunity should contact the FSSDC research team at fssdc@champaign-il.org by August 31, 2017.
Cases Active as of 7/1/2015

Cases Active as of 7/1/2015 closed before 6/30/2016

Cases Closed Due to Earnings
- Reopens in less than 3 months
- Reopens after 3-6 months
- Reopens after 6-12 months
- Remains closed

Cases Closed Due to Reasons Other Than Earnings
- Reopens in less than 3 months
- Reopens after 3-6 months
- Reopens after 6-12 months
- Remains closed

Wage Data

Enhanced analysis of outcomes for closed cases
Primary Dataset

This dataset is unique on case ID/month, which is to say that it includes one record per case per month. It should include all active case months within the study window (June 2015 – August 2017); in other words, for any case that is active any point in this time frame, there should be a record for every month for which that case was active. Optionally the case can include records coded as “inactive” for months where the case was not actively receiving TANF.

The data dictionary presented in Table 1 captures a minimum dataset to facilitate analysis. This dataset should be created using code that can be saved to be easily amended and re-run. That will make it possible to add additional case- and even individual-level characteristics to allow for further rich analyses.

Table 1. Primary Dataset Data Dictionary

<table>
<thead>
<tr>
<th>Field</th>
<th>Data Type</th>
<th>Code Values and Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>caseid</td>
<td>character</td>
<td>State-specific -- needs to uniquely identify a case over time.</td>
</tr>
<tr>
<td>year</td>
<td>integer</td>
<td>Four digit year corresponding the (calendar) year when the data was subset.</td>
</tr>
<tr>
<td>month</td>
<td>integer</td>
<td>Integer value corresponding to the month when the data was subset.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Indicates whether the case was or was not receiving TANF benefits during the month.</td>
</tr>
<tr>
<td>tanf</td>
<td>boolean</td>
<td>If the dataset only includes observations for active months, this indicator should always be set to 1.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0 = inactive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 = active</td>
</tr>
</tbody>
</table>
Frequently Asked Questions (FAQ) about Data
Moving from Employment to Self-Sufficiency Multi-State Research Project

This document presents common questions that have arisen as states prepare data for this project. We provide responses to these questions based on conversations between individual states and the FSSDC project team. We share these answers in the interest of enabling consistency; if you disagree with one of these answers or have made a different decision in your state please let us know and we can discuss.

We plan to continue populating the FAQ as additional questions arise and will share updates with the group periodically.

Defining Population:

Question: Should state diversion programs and other cash assistance programs be included in the analysis?

Answer: Each state can define "TANF recipients" as narrowly or as broadly as appropriate for program management purposes. However, we recommend including a flag to indicate cases that are not federally reported to allow for maximum flexibility in the analysis. In the summary statistics they can be added and subtracted from the overall analysis for comparison.

Retroactive Payments and Recoupments:

Question: Should the time periods of retroactive payments be counted as active time periods for the case? (i.e. if a retroactive payment for May is made in June, should May be coded as active or inactive?)

Answer: Do not include months when a case received payments retroactively as active time periods of the case.

Question: If a payment for a month was later recouped, should the time period be changed as inactive?

Answer: Ignore recoupments, count these periods as active time periods for the case.

Work Eligibility:

Question: What fields should we use to classify a case as work eligible?

Answer: We propose using your state’s version of the federal work eligibility definition. However, we understand that states have different approaches to defining work eligibility. We encourage sharing your definitions and developing a consistent framework for the project.

Defining Income:

Question: How should we define income for the purpose of the project?

Answer: Each state can define "income" as broadly or as narrowly as appropriate for their program management purposes. However, we recommend developing a consensus on a common definition to enable comparison across states.
Spells Creation Demonstration Guide

This document is a reference guide for the spells creation code provided by the FSSDC team for the research project on self-sufficiency outcomes. We will introduce this material on the phone to each state, and you are not expected to read this document prior to that demonstration. However, please refer to the “Installation Guide” document and install the necessary software prior to the demonstration.

Description of Files

The R File

The “calculate_spells.R” file is the code file written in R to transform monthly extracts of case information into a spells format more suited for longitudinal analysis. A spell represents a period of time when the characteristics of a case remain the same. At a minimum, a spell can be defined as the period for which a case is active, but other case characteristics may be optionally considered, such as if a case changes case type or the size of the household changes.

The Input File

For the demonstration, we provide an input file of mock monthly extracts called “input.csv”. It has the following basic format which should match the format of the file you are preparing for analysis:

<table>
<thead>
<tr>
<th>caseid</th>
<th>year</th>
<th>month</th>
<th>tafn</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2016</td>
<td>01</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2016</td>
<td>02</td>
<td>1</td>
</tr>
</tbody>
</table>
Resources for program administrators and researchers looking to use data for family self-sufficiency research.

| Branch: master | New pull request | | Find file | Clone or download |

| pmack1 | Merge pull request #2 from chapinhall/doc_fix | Latest commit #f555a on Jul 31, 2017 |

- create_spells: fix README
- sample_TANF_data: making edits to main page
- tutorials: making edits to main page
- LICENSE.md: Changes to repo structure, expanding READMEs
- README.md: making edits to main page
- logo.png: Changes to repo structure, expanding READMEs

---

**Family Self-Sufficiency Data Resources**

This repository has been created as a way to share resources for program administrators and researchers looking to use data for family self-sufficiency research.
Family Self-Sufficiency Data Center
Moving from Employment to Self-Sufficiency Multi-State Research Project
Specific Research Questions – 2/21/2018

Note that some of these will require modifications or additions to the current data structure to comprehensively answer.

What working recipients earn sufficient income to close cases?
- What percentage of cases close within a certain period of time after opening?
- Of cases that close, how long were they open before closing? (**raises questions about first closing v. repeat closings)
- What are the characteristics of TANF households (all households or certain subpopulations of households) at case closing?
- What are the reasons for case closing, and how do they compare to household or case characteristics, or the duration for which the case has been open?
- What is the distribution and what are the characteristics of TANF households that have earnings? That don’t have earnings?
- What does the earnings history/trajectory of households look like?
- How do closed cases with earnings differ from closed cases without earnings?
- How do any of the above questions vary by household size or other case or individual characteristics?

Do households that earn their way off benefits main self-sufficiency, or do the cases reopen?
- What proportion of cases (all cases or certain subpopulations of cases) that close reopen within ___ period of time?
- What are the characteristics of cases that do or do not reopen within ___ period of time?
- What is the rate of churn in case closings? Are there certain patterns of cases that close for short windows only to subsequently reopen?
- Do time limits affect any of these patterns?

Do the workers in households closed due to earnings maintain a stable income after case closings?
- What are the earnings of closed cases (all case closings or certain subpopulations of case
Research Project on Self-Sufficiency Outcomes: Analysis Design

This survey solicits feedback from the participating states about initial priorities and structure for the FSSDC analysis. We appreciate your feedback and ask that you reply by Friday, March 9.

Please contact Emily Wiegand (ewiegand@chapinhall.org) or another FSSDC team member if you have any questions or comments.

Which state do you represent?

Who is completing this survey?

The first step in the analysis will be to characterize the cases in each group based on a variety of characteristics. Which of the following characteristics are you interested in assessing, and for which do you have readily accessible data? Check all that apply.

- Family size
- Household composition (i.e., number of adults)
- Head of household’s employment history
- Head of household’s educational attainment
- Quarterly wage history
- History of TANF receipt
- Geography/location
- Other - please list:
Transformation Steps:

1. Transform the Primary Dataset into spells with wage information appended. The FSSDC team has provided code that creates a spells view of the monthly case information in the Primary Dataset appended with the wages of the household at the start and end of the spell from the Wages Record Dataset. This spells transformation allows for a flexible approach to all of our case closings analyses.

2. Identify an exit cohort population. We will focus on cases that closed between July 2015 and December 2015 limiting our analysis to the first time a case closed in the cohort timeframe. The spells created in step 1 are sorted by the startMonth and startYear fields for each caseid. Select the first row for each caseid where the spells is active with an end date between July 2015 and December 2015. Add an indicator field called “wageThreshold” where the field is true for rows where the end wage is above the analysis threshold (TBD) and false otherwise.

Sample Outputs:

1. Examine the variation within the exit cohort’s closing wages (active end wage). Below are sample outputs for this analysis:

<table>
<thead>
<tr>
<th>Quartile</th>
<th>Wage at Closing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>0</td>
</tr>
<tr>
<td>2nd</td>
<td>0</td>
</tr>
<tr>
<td>3rd</td>
<td>200</td>
</tr>
<tr>
<td>4th</td>
<td>400</td>
</tr>
</tbody>
</table>

2. Examine the length of time a case remains closed by whether it was above or below the wages threshold at the time the case closed.
Things that may surprise you

• It’s not about complex analytics.

• Capacity and sophistication of data use are not a clean continuum.

• States want research partners.
Recommendations

• Encourage ongoing collaborations among state and local agencies and researchers to jointly address the barriers in using administrative data across programs and agencies;

• Build collections of data in secure facilities with the proper controls to ensure that only those individuals with the proper permission have access to data in a quick, manageable fashion;

• Develop and hire agency leadership that understands the need for evaluation and research;

• Train state and local government staff in the use of administrative data for program management and evaluation; and