Guidebook for Implementing a Study on the Dynamics of Child Care Subsidy Use

Deana Grobe
Roberta B. Weber
Elizabeth E. Davis

Oregon State University · Family Policy Program,
Oregon Child Care Research Partnership

March 2003

Guidebook for Implementing a Study on the Dynamics of Child Care Subsidy Use

This publication was funded partially by Oregon State University (OSU) Family Policy Program, Oregon Child Care Research Partnership, and the federal Child Care Bureau of the U.S. Department of Health and Human Services' Administration for Children and Families, Administration on Children, Youth and Families.

Principal authors of this document are Deana Grobe and Roberta B. Weber, OSU Family Policy Program; and Elizabeth E. Davis, University of Minnesota.

Acknowledgments

The Dynamics Study Guidebook has benefited greatly from the thoughtful comments of the Dynamics Study advisors who reviewed the guide at various stages in its development. The following individuals' assistance is gratefully acknowledged: Deanna Schexnayder and Daniel Schroeder, University of Texas at Austin; Laura Peck, Arizona State University; and Annie Georges, National Center for Children in Poverty. The authors would also like to acknowledge that the original Child Care Dynamics Study was carried out as a project of the Child Care Bureau's Child Care Policy Research Consortium.

Contents

Introduction

Why study the dynamics of child care subsidy use?

To make child care more affordable, states provide financial subsidies for low-income working parents and for those trying to transition from welfare to employment. Passage of the 1996 federal welfare reform law has resulted in a sizeable increase in the amount of funding and the number of children receiving child care subsidies. Child care subsidies are a key strategy in helping families find alternatives to relying on welfare payments.

For any government-funded program, information about participants in the program, services they receive, and how the program operates is crucial for policymakers. Although some states can use administrative data to obtain basic information about the population using child care subsidies at a point in time, they have been unable to extend these analyses to probe the characteristics of families or their patterns of child care use. In particular, information on the "dynamics" of child care subsidy use can illuminate families' complex interplay of employment, child care arrangements, and welfare program participation. The "dynamics" of child care subsidy use includes time-related elements of participation such as measures of the number of months a family receives a child care subsidy, and whether they "cycle" in and out of the program or participate in an unbroken stream of consecutive months. Studying the dynamics of child care subsidy use enhances our understanding of subsidy program participants and outcomes for families and children.

Information about patterns of child care subsidy use can assist both federal and state policymakers in designing and developing child care subsidy programs. For example, better understanding of the impact of policy interactions on the people served will allow states to decide which families they want to target for the child care subsidy program. Information on participation dynamics may also guide development of policies to impact the continuity and stability of child care arrangements.

To date, limited research has been conducted on the characteristics of participants in the child care subsidy program, the services they receive, and the dynamics of subsidy program participation. By using this guidebook to replicate one five-state study of the dynamics of child care subsidy use, policymakers and researchers in other states can improve knowledge of subsidy program participants and outcomes and conduct cross-state comparisons.

Terminology

Throughout this report, the following distinctions are made for the terms "spell," "spell-length" (duration), "reentry," and "dynamics":

> A **spell** of subsidy receipt is defined as the number of consecutive months a family received a subsidy without a "break"; that is, a month (or more) in which they did not receive a subsidy.

Spell-length measures the length of a spell of participation, which is the number of months a family receives a subsidy for child care without interruptions (continuous subsidy receipt).

Reentry is when a child who had ended a spell begins another spell of subsidy.

Dynamics refers to various time-related elements of participation in the child care subsidy program (for example, spell-length, reentry, continuity, cumulative months of participation).

Purpose of this guidebook

The purpose of this guidebook is to enable states and researchers to conduct their own studies on the dynamics of child care subsidy use. To accomplish this goal, this guidebook describes the methodology developed through a five-state study on this topic.

Although there are various ways one could approach a study on child care subsidy use, replicating the same methodology will enable states to compare their findings with those of other states. These comparisons increase the usefulness of the states' findings. Further, by sharing states' findings, we will begin to create a national picture of the dynamics of child care subsidy use.

Background – Study on the dynamics of child care subsidy use

The content of this guidebook is based on a five-state research study on the dynamics of child care subsidy use (referred to henceforth as the Dynamics Study[1]) (Meyers, et al., 2002). The five states participating in the study were Illinois, Maryland, Massachusetts, Oregon, and Texas. The project built upon relationships established by the Child Care Policy Research Consortium between university researchers, state child-care staff, and community child care practitioners. This consortium is a national alliance of research projects sponsored by the Child Care Bureau in the Administration on Children, Youth and Families, U.S. Department of Health and Human Services. The consortium helps the Child Care Bureau increase the national capacity for sound child care research, identify and respond to critical issues, and link child care research with policy and practice.

The main objectives of the Dynamics Study were to provide information on the dynamics of child care subsidy use; explore how subsidy spell-length varies by characteristics of the family, child, and type of child care; and assess the stability of child care arrangements purchased with subsidy dollars. In addition, the research analyzed differences in state policy related to child care subsidies that may affect dynamics of subsidy use.

Analysts constructed longitudinal datasets using administrative data from state child care subsidy systems in the five states. By creating these longitudinal datasets, the data included all (or a significant proportion) of the child care subsidy assistance that was delivered through each state's voucher program. This design allowed the team to study the:

- characteristics of children and families who received subsidies
- services they received

[1] The Dynamics Study was also known as the "Duration Study."

- length of time they received services
- frequency with which families returned to the subsidy system
- stability of children's care arrangements while in the subsidy system.

Although the primary goal of the Dynamics Study was to understand the continuity and stability of child care subsidy receipt, the study further benefited state agencies by providing a basic understanding of their subsidy program. The study provided important information on the characteristics of families and children using subsidies and the services they were receiving.

A cautionary note

This guidebook is intended to make it easier for states and researchers to replicate the Dynamics Study. It is critical to remember, however, that each state's data will contain nuances that will require time and effort to understand. Each state's data system is different, and unforeseen problems and issues may arise. It is important to recognize at the outset that any dynamics study will require substantial personnel time to form partnerships, acquire the data, become familiar with and document the data, clean and format the data, conduct the data analysis, and produce reports. It is difficult to provide a cost estimate of replicating this study because many of the costs are hidden in the operational budgets of organizations. In our experience, cleaning and reorganizing the administrative data constitutes the bulk of the time and costs.

Who should use this guidebook?

The guidebook will be of value to anyone interested in conducting a study on the dynamics of child care subsidy use. Ideally, research partnerships would be formed that join the resources and expertise of researchers from universities and private research organizations, child care practitioners, and state (or county) child care agency administrators and staff. (For more information on *Forming a Research Partnership,* see *Section I: Getting Started*.) These types of partnerships are an effective strategy for conducting policy-relevant research.

What's inside

The guidebook is divided into five sections, each describing a stage of the research project. Although the sections are presented in a linear fashion, the actual application of the research project is more circular, requiring many feedback loops.

Section I: Getting Started

Section II: Selecting Data

Section III: Assessing Data Quality and Documenting Data

Section IV: Creating an Analysis Data Set

Section V: Analyzing Data and Presenting Findings

Section I begins with ideas on forming a research partnership, identifies the key research questions addressed throughout the guidebook, and includes a discussion on research design.

Section II discusses the strengths of the Dynamics Study by focusing on the use of administrative data and the creation of longitudinal data files.

Section III provides specific information on how to assess the quality of the data and document the source data. It also includes a checklist for cleaning the data.

Section IV details the creation of a longitudinal data analysis file. Discrete steps are presented, with each step building on the previous steps.

Section V details the elements of the data analysis and presentation of findings.

The **glossary** includes definitions of key concepts referenced throughout the guidebook.

Appendix A provides further discussion on why a longitudinal design was chosen to study the dynamics of subsidy use instead of a cross-sectional design.

Appendix B provides examples of data sharing and confidentiality agreements. These documents may be used as templates for creating agreements between research institutions and state agencies.

Appendix C includes the Oregon Dynamics Study codebook that illustrates one approach to documenting source data.

Appendix D provides the statistical code used in the Oregon Dynamics Study that is referenced throughout the text. This code is written for use with the SAS software program and is provided for reference purposes. For more information on using SAS software program see Allison (1995) and Spector (1993).

Section I: Getting started

Forming a research partnership

No one model can define the creation of partnerships. Partnerships grow out of their own ground. The relevance, strength, and credibility of child care research partnerships arise from the collaboration of many participants – administrators, service practitioners, parents, and researchers. An effective partnership joins different perspectives and skills, yet strives for shared understanding and common purpose.

The core value that guides a research partnership is respect for the unique contributions, views, and vital interests of all parties. In the initial stage, the partners should come together and reach consensus on the needs of each partner, the role of the funders, research questions, data sources, and analysis and dissemination plans.

The following are important issues to consider when developing and maintaining partnerships.

- How will the group move beyond self-interest to common understandings and a shared vision for future research?
- How will the partnership facilitate communication among all partners from inception to dissemination and follow-up?
- What will the products be?
- How will credit be given for authorship, work, and/or participation in the partnership?
- Who has authority to speak or sign off on behalf of participating organizations?
- Which decisions will be made by the group, and which by the experts in different areas?

Although the creation of a research partnership is demanding, the benefits of this hard work are many (see box). In addition to providing quality, credible, and timely policy-focused research, research partnerships can expand the capacity for, and quality of, research.

Policy-makers and practitioners bring insight regarding the data, ensuring that the right questions are asked and the right data sets used.

Researchers can assess the sufficiency of datasets relative to proposed research questions and can propose different questions if needed. The objectivity of university researchers adds to the credibility of the findings with policy-makers.

Benefits of Research Partnerships:

- Policy is informed by research
- Research is respected and relevant
- Stakeholders are engaged
- Partners learn to speak a similar language
- Cooperation, coalition, and relationships are built
- Problems are approached holistically by incorporating multiple perspectives
- Opportunities are provided to leverage resources
- Use of existing data resources is improved
- Data infrastructure is improved

The Child Care Bureau is fostering these kinds of research partnerships. The Bureau supports nine research partnerships through its Child Care Policy Research Consortium. Common characteristics of all partnerships include partners' joint determination of the research agenda and use of primary and secondary data sources. These partnerships have contributed information on the effects of state and local policies on (1) parents' choice of care, (2) the supply of care, (3) the duration of subsidy utilization and child care arrangements, and (4) the relationship between availability of subsidized care and entry into the job market. Weber and Wolfe's (2002) paper includes a more comprehensive discussion of child care research partnerships.

Research questions

A successful research study begins with clarity on key research questions. The chart below shows the research questions addressed by the Dynamics Study. The remainder of this guidebook describes the processes used to frame these research questions and discusses the structure and definition of key variables. If other state partnerships seek to address similar questions, it will be important for states to compare datasets and variables available in that state with those used in the five-state study.

Research Questions
What are the characteristics of children and families who receive subsidies? What services do these children and families receive? What is the average ***length of spell*** of subsidy receipt; that is, how long do spells of subsidy receipt last? What is the probability of ***reentry*** into the subsidy system; that is, how likely is it that children who end a spell of subsidy receipt subsequently begin another? How ***stable*** are children's care arrangements while they are in the subsidy system?

Research design

Policy-makers are keenly interested in the length of time that families participate in a government program; in this case, for example, for how many months does the family receive a subsidy to pay for a child's care? A "spell" of subsidy receipt is defined as the number of consecutive months that a family received a subsidy without a "break," that is, a month (or more) in which they did not receive a subsidy.

Measuring the length of a spell of participation can be tricky: families may participate for a period of months, leave, and then return to the program. Some families will be receiving subsidies before the start of a study and some will continue receiving subsidies after the study ends.

Understanding the different ways to measure the spell-length of participation is crucial to understanding and interpreting the results of any dynamics study. "Hazard analysis" is the general name given to statistical methods used to measure and study events over time. Others might know it as "survival" or "event history" analysis.

Distinguishing time

There are two ways to distinguish time in research design: cross-sectional and longitudinal. The choice of design depends on the research questions of interest. A longitudinal design approach was used in the

Dynamics Study because it allowed the team to study directly the research questions on dynamics of subsidy use (see *Appendix A* for further discussion on why a longitudinal design was chosen):

- the length of time children and families received services,
- reentry of subsidy use, and
- the stability of children's care arrangements while in the subsidy system.

A longitudinal study is one that includes repeated measures or waves of measurement of the same variables for the same group of people. Data may be collected daily, weekly, monthly, quarterly, or yearly. Longitudinal datasets (sometimes referred to as panel data) typically have an observation for each family for each time unit (typically a month) of subsidy use.

Measuring spell-length with a longitudinal sample: How long does an average family receive subsidy assistance?

Suppose we are interested in the average length of time a family participates in the subsidy program without a break (that is, the average length of a continuous spell of participation). We can measure this using a longitudinal sample. Let's illustrate with a simple example.

Suppose we have a center care provider with space for two children at a time (2 slots), and we monitor who is using each slot for 12 months (see diagram below). The longitudinal sample includes a sample of 13 children and includes everyone who was in the two center care slots for any length of time during the 12 months. One of the 13 children has a spell-length of 12 months (X_1), and the other 12 children have spell-lengths of one month each (X_2-X_{13}). Based on the longitudinal sample, the average spell-length is ((12 months*1 person)+(1 month*12 persons))/13 = 1.85 months.

Longitudinal sample

Months	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Slot 1	X_1 ———→											
Slot 2	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}

If we took a longitudinal sample of all 12 months, we would have a sample of 13 children, one with a final spell-length of 12 months (X_1) and 12 with a final spell-length of 1 month (X_2-X_{13}). Given this sample,

average spell-length
= ((12 months*1 person)+(1 month*12 persons))/13
= **1.85 months**.

The longitudinal sample tells us the average spell-length using a sample of all those who participate during the time period (that is, percent or average of those ever served). ***Using a longitudinal sample is most appropriate if policy-makers are interested in the average spell-length experience of all families.*** Again, it is important to be clear about research goals and priorities so that the analysis can be tailored to best answer specific research questions.

Section II: Selecting Data

Data selection

The main source of data used in the Five-State Dynamics Study was administrative data from the child care subsidy program. Administrative data are collected on a regular basis in support of an agency's function and kept within the agency's database or information system. The child care subsidy program data were selected as the source data for this study because:

1. The data represent complete and cumulative information on the service population. There is a higher degree of reliability with data from the population versus data from a sample of the population.
2. The subsidy program data was usually collected in monthly files. These monthly data files can be used to construct a longitudinal research data set. A longitudinal data set is crucial to being able to analyze time effects; in this case the length of time a family receives a subsidy for child care.
3. The data are timely and data collection is continuous.
4. Administrative data are cost effective. The costs of collecting longitudinal survey data on all child care subsidy clients in a state, or even a large sample, would be prohibitive for most research projects.
5. Administrative data can be linked with other data sources to improve data quality and permit more extensive analysis (see *Further Analysis* for details).

The disadvantage of administrative data is that they are collected by systems designed to determine eligibility and process payments; they are not designed for research purposes. Because of this, an administrative dataset commonly lacks documentation on the quality of the data, is available only for periods that the client is in the program, and lacks control variables (for example, education, prior employment).[2] Despite these disadvantages, administrative data are a convenient and inexpensive source of child care subsidy client data.

The subsidy data in the Dynamics Study included almost all of the families who received child care assistance that is delivered through each state's voucher program (that is, child care that was paid for through vouchers given to families to use in choosing their preferred provider), but did not include children receiving child care that was paid for through a state's system of child care contracts (that is, child care that was paid for through a direct contract with a provider). Only the voucher data were used in the Dynamics Study. In general it is important to understand and acknowledge the coverage of any study. In this case, any conclusions from the Dynamics Study refer to the voucher systems.[3] Characteristics of children, child care providers, and spell-length of assistance might look different if one were studying care funded through contracts with child care providers.

---

[2] Many of these limitations can be overcome by linking administrative data files across various programs. See the *Further Analysis* section for more information on this topic.

[3] Not all voucher care was included in all states. In Texas, for example, voucher-based care purchased for children in child protective service cases was not included, nor was care purchased through JTPA / WIA funds.

Acquiring agency data

State agencies may be willing to share their data if researchers articulate how they want to use the data and these uses are compatible with the agencies' goals and mandates. Agencies are interested in gaining information to improve their program, and because of scarce resources they may see value in forming partnerships. After establishing an informal agreement of cooperation from a source agency, a formal agreement outlining the data sharing and confidentiality assurances should be negotiated with each participating human service agency before the researcher receives the data.

Formal agreements

Reaching an agreement will depend on the structure and rules and regulations governing the agency. Elements to consider including in a formal agreement (see *Appendix B* for examples of data sharing and confidentiality agreements) are:

- **Ensure the confidentiality of client data**. One way is to have the research team sign an agreement to protect the confidentiality of data.[4]
- **Ensure data security and control access to data.** Indicate what precautions will be made to protect the data or other information from being accessed by persons outside the project. Precautions may include: inventory confidential records when received, store data files in a locked facility or cabinet, protect files in the database with passwords, remove any individual identifying information, train staff about data security, use encryption software when transferring files electronically, use computers that are not connected to the Internet when processing data, and use a physical security system such as an alarm.
- **Share the research results with the agencies.** Agencies should be given the opportunity to review and provide input concerning the research findings before they are made available to the public. While it is important that researchers remain in control of the research findings for credibility purposes, the research is usually improved by incorporating state staff and other practitioners' interpretations of the findings.

[4] In order for a government agency to obtain and/or use a client's Social Security Number for a particular purpose, the Federal Privacy Act of 1974 (5 U.S.C. 552c) requires the government agency to establish its authority to request the client to disclose his or her Social Security Number and to disclose to the client whether that disclosure is mandatory or voluntary, by what statutory or other authority such number is solicited, and what uses will be made of it. This often can be handled when a client enrolls in a program that provides government assistance.

Section III: Assessing Data Quality and Documenting Data

Checklist for cleaning the data

After obtaining the data, it is important to check the data to make sure the values make sense. This process is generally referred to as cleaning the data. Past experience with administrative data has shown that the quality of the data varies, as some variables may be quite reliable while others are full of errors and missing data. Thus, the administrative data should be carefully reviewed, checking for outliers and missing values. It is also important to work with the source agency to improve the data quality for future use.

This process of cleaning the data should be revisited throughout the preparation of the data analysis file. The various approaches to cleaning data are indicated below.

√ **Look at a printout**
Simply print out a subset of the data. Examine the printout and look for alignment across the cases.

√ **Examine the frequencies of variables**
Another approach is to run a frequency check on all the variables in your data set. Things to look for in the frequencies:

- Is the number of responses correct for each variable? That is, if there were 14,119 families in the April data file, are there 14,119 responses to each variable?
- Are all of the responses in the correct range? That is, if a variable has a 1–5 scale, does the data reflect that range? Or if AS is a code for a variable but the frequency for that variable included codes such as A, AD, and ASE, the analyst may need to ask the source agency what they think those codes mean or make a judgment call about what is correct.
- Does the data make sense? For example, a figure of 95% of child clients under the age of 2 would be suspicious.

√ **Other things to check for:**
- Check numeric fields to determine whether, for example, a response of 2 indicates 02 or 20.
- Check to see whether a unique number (for example, 0, 9, 99) is assigned for missing data. This unique number should not be a valid response in any variable.
- Check the data system rules for implied decimal places. Sometimes 1000 may mean 10.00 or 100.0.
- Check how dates are recorded in the system.
- Check for consistency across fields. For example, in one data system, each individual in the household was given a number (1, 2, 3, …, 20) and 1's and 2's were supposed to indicate which individuals were the parents. However, when checking birth dates, we discovered that a sizeable percentage of those coded as "2" (supposedly a parent) were under 10 years old. This miscoding would matter if we tried to use this variable to identify children versus adults or to calculate the percentage of one-parent versus two-parent families.

Assessing the data quality

Assessing the quality of data is an ongoing process that includes detecting unreliable data, identifying systematic errors, or correcting inaccuracies. Because source agencies seldom document information on the quality of the data, it is up to the researchers to interview those who maintain, use, or train others to use the database. Questions to inquire about include (Goerge, et al., 1994):

- What motivation does the source agency have for collecting the data?
  The data are more likely to have a high degree of accuracy if the source agency heavily relies upon the information generated from the data. Similarly, data that are used to satisfy state and federal compliance laws tend to be more complete and reliable.
- Is there a system in place for checking the accuracy of data?
- How often are errors corrected? What are the processes for updating changes to data?
- What are the rules for purging records from the system? *This may not be obvious just from examining the data, as some data systems might remove records after a client has been inactive for a certain period of time.*
- Are data entered directly by the agency staff? Was the staff trained to use the database?
  We would expect data to be more reliable and accurate if workers receive training on data entry.
- Do forced range of allowable entries exist in the information system? If so, what are they? Are there fields that do not require an entry for the system to approve a client? *If so, these will probably have a higher degree of missing values.*
- What analyses have been conducted with these data in the past?
- Are the data fields missing critical to the study? Are there other ways of obtaining these data?

In some cases, "learning about the reliability, validity, and accuracy of data may be possible only after the data have been analyzed," such as when comparing results across states (Goerge, 1997). For example, only after comparing family income across the five states in the Dynamics Study did we realize that "income" was recorded differently in the data systems of the five states. Several states included cash (TANF) grants, some included only earned income, and some included income from other household members (other than the child's parent).

Understanding the history of the administrative data, including the purpose(s) for which data are collected and any changes in the method of collection or definition of variables over time, can help to prevent misinterpretations. With changes in personnel it is not always easy to track down changes in data meanings that have occurred in the past but may be included in your data. Other things to learn about the data are:

- Policy changes over time that may influence the meaning of the values collected.
- Whether changes/corrections to the data system are recorded historically or over-written (particularly important to know if you are getting a file a year or two after the time period it represents).
- Whether a time lag exists after the end of a given time period before the data files contain complete information (for example, billing data may take several months to be added to the system).

In addition, it is important to understand the process used by the agency for (a) correcting errors in the data, and (b) updating data with new values. For example, certain data files may be corrected or updated only at redetermination, which may be as infrequently as every 6 or 12 months. Errors may be corrected in the current month's database, when discovered, but may not be corrected in prior months. As a result, one might interpret the data as showing a change when in fact it is only correcting an old error.

Understanding when and for which variables corrections are made can also help to prevent misinterpretation. A good example is employment changes that may occur between redeterminations. If they are recorded only at redetermination (or on the off chance that the client calls in and the case worker makes the change on the system), one needs to be careful not to confuse when the data was corrected with the actual timing of such changes.

Documenting source data

An important step at this point in the process is to document the original data's content and structure. This document is referred to as a codebook and is designed to serve two purposes: (a) to document how each data item was defined by the source agency and where this variable is located, and (b) to serve as a reference when you begin your analysis of the data. An example of a codebook created for the Oregon Dynamics Study is found in *Appendix C*.

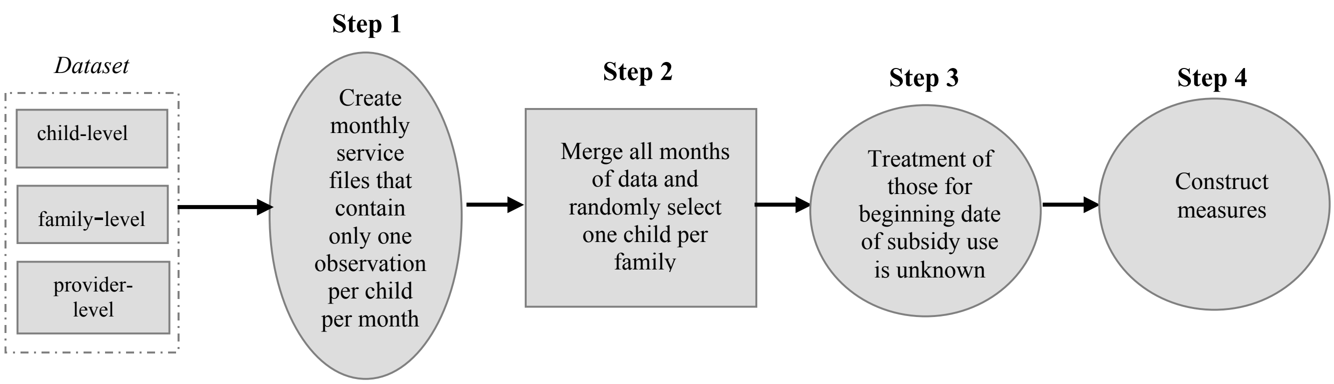
Documentation could include variable names, variable definitions, value codes, and comments explaining either agency protocol or redefinition of variables. For example, if the agency adds new values to a variable such as provider type, it would be important to document the new values as well as the effective date for this new coding. In addition, the original data layouts used by the source agency should be kept on file. This process of documenting the data should continue throughout the study as new data are added, changes are made by the source agency, and new variables are created.

Section IV: Creating an Analysis Data Set

For the Dynamics Study, raw data from the states' subsidy payment files were obtained for 24 consecutive calendar months in 1997–1999. This length of time was not necessarily the optimum number of months for conducting a study of this type; in fact, more would be better. Twenty-four months was selected because all five states could provide data for that length of time. If a longer time frame is chosen, it will be important to document policy changes throughout the time period.

The analysis data set was structured in a way such that additional months of data could be added to lengthen the study period throughout time. Figure 1 illustrates the general steps involved in creating the data analysis data set to study the dynamics of child care subsidy use, and the steps are described in more detail below.

Figure 1



Three levels of data are needed for this study: child-level, family-level, and provider-level. Each level (child, family, provider) requires a unique identifier to permit tracking of changes over time. For example, to analyze the number of providers or changes in provider, each provider must have a unique identifier. Examples of variables that are important at each level are provided in the table below.

Data Level	Description	Examples of variables
Child-level	Information on the children participating in the child care assistance program	Age or birthday, race/ethnicity, sex, special needs (if identified)
Family-level	Information on the families of children participating in the child care assistance program ***Note:*** Many of the family level variables need to be created	*Characteristics of parents:* employment status, age, sex *Characteristics of the family:* TANF receipt, amount of copay (for all children), amount paid to all providers on subsidy (for all children), household income, number of children, number of children in subsidy
Provider-level	Information on providers who provide service to a specific child who is in the child care assistance program	Type of provider, licensed/regulated

Step 1: **Create monthly service files that contain only one observation per child per month**

The objective of this step is to create monthly service files (that is, the month the family received services, rather than the month in which the provider was paid, if different) that contain only one observation per child per month. These service files should contain subsidy payment data and information about the child, family, and the providers. Because each state organizes its subsidy payment data differently, this guidebook provides a checklist of tasks for two general cases: (1) when all the subsidy data are provided in one file each month, and (2) when subsidy data are provided in multiple files (child-level, family-level, provider-level) each month.

An additional objective of this stage is to retain information at the monthly level. For example, some variables such as number of providers, family copayment, and family subsidy value need to be created each month because their values may change from month to month.

Note: Remember to document the specific procedures used to create the data analysis files in your state for future reference.

Case 1: Subsidy data are available in one file each month

In this case, the subsidy payment files are received monthly and contain child-level data along with information on families and providers.

Checklist of tasks:

√ Drop any lines of data that are exact duplicates for all variables. One should, however, always document the nature and number of dropped records and attempt to discover the source of duplication, especially if the number dropped is substantial relative to the number kept.

- √ A unique identifier is needed for each child; that is, a number that refers only to that child and does not change. Some states have child-specific case numbers that are unique. If no unique child identifier exists or if it is missing for some children, one should be created. Other ways to match without a unique identifier are to use the child's social security number or a combination of identifiers such as name and date of birth. (Using date of birth by itself is usually not sufficient because of the possibility of twins.) It may be important to determine from the state agency why some children are missing unique identifiers.
- √ Count the total number of providers for each child in the month and keep this variable. One way to do this is to sort by provider with the highest payment and hours. Count the number of providers by deleting lines with the same child ID and same provider ID. Count the number of providers for each child in the month. (Many children will have only one provider in a month.)
- √ In cases where a child has more than one provider in a month, data on each provider may be retained. To create a data file with only one observation per child per month, separate variables are created to refer to the "first" provider and "second" provider. Divide the sample into first providers and second providers (that is, sort by provider with the highest payment and hours). Use the data set with second providers to create the second provider information variables such as provider ID, type of care, relative care, claim payment, and copayment. ***Note:*** In a few cases, children may have three or more providers. Keeping information on only two providers was a compromise between retaining more information and keeping the size of the file manageable.
- √ Count the number of children and/or number of children in subsidy (if this variable is not already available at the family level).
- √ For each family, sum up copayments for all providers, across all children in a family to create a "family co-payment" variable. This family copayment value will be the same for all children in a family. ***Note:*** This step is performed if the family pays just one copayment amount, but this amount is distributed between the children in the family and in some cases between various providers. Thus, family copayment would reflect copays paid for every child to all providers in that month.
- √ For each family, sum up subsidy amounts paid to all providers, across all children in a family to create a "family subsidy" variable. This family subsidy value will be the same for all children in a family. ***Note:*** Similarly to the previous task, the total value paid to providers (family subsidy value) would be the sum of all amounts paid to all providers for all children in the family for the month.

Case 2: Subsidy data are provided in multiple files each month

In this case, the subsidy payment files arrive in multiple files such as a child file, a family file, and/or a provider file each month. All three files must be merged together on a monthly basis to create a monthly service file.

Checklist of tasks:

All the tasks listed for *Case 1* are also relevant for *Case 2*. The additional tasks for *Case 2* are as follows:

- √ For each month, merge the different service files (for example, child-level, family-level, provider-level) into one service file. The files need to be merged by unique identifiers to match children with their parents and providers.

√ Check for observations without matching observations from the other data set(s). For example, do the children in the child-level data set match up with a family in the family-level data set? If not, investigate why the matching observation is missing, if possible. Sometimes there could be a child or family record but no payment records because the application was incomplete or was denied.

√ Delete cases that are missing the matching observation. Again, always document the nature and extent of dropped records. The results could be biased if those dropped are systematically different from those kept.

Reference Note: In Oregon, we receive both child- and family-level service files each month. *Appendix D* contains three SAS statistical program files Oregon used to carry out the tasks for Step 1, Case 2. These files are labeled "*D.1 Child-level*," "*D.2 Family-level*," and "*D.3 Child-Family merge*" and are included for reference purposes.

Step 2: **Merge all months of data and randomly select one child per family**

The objective in this step is to create a longitudinal service file that randomly selects one child per family. In the Dynamics Study, a random child was selected from each family to prevent over-weighting the experience of large families. The pros and cons of this approach are discussed in detail below.

The tasks to accomplishing this objective are:

√ Pull together all of the monthly service files with one line per child per month (an example of this data layout is provided below). ***Note:*** This data was sorted using unique family and child identifiers to capture all the monthly data during which the family received a subsidy.

Example: Monthly Service File Layout

Family id	Child ID	Prov ID	Service Month	Birthdate	Type of care	Race	Single parent	TANF status	Family Copay
AA2	FG5J	ZAL8	199807	19980205	FAM	W	1	0	$89
AA2	FG5J	ZAL8	199808	19980205	FAM	W	1	0	$89
AA2	FG5J	ZAL8	199809	19980205	FAM	W	1	0	$89
AA2	FG5K	DEG4	200103	20001123	FAM	W	1	0	$89
AA2	FG5K	DEG4	200104	20001123	FAM	W	1	0	$89
PY4	UK0K	LAG5	199802	19940914	FAM	B	0	0	$100
PY4	UK0K	LAG5	199803	19940914	FAM	B	0	0	$100
PY4	UK0K	LAG5	199804	19940914	FAM	B	0	0	$100
PY4	UK0K	LAG5	199805	19940914	FAM	B	0	0	$100
PY4	UK0K	EIA7	199809	19940914	CNT	B	0	0	$100
PY4	UK0K	EIA7	199810	19940914	CNT	B	0	0	$100

Family with two children with a family copay of $89.00 (same for both children). Thus, when one child is randomly chosen, the family information remains with that child.

Illustrates one line per child (from one family) per month. In 1998, this child has received child care subsidy assistance in 6 different months. Each month is represented on a different line.

Data may be organized in other ways, such as one long string of data per child for all months in the observation period. However, an advantage of the "one line per child per month" method is that it becomes easier to add additional months of data if there is a line for each child for each month. If data were organized as one long string of data per child for all months, new variables would need to be created each time a new month was added to the longitudinal data set.

√ Transform the data to a family analysis sample by randomly selecting one child from each family that received any subsidy assistance during the observation period. This sampling approach retains all families who ever received subsidies but analyzes outcomes for only one child per family.

Why include only one child per family? The reasons to include only one child per family are: (a) families with more children would weigh more heavily in the results if all children in the family were included as separate observations, and (b) the experiences of children from the same family are likely to be highly correlated, thus they are not really independent observations. For instance, the children are more likely to leave the subsidy program at the same time if their mother loses her job. Thus, results using one child per family are interpreted as the experience of all *families* served in the subsidy system, with equal representation of families regardless of the number of children who receive subsidies.

How to choose the one child per family? Many studies of child care use the youngest child, but the Dynamics Study used a random selection of one child per family to prevent distortion of the age distribution.

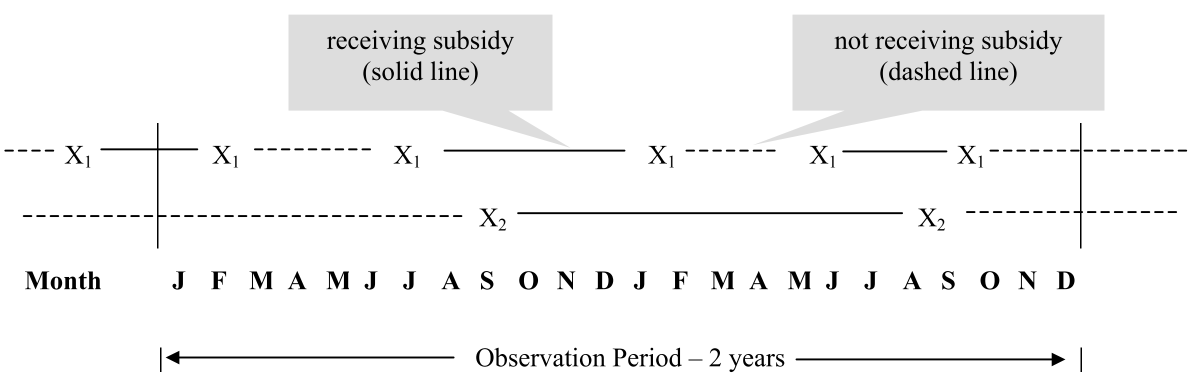
Reference Note: "*D.4 Longitudinal-RC file*" in *Appendix D* lists the SAS program code created for the Oregon data to implement the tasks in Step 2.

Step 3: **Treatment of those for whom the beginning date of subsidy use is unknown**

Once the longitudinal service file is created, further preparation is needed to create the analysis data set. The first element is to determine how to treat left-censored cases.

Left-censored cases

An observation is considered left-censored if the interval of time since the preceding event (looking to the *left* on the timeline) is unknown. If we consider the two-year observation period of the Dynamics Study, the pattern of subsidy use for child X_1 and child X_2 might look like this:



where two X's with a solid line between them denotes the occurrence of a spell (a spell is 1 or more consecutive months of subsidy receipt that are preceded and followed by 1 or more months of non-receipt), dotted lines represent a period of non-receipt, and the solid vertical lines indicate the beginning and end of the observation period. The first spell for X_1 began prior to the start of the study's observation period as shown by the leftmost X_1 that is to the left of the vertical line.

The two spells occurring during the observation period for X_1 and the observed spell for X_2 are clearly uncensored, while the first spell for X_1 is problematic. Although the first spell ends with an event (exiting from receiving a subsidy receipt), the spell is left-censored because the time when the event begins is unknown; it occurs before the observation period.

Treating the first spell as if it began at the beginning of the observation period will introduce bias because the researcher does not know when that child began receiving subsidy for that spell. A practical solution is to discard these left-censored cases. For the Dynamics Study, cases in which the randomly selected child was receiving a subsidy during the first month of the observation period (left-censored cases) were excluded from the analysis file. Although removing left-censored cases is a common solution, one should understand that this method excludes many of those with the longest spells, so the actual spell-lengths will be underestimated.[5]

How to treat later spells for those whose first observed spell was left-censored

The second issue with left-censored cases is how to treat later spells for those whose first observed spell was left-censored. The Dynamics Study team made the decision to drop all monthly observations of the child if their first observed spell was left-censored. This is because for these children the study period would be shorter, since one would have to wait for their first spell to end before observing the second. Thus, for example, all spells for Child X_1 in the example above would be dropped from the analysis. If, however, there were more years of analysis data, this issue would be less critical and the researcher could consider leaving in the later spells for those whose first observed spell was left-censored.

This exclusion of all monthly observations for those whose first observed spell was left-censored resulted in a loss of between 24 and 36 percent of observations in the states participating in the Dynamics Study. Because the exclusion of these cases could bias results by eliminating cases with the longest spell-length of receipt, a separate analysis compared the characteristics of families in the full and analysis samples. Results confirmed that, at least for characteristics that are recorded in the data system, the analysis sample was similar to the entire population.

To recap, this step further restricted the analysis data set in two ways:

(1) excluding cases receiving a subsidy during the first month of the observation period (left-censored cases), and
(2) dropping all monthly observations of the child if their first observed spell was left-censored.

[5] Other methods of estimating spell-length are available. Some procedures, for example, allow the inclusion of left-censored observations in the estimation, while others allow testing for statistical association between spell-length and other variables (see Allison, 1995).

Step 4: **Constructing measures**

The second element in creating the analysis data set is to construct measures that describe the population served and the length of subsidy receipt.

√ Construct measures of child and family characteristics. Listed are examples (see box) of some of the key variables constructed in the Dynamics Study and how they are defined. The variable definitions were determined through a compromise across the five states depending on the information available in each state. For further discussion on the measures, see Meyers, et al. (2002).

Each state will need to decide what variables they are interested in, and whether those variables are available in their administrative data sources. In addition, because of the variability of variable definitions across states, it will be important that states construct measures similar to those used in the Dynamics Study. This will enable states to compare their findings with those of other states. Two examples of variables that differed considerably among the Dynamics Study states are: (a) types of care arrangements, which have different names across states, not to mention different regulations, and (b) interpretation of the meaning of the TANF variable.

√ Construct a variable that measures the length of subsidy receipt by spells.

Measures as Defined in the Dynamics Study

Family income—includes both earned income and transfer payments (exclusive of Food Stamps) as reported in the child care subsidy data. Income measures varied across the states due to differences in data availability (for example, Texas income variable was constructed from Unemployment Insurance wage data).

Age of the randomly selected child as both a continuous and categorical variable (see *Appendix D—D.5 Age* for SAS code). Categories differed across states because of large differences in child care licensing standards by age of child and in TANF state policies regarding work exemptions relating to children's ages.

Welfare and employment status of the family—were measured as an indicator of the reasons why the target child was eligible to receive subsidies, such as earnings and/or employment or TANF receipt.

Type of care—coded as center care, family child care (non-relative), in-home care, or relative care. States differed in their categorization scheme for care arrangements as well as regulations for the different types of care.

Definition of a spell
A spell was defined in the Dynamics Study as 1 or more consecutive months of subsidy receipt that were preceded and followed by 1 or more months of non-receipt. Although an interruption of services for a longer period of time (for example, 2 to 3 months) has been used in many analyses of welfare dynamics, the shorter period of 1 month was used in this study for two reasons:

1. Monthly data corresponds to service receipt, thus even a 1-month break indicated a break in the continuity of subsidized care. Administrative data that records only when payment is made to the provider rather than the service date (that is, the month when the child received care) generally cannot be used to reliably measure length of subsidy receipt.

2. The cost of child care is usually substantial, relative to the income of families poor enough to qualify for subsidies, thus even a 1-month interruption in child care subsidies is a significant event.

An analysis using two different definitions of spell breaks (1 and 2 months) produced roughly comparable estimates of subsidy spell-length (Meyers, et al., 2002).

Measuring spell-length

The spell-length measure is intended to add up all the months of continuous receipt. What the analyst needs to do is:

First: Identify the beginning of a spell (the child received a subsidy in this month, but did not receive one in the prior month).

Second: Identify the end of the spell (the child did receive a subsidy in this month, but did not receive a subsidy in the next month).

Third: Count up the number of months from the start month to the end month to record the length of one spell. (Or subtract end month from start month minus one, if the months are numbered consecutively.)

Fourth: Create a variable that indicates whether the spell is still ongoing at the end of the observation period, or what is known as right-censored (being unobserved on the *right* side of a timeline). ***Note:*** This variable is used in the Kaplan-Meier procedure (see *Analysis 2* in the next section).

Fifth: Create an indicator variable within each observed spell that identified whether the spell was the first, second, or later spell after the beginning of the observation period. This would be useful for those interested in studying cumulative months over multiple spells.

Reference Note: "*D.6 Analysis data file*" in *Appendix D* lists the SAS program code created for the Oregon data to implement the tasks in Step 4.

Section V: Analyzing Data and Presenting Findings

This section describes how the Dynamics Study analyzed the data to answer the research questions outlined in *Section I*. In addition, this section describes what policy information, relevant to the time period of the administrative data files, was collected from each state. The following details are included as a guide, but might be altered or expanded depending on the research interests of the state.

Analysis 1—Characteristics of recipients and services received

Objective	Describe families and children who receive subsidized child care (such as the age of the child, the activity status of the parent, and family income) and the services they receive (such as the type of care, value of subsidy, and size of copayment).
Key research questions addressed	· What are the characteristics of children and families who receive subsidies? · What services do these children and families receive?
Analyses sample	Measures of child/family and service characteristics were constructed in *Section IV, step 4*, using the first month (typical when using a longitudinal sample) of the first observed spell of subsidy receipt for the randomly selected child. Analyses were conducted on those spells of subsidy receipt that started during the observation period. All data reflected the first month in which the family received services.
Analytic approach	Descriptive statistical analyses
Procedures	· Obtain the percent of subsidy recipients in the different age categories · Report the activity status of the subsidized child's parent: percent of parents' working (not receiving TANF/receiving TANF) and not working (receiving TANF/not receiving TANF) · Income - calculate recipients' median family income and percentiles of family income - compare the median incomes among subsidy recipients (at the start of the spell) to the relevant state ceiling for initial and continuing eligibility (see *Analysis 5* for information on how policy variables were collected in the Dynamics Study) · Obtain the percent of subsidy recipients by type of care arrangement · Calculate the mean and median of total provider payments and total monthly co-payments at the family level · Other variables of interest

Presentation of results

Use of appropriate types of charts is essential for effective presentation of results. This section introduces some of the graphs developed to display the above findings from the Dynamics Study. Figure 1 compares both the dollar amount of the income limit and this limit as a share of the state median income for the same year; Figure 2 displays the activity status of mothers of subsidy recipients.

Figure 1. Subsidy eligibility ceilings relative to state median monthly income (1998)

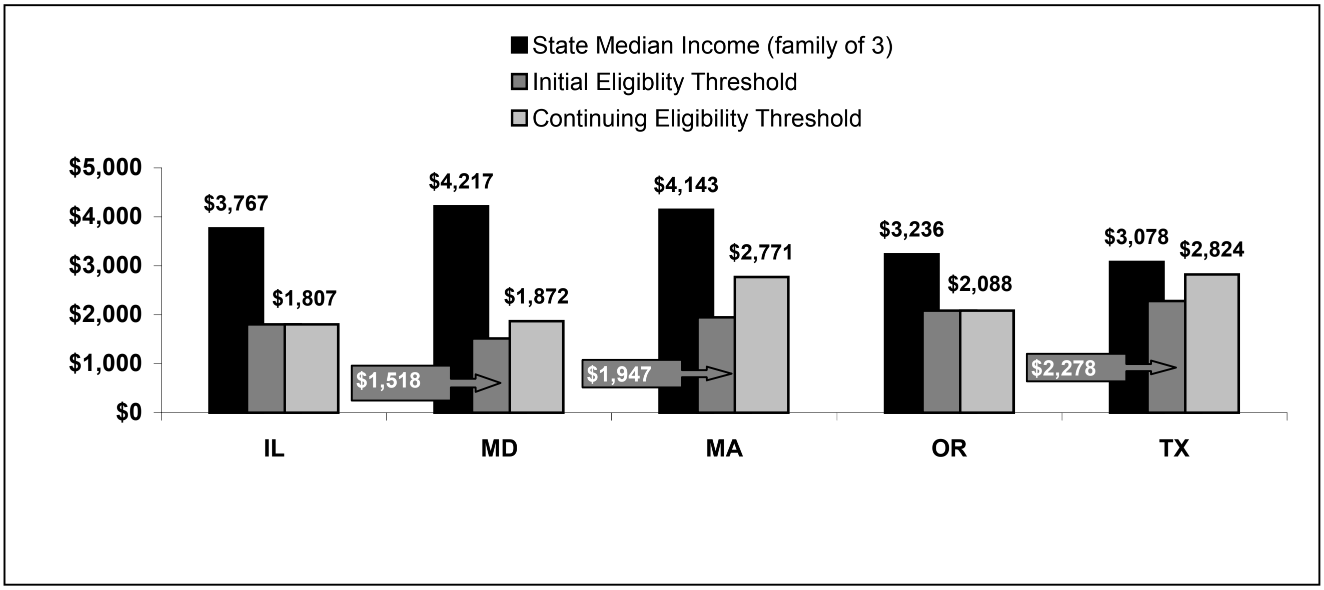
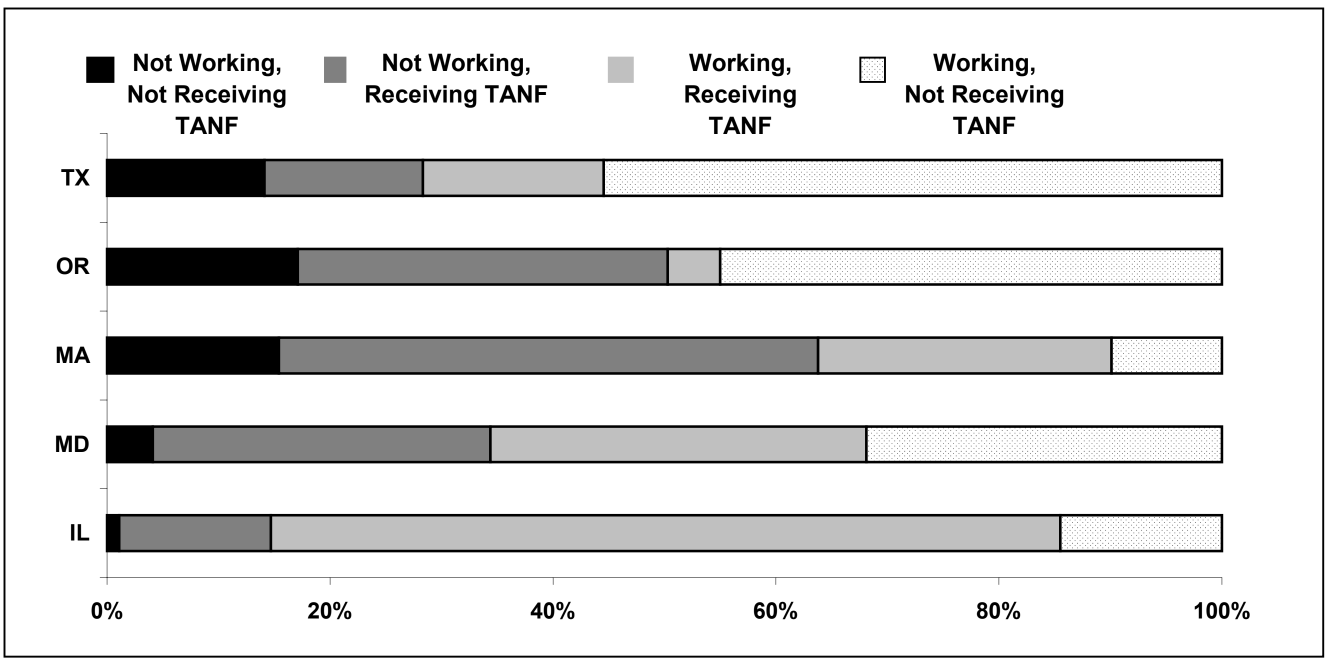


Figure 2. Activity status, mothers of subsidized children



Note: The above figures are intended for illustrative purposes only, thus the information should not be cited. Refer to Meyers, et al. (2002) for a full interpretation of the data in the figure.

Analysis 2—Spell-length of subsidy assistance

Objective	Calculate the average length of a continuous spell of subsidy receipt for the randomly selected child, as a measure of the continuity of subsidy assistance.
Key research questions addressed	What is the average spell-length of subsidy receipt; that is, how long do spells of subsidy receipt last?
Analyses sample	The first observed spell of subsidy receipt for the randomly selected child. The analysis sample was restricted to periods of subsidy receipt that began during the 24-month observation period (left censored cases were omitted).
Analytic approach	Hazard models (Allison, 1995)
Procedure	· The Kaplan-Meier[6] procedure was used to estimate spell-length in the Dynamics Study. This statistical procedure estimates the conditional survival rate at each month (the proportion of cases that continue to the observed month, given that they survived to the prior month), using spells for which data are available in the observed month (correcting for right-censoring of spells that extended beyond the observation window). · Comparison of Kaplan-Meier estimates of the median spell-length by characteristics of families and children who receive subsidized child care (including the age of the child, the activity status of the parent) and the services they receive, such as type of care.
Interpretation of this measure	Results should be interpreted as the first spell-length of families entering the subsidy system within the observation period, not the total spell-length of subsidy receipt for all families.
Reference note	"*D.7 Spell-Length*" in *Appendix D* lists the SAS program code created for the Oregon data to implement Analysis 2.

[6] Other methods of estimating spell-length are available. Some procedures, for example, allow the inclusion of left-censored observations in the estimation, while others allow testing for statistical association between spell-length and other variables (see Allison, 1995). The Kaplan-Meier procedure, however, is straight forward and widely used, so its results are easily compared to estimates produced by other researchers. This method also adjusts for the problem of right-censored observations, those whose spells continue after the end of the observation period.

Presentation of results

The next graph, Figure 3, displays the probability that families continued to receive subsidy beyond a given number of months (the "survival" rate) in the state subsidy systems. A more steeply sloped curve indicates a lower probability of continuation of subsidy or a higher probability of exit. While the survival rates suggest that the probability that children exited subsidies was very high in all states during the first few months of subsidy receipt, the median spell length ranged from 3 to 7 months in the three states shown in Table 1.

Figure 3. Survival Rate – First Observed Spell Starting During Observation Period

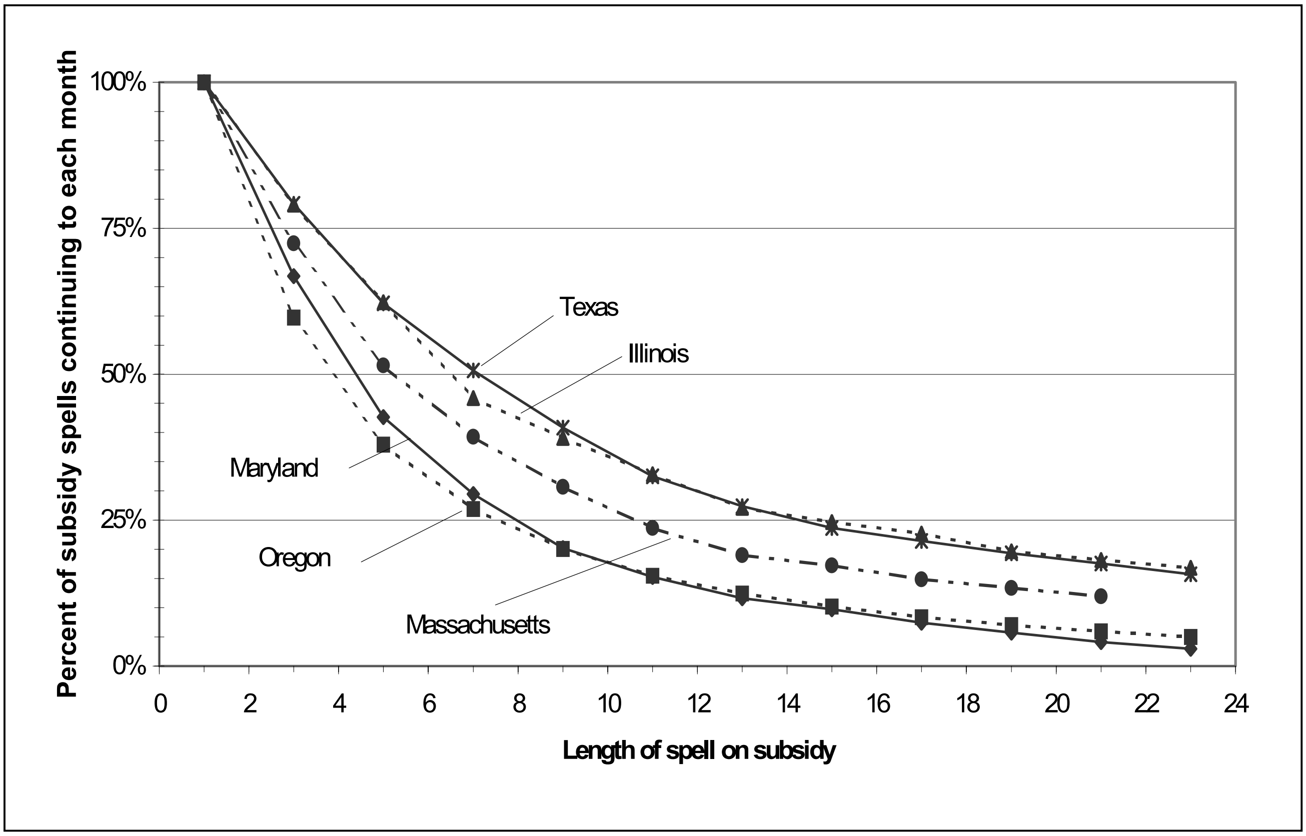


Table 1. Kaplan-Meier Estimates of Length of Subsidy Receipt in Months (First Observed Spell Starting During Observation Period), by State

	Massachusetts	Oregon	Texas
25th percentile	3	2	3
Median	5	3	7
75th percentile	11	7	14

Note: The above figure and table are intended for illustrative purposes only; the information should not be cited. Refer to Meyers, et al. (2002) for a full interpretation of the data in the figure and table.

Analysis 3—Reentry into the subsidy system

Objective	Determine the rate at which children re-enter the subsidy system after the end of a subsidy spell. Given the short observation period in the Dynamics Study, it was difficult to accurately estimate children's total time in the subsidy system.
Key research questions addressed	What is the probability of reentry into the subsidy system; that is, how likely is it that children who end a spell of subsidy receipt subsequently begin another?
Analyses sample	The sample was restricted to spells with a sufficient number of months of data to calculate the reentry rate (for example, to be included in the calculation of the rate of reentry within 6 months, there must be at least 6 months of data for the child following the end of the subsidy spell). The analyst can create dummy variables for reentry at any interval of interest.
Analytic approach	Descriptive statistical analysis
Procedures	▪ Determine whether the observation represents an exit from the subsidy system. ▪ Create two new variables to determine whether a sufficient number of months are observed to include the child in the calculation: - the number of months between the child's exit and the end of the observation period - the number of months between the child's exit and a subsequent return to the subsidy program ▪ Using these two new variables, create dummy variables for reentry at 3, 6, 9, and 12 months.
Reference note	"*D.8 Reentry*" in *Appendix D* lists the SAS program code created for the Oregon data to implement Analysis 3.

Presentation of Results

Table 2 reports the rate at which children re-entered the subsidy system within 3, 6, 9, and 12 months after the end of a subsidy spell. It illustrates that the return to subsidy receipt was common in all states, but the rate varied across the five states.

Table 2. Cumulative Percent of Children Exiting Subsidy Spell who Return within 3, 6, 9, and 12 months

	Maryland	Massachusetts	Oregon	Texas
Interval				
3 months	46%	28%	26%	20%
6 months	56%	36%	37%	27%
9 months	60%	38%	43%	32%
12 months	58%	40%	44%	35%

Note: This table is intended for illustrative purposes only; the information should not be cited. Refer to Meyers, et al. (2002) for the full interpretation of the data in the table.

Analysis 4—Stability of care arrangements

Objective	Examine the stability of children's care arrangements by determining the number of providers the child had during his/her time in the subsidy system and the length of time with the child's primary provider. The child's primary provider is defined as the provider who provided the most months of care while the child was in the subsidy system.[7]
Limitations	Because the data for this analysis included only months during which children received subsidies, it was impossible to measure the total length of time children spent in a single arrangement (because children may have been in the same arrangement before or after the period of subsidization). Instead, the data were used to measure the stability of providers for each child during the entire period of subsidization.
Key research questions addressed	How *stable* are children's care arrangements while they are in the subsidy system?
Analyses sample	All the monthly observations for one random child per family—not just the first spell (data set was created in *Section IV, Step 2*). Left-censored cases should also be excluded.
Analytic approach	Descriptive statistical analysis
Procedures	· Create the following variables: - a count of all months of subsidy receipt for the child—not just the first spell - a count of how many providers the child had in all months on subsidy - a count of the number of months (duration) with the primary (longest)

[7] The primary provider could be selected based on the most hours of care provided for the child if data on hours of care are available.

provider, ignoring any breaks in time. That is, count all the months with the primary provider even if there is a break in time.

- Drop all monthly observations for a child if that child is considered a left-censored case—in the Dynamics Study, the child was considered a left-censored case if he/she received a subsidy in the first observation month.
- Calculate a primary provider ratio:

$$\frac{\text{number of months with primary provider}}{\text{total number of months of subsidy receipt}}$$

Interpretation of Primary Provider Ratio

A primary provider ratio of 1 is interpreted to mean that all of the months of subsidized care for the child were spent with the primary provider. A ratio of .75 is interpreted to mean that during 75 percent of the months of subsidized care, the child was with the primary provider, and so on.

Reference note

"*D.9 Stability*" in *Appendix D* lists the SAS program code created for the Oregon data to implement Analysis 4.

Presentation of results

Table 3 shows the stability of care arrangements while children are receiving subsidies, for two of the five states that participated in the Dynamics Study. This table illustrates that while care arrangements were relatively stable for children who remained in the subsidy system for a short time, the stability of providers declined sharply in all states as the cumulative months of subsidy receipt increased.

Table 3. Stability of Provider-Child Relationships, by State

	Cumulative duration of subsidy receipt (months)	Mean primary provider ratio	Percent remaining with primary provider for entire subsidy period (PPR=1)
Illinois	3	0.96	93%
	6	0.91	83%
	9	0.84	69%
	12	0.78	60%
Texas	3	0.96	90%
	6	0.91	73%
	9	0.89	67%
	12	0.88	57%

Note: This table is intended for illustrative purposes only; the information should not be cited. Refer to Meyers, et al. (2002) for the full interpretation of the data in the table.

Analysis 5—Policy Variables

Objective	Examine state choices about TANF and child care policies to (a) understand the relationship of policies on the characteristics of families served in the state subsidy system, the services that families are provided, and the length of assistance; and (b) build a policy framework so that states can compare their policies with other states that complete the Dynamics Study.
Analytic approach	Qualitative analysis

Variables used in the Dynamics Study

Child care subsidy eligibility rules

- Income eligibility ceilings per month as % of state median income
- Income exclusions for determining eligibility (for example, children's earned income; non-related adults' income; Food Stamp benefits)
- Eligible activities (for example, employment; education; training)
- Frequency of recertification
- Service rationing – how states determine what services eligible claimants receive (for example, committed to serving all eligible families; waiting lists maintained)

TANF policy

- Maximum TANF grant for a one-parent family of three
- Earned income disregard policies for TANF eligibility
- Maximum TANF grant for a family with adult working full time at minimum-wage job
- Age of youngest child that exempts parent from TANF work requirements

Copayment policy

- Copayment rules (for example, adjusted for income, family size)
- Copayment exemptions (for example, TANF recipients, Food Stamp recipients)
- Monthly copayment amount for TANF family earning $2,000 per year, as well as other scenarios

Provider payment rates

Payment rates for a 4-year-old child in full-time care

- Maximum rate for center, family, in-home, and relative care
- For center and family child care
  - 75th percentile of market rate
  - maximum as % of 75th percentile of market rate
  - estimated proportion of market that the center/family payment rate purchased

Payment rate comparisons

- Rate for in-home care as % of rate for family child care
- Rate for relative care as % of rate for family child care
- Rate for family child care as % of rate for center care
- Are providers allowed to collect additional charges from parents?

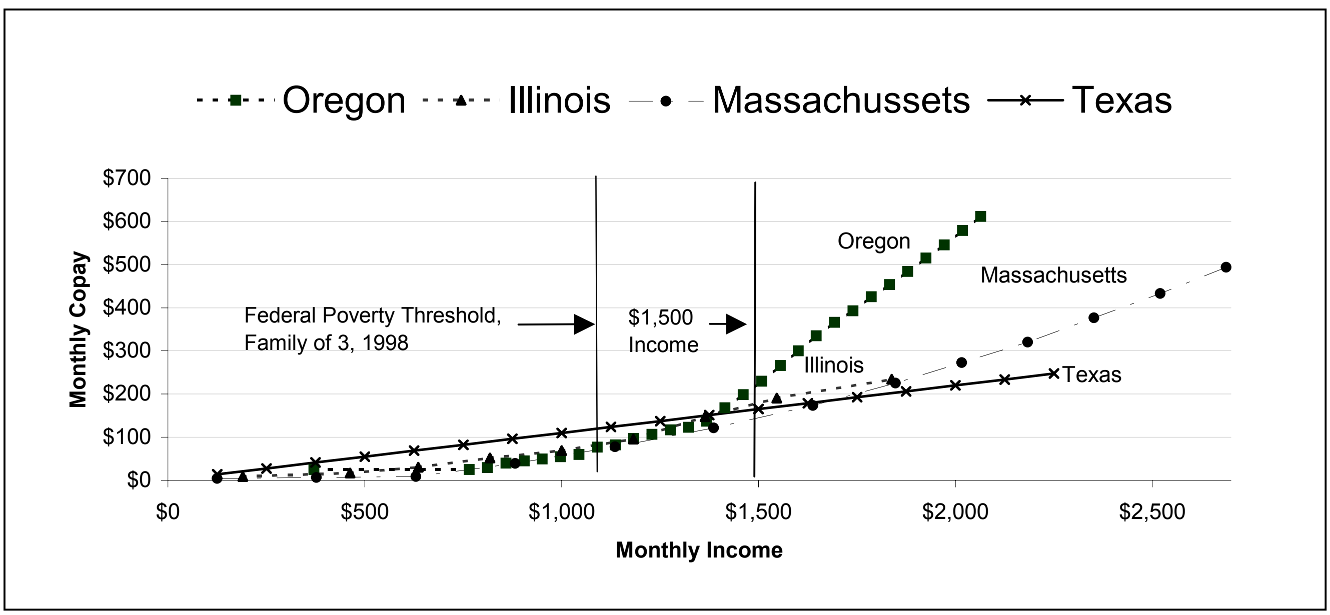
Presentation of Findings

Child care policies vary widely between states—and in some states, between counties. If one is comparing findings across geographic areas (state to state, county to state, or county to county) it is important to know how policies differ across those geographical areas. There are many purposes for collecting policy information. Qualitative data on state TANF and child care policies can be used to examine the consistency of the empirical results with state policy choices. One of the real values of comparing policy and findings is to determine whether a policy is having the desired effect. One sometimes learns that the policy effect is not as expected or desired because policies interact in a way that results in unintended effects.

Figure 4 examines copayment policy across four states. This exercise demonstrates cross-state variation in how steeply, and at what points, copayments change with income. States appear to have different priorities regarding the use of subsidy funds. For example, Oregon favors using subsidy funds for extremely poor families through its low copay for those families, compared to steeper copays for families with higher incomes. The reverse is true in Texas, whose copay is a straight percentage of earnings, thus imposing a higher burden on families with the very lowest incomes relative to those with somewhat higher incomes.

When findings on household incomes of families served (see Figure 1 on p. 24) are compared across the four states, the effect of policies on who is and is not served can be observed and evaluated. Observing state-level variation that may result from child care, TANF, and other public policies takes on more meaning when you compare differences in both policy and findings.

Figure 4. Copayment rules relative to monthly family income (three-person family with one adult and two children)



Note: This graph is intended for illustrative purposes only; the information should not be cited. Refer to Meyers, et al. (2002) for a full interpretation of the data in the graph.

Further Analysis

The type of data available to a researcher influences the scope of the study. One of the principal limitations of the Dynamics Study was that the administrative data used (typically, only data from the child care subsidy system) did not permit more extensive analysis of the factors that explain variation in subsidy spell-length among families. That is, the study team could not determine whether a spell of subsidy receipt ended for a positive reason such as an increase in earnings, or for more problematic reasons such as the loss of a job, the loss of a child care provider, or difficulties with the recertification process. To answer these questions more completely, data are needed on variables likely to influence participation in the subsidy system (for example, education or prior employment), as well as data on changes in family circumstances during the months when they are not receiving subsidy assistance.

To address these unanswered questions, studies are being conducted that link other administrative data with child care subsidy data to conduct more in-depth analyses. For example, a more extensive analysis is being conducted in three states – Illinois, Maryland, Massachusetts –using linked-level administrative data on TANF receipt, use of child care subsidies, wage reports, and child care providers. Chapin Hall's Center for Children at the University of Chicago is leading this project. They are examining the patterns of child care subsidy use and the effects of that use on welfare and employment outcomes among current and former TANF single mothers.

In Texas, subsidy data, market rate data, TANF, and Unemployment Insurance wage data files will be combined with changing policy variables over time to determine whether the devolution of subsidy policy to the local level has an effect on the duration of care, employment outcomes, and availability of care within a given geographic area.

In a similar study, Oregon is working on linking several administrative data sets to address critical questions about the relationship between use of child care subsidies, employment, and stability of child care. Understanding the interrelationships among employment, subsidy use, and child care arrangements is crucial to improving outcomes for children and families. The key to the study is to explore families' transitions from the subsidy program in greater depth.

In addition, a new study is now underway by Abt Associates and the National Center for Children in Poverty. This research will further explore relationships in administrative data and carry out a new parent survey to better understand issues affecting the duration of child care subsidies.

We hope that other partnerships will replicate the Dynamics Study in their county or state. We also hope they will extend the study. We are confident that others can design even more ways to increase understanding of the effects of child care subsidies on the lives of children and families. Another hope is that partnerships around the country will share their findings and methods/processes as part of a larger effort to build a more comprehensive picture of the dynamics of child care subsidy use. Such sharing will allow for more geographic comparisons and opportunities to understand changes over time. We request that you send findings of similar studies to:

Patricia L. Divine, Research Coordinator
Child Care Bureau
Administration on Children, Youth and Families
330 C Street SW, Washington, DC 20447

Email: pdivine@acf.dhhs.gov
Phone: 202.690.6705

Glossary

Administrative data – data that are collected on a regular basis in support of an agency's function and kept within the agency's database or information system.

Child care contracts – child care that is paid for through a direct contract with a provider.

Child care vouchers – child care that is paid for through vouchers given to families to use in choosing their preferred provider.

Child care research partnership – stakeholders such as administrators, service practitioners, parents, and university-based researchers joining together to conduct research related to child care policy at both the state and national level.

Codebook – any information on the structure, contents, and layout of a data file. Codebooks vary widely in quality and amount of information included.

Cross-sectional study – measurement is done at one point in time. Data are gathered for a particular point in time, taking a slice or cross-section of a population.

The Dynamics Study – Five-state research study on the dynamics of child care subsidy use. (Meyers, et al., 2002).

Dynamics – includes different elements associated with participation in the child care subsidy program (for example, subsidy receipt, provider stability, spell-length, reentry).

Duration – see spell-length.

Hazard analysis (hazard models) – models used for studying the occurrence and timing of events.

Kaplan-Meier – method for estimating spell-length. This statistical procedure estimates the conditional survival rate at each month (the proportion of cases that continue to the observed month, given that they survived to the prior month), using spells for which data are available in the observed month (correcting for right-censoring of spells that extended beyond the observation window).

Left-censored – spells that begin prior to (and continue into) the observation period.

Longitudinal study – includes repeated measures or waves of measurement of the same variables for the same group of people.

Primary provider ratio – number of months with primary provider divided by the total number of months of subsidy receipt.

Random selection – each element of the population has an equal chance of being selected.

Right-censored – spells that extended beyond end of the study observation period.

Service month – the month the family received services (rather than the month in which the provider was paid, if different).

Service file – data file containing subsidy payment data and information about the child, family, and the providers for the month(s) the family received services.

Spells – 1 or more consecutive months of subsidy receipt that were preceded and followed by 1 or more months of non-receipt.

Spell-length – length of time of participation. In the Dynamics Study, spell-length was defined as the number of months that a family receives a child care subsidy without interruption.

Stability – refers to the children's child care arrangements. It was measured by determining the number of providers the child had during his/her time in the subsidy system and the length of time with the child's primary provider.

Subsidy use – families receiving childcare subsidy assistance.

References

Allison, P.D. (1995). *Survival Analysis Using the SAS® System: A Practical Guide*. Cary, NC: The SAS Institute, Inc.

Goerge, R.M. (1997). Potential and Problems in Developing Indicators on Child Well-Being from Administrative Data. In *Indicators of Child Well-Being* by the Russell Sage Foundation.

Goerge, R.M., Voorhis, J.V., and Lee, B.J. (1994). Illinois's Longitudinal and Relational Child and Family Research Database. *Social Science Computer Review*, 12(3): 351–365.

Meyers, M., Peck, L.R., Davis, E.E., Collins, A., Kreader, J.L., Georges, A., Weber, R., Schexnayder, D.T., Schroeder, D.G., and Olson, J.A. (2002, July). *The Dynamics of Child Care Subsidy Use: A Collaborative Study of Five States*. Child Care Bureau, Administration on Children, Youth and Families, Administration for Children and Families, Department of Health and Human Services.

Spector, P.E. (1993). *SAS Programming for Researchers and Social Scientists*. Newbury Park: Sage Publications.

Weber, R., and Wolfe, J. (2002, summer). *Creating Research that Informs State Child Care Policy: Building and Maintaining Child Care Research Partnerships*. Albany, OR: Oregon Child Care Research Partnership, Linn-Benton Community College.

Appendices

There are two ways to distinguish time in research design: cross-sectional and longitudinal. The choice of design depends on the research questions of interest. A longitudinal design approach was used in the Dynamics Study because it allowed the team to study directly the research questions on dynamics of subsidy use. The discussion below illustrates why a longitudinal design was chosen.

A **cross-sectional** study is one where measurement is done at one point in time. Data are gathered for a particular point in time—taking a slice or cross-section of a population. Each family's spell of subsidy is measured only once. Selecting only one month of subsidy service data would be considered a cross-sectional design.

> *Pros/Cons:* Cross-sectional studies provide a snapshot of what is happening at a certain point in time and can be obtained on a relatively small budget. Cross-sectional designs, however, can only provide indirect evidence of changes over time.

A **longitudinal study** is one that includes repeated measures or waves of measurement of the same variables for the same group of people. Data may be collected daily, weekly, monthly, quarterly, or yearly. Longitudinal datasets (sometimes referred to as panel data) typically have an observation for each family for each time unit (typically a month) of subsidy use.

> *Pros/Cons:* Longitudinal studies have the capacity to look at trends across time, as process and change are actually measured. The disadvantage of this type of design is that it is resource intensive and the methodology is more difficult.

Measuring spell-length: How long does an average family receive subsidy assistance?

Suppose we are interested in the average length of time a family participates in the subsidy program without a break (that is, the average length of a continuous spell of participation). We can measure this using either a cross-sectional sample or a longitudinal sample, and in general we will get different results. Let's illustrate with a simple example.

Suppose we have a center care provider with space for two children at a time (2 slots), and we monitor who is using each slot for 12 months (see diagram below). In one of the slots, there is a child who is there for all 12 months (X_1) and leaves after the 12^{th} month. The other slot is occupied by 12 different children (X_2-X_{13}), each for one month, and the 12^{th} child (X_{13}) leaves after that month.

Cross-sectional sample

Months	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Slot 1	X_1 ————————————————————————————————◆											
Slot 2	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}

If we took a cross-section sample of month 12, we would have a sample of two children, one with a final spell-length of 12 months (X_1) and one with a final spell-length of 1 month (X_7). Given this sample,

average spell-length = (1 + 12) / 2 = **6.5 months**

In any one month, one of the slots is occupied by someone whose spell-length is 12, and someone whose spell-length is one month occupies the other slot. The average of these is (1+ 12)/2=6.5. Thus, based on the cross-sectional sample, we would conclude that the average spell-length is 6.5 months.

If, instead, we use a longitudinal sample, we would have a sample of 13 children. The sample now includes everyone who was in the two center care slots for any length of time during the 12 months (see diagram below). One of the 13 children has a spell-length of 12 months (X_1), and the other 12 children have spell-lengths of one month each (X_2-X_{13}). Based on the longitudinal sample, the average spell-length is ((12 months*1 person)+(1 month*12 persons))/13 = 1.85 months.

Longitudinal sample

Months	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Slot 1	X_1 ————————————————————————————————◆											
Slot 2	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}

If we took a longitudinal sample of all 12 months, we would have a sample of 13 children, one with a final spell-length of 12 months (X_1) and 12 with a final spell-length of 1 month (X_2-X_{13}). Given this sample,

average spell-length
= ((12 months*1 person)+(1 month*12 persons))/13
= **1.85 months**.

The longitudinal sample tells us the average spell-length using a sample of all those who participate during the time period (that is, percent or average of those ever served). ***If policy-makers are interested in the average spell-length experience of all families, the longitudinal sample is more appropriate.*** On the other hand, if one is more concerned with costs and benefits, then the families who receive subsidies for longer periods of time account for a disproportionate share of the funding, and perhaps they should carry greater weight in the analysis (shown in the cross-sectional example above). For this and other decisions it is important to be clear about research goals and priorities so that the analysis can be tailored to best answer specific research questions.

Describing families: Who is being served through the subsidy program?

The same logic can be expanded to understand why characteristics of the families served in a cross-sectional sample look different than when using a longitudinal sample. To illustrate the difference (see table below), we have used hypothetical characteristics for the children (X_1 – X_{13}) in the slot example above.

Child	Child's Age	Child's Ethnicity	Child's Gender	TANF receipt of the family
X_1	2	B	F	0
X_2	4	W	F	0
X_3	1	W	F	0
X_4	3	H	M	0
X_5	3	W	F	0
X_6	5	H	F	1
X_7	1	W	F	0
X_8	2	W	M	0
X_9	2	W	M	0
X_{10}	2	B	F	0
X_{11}	5	W	M	1
X_{12}	4	W	F	0
X_{13}	4	H	M	0

CROSS-SECTIONAL SAMPLE

The sample consists of Child X_1 and Child X_{13}. The characteristics for this sample are:

Average Age	3 yrs old
Ethnicity	50% Black 50% Hispanic
Gender	50% Female 50% Male
TANF receipt	100% no TANF

LONGITUDINAL SAMPLE

The sample consists of all the children (X_1 through X_{13}). The characteristics for this sample are:

Average Age	3.1 yrs old
Ethnicity	62% White 23% Hispanic 15% Black
Gender	62% Female 38% Male
TANF receipt	85% no TANF 15% on TANF

Although these results are exaggerated because of the small sample size, this simple example makes the point that how the analysis is conducted will affect the description of the population. That is, the decision to analyze a cross-section of the population will produce different characteristics of the children and families served than analyzing a longitudinal sample. At a given point in time, the subsidy caseload is a cross-sectional view of families who receive subsidies. For reasons explained above, the description of families served in a given month will differ from the description of those served in a longer time period (that is, one or more years).

This agreement is between the _________________ and Oregon State University regarding data sharing and the confidentiality of the data collected by the ________________ to be provided to Oregon State University (OSU). The data covered by this agreement are the computer files of the ________________ systems, which contains the service records of the clients served and services funded by _______. The articles listed below dictate the use of the data and the dissemination of information that is developed from the data.

[A paragraph can be added here to describe the purpose and scope of the research.]

1. All materials produced by investigators at OSU will contain sufficient protections to ensure client anonymity. No data or analyses will be disseminated that will describe or identify an individual client. No names, street addresses or other information that could be used to identify a particular client will be disseminated.

2. The magnetic tapes and computer files which contain the names and other identifying information will be stored in a secured, limited-access area at the Oregon State University offices, which are housed at the Family Study Center, a well-guarded building with limited access.

3. No computer files will be transferred by OSU to any other individual or organization without the expressed written permission from the Director of the __________________.

4. The data will be exclusively used for research purposes. All research materials produced from the data shall be shared at no cost with the ________________ prior to any dissemination for the purpose of reviewing the information that is developed from the data to ensure compliance with this agreement and the protection of client confidentiality.

____ __________
Signature of Center Director Date
Oregon State University

____ __________
Signature of Agency Director Date

SAS Variable Name	Reference to SAS Code	Dropped —Not Needed	Raw Data Variable Name	Description	Value or Units	Recoded	Comments
age	Nagerc3			age in current month	(years)		
agecat0 agecat1 agecat2 agecat3 agecat4 agecat5 agecat6 agecat7 agecat8 agecat9 agecat10	Nagerc3			child age dummies	1=0-11 months; 0=otherwise 1=12-23 months; 0=otherwise 1=24-35 months; 0=otherwise 1=36-47 months; 0=otherwise 1=48-59 months; 0=otherwise 1=60-71 months; 0=otherwise 1=72-83 months; 0=otherwise 1=84-95 months; 0=otherwise 1=96-107 months; 0=otherwise 1=108-119 months; 0=otherwise 1=120+ months; 0=otherwise		
agefm	newORscc			child's age at first observed subsidy month	(months)		
agemiss	newORscc			child age is missing	1=missing age; 0=otherwise		
agemm	Nagerc3			age in current month	(months)		agemm is used to create age categories (agecat0-10)
agemmfm	newORscc			child age in months at the first observed subsidy month	(months)		
asian	Child Data	X	asian	*Ethnicity:* Asian Data not included in 1999 (Oct. 1998-Sept. 1999) or 2000 (Oct. 1999-Sept. 2000)	1=Asian, otherwise blank		used ' race' instead because separate race indicators were frequently missing
black	Child Data	X	black	*Ethnicity:* Black or African American Data not included in 1999 (Oct. 1998-Sept. 1999) or 2000 (Oct. 1999-Sept. 2000)	1=African American, otherwise blank		used ' race' instead because separate race indicators were frequently missing
carecat	newORscc			care type categories which combine family and group home care	2=non-relative family and group home care 4=non-relative home care 5=center care 7=relative family, home, and group home care		
carefm	newORscc			care type (categories) at first observed subsidy month	see categories for caretype		

SAS Variable Name	Reference to SAS Code	Dropped —Not Needed	Raw Data Variable Name	Description	Value or Units	Recoded	Comments
carefm2	newORscc			care type (dummies) at first observed subsidy month	see categories for caretyp1-caretyp8		
caremiss	newORscc			care type is missing	1=missing care type; 0=otherwise		
caretyp1 caretyp2 caretyp3 caretyp4 caretyp5 caretyp6 caretyp7 caretyp8	newORscc			care type dummies	1=relative family care; 0=otherwise 1=non-relative family care; 0=otherwise 1=relative home care; 0=otherwise 1=non-relative home care; 0=otherwise 1=center care; 0=otherwise 1=non-relative group home care; 0=otherwise 1=relative family, home, & group home care; 0=otherwise 1=relative group home care; 0=otherwise		
caretype	newORscc			care type categories	1=relative family care 2=non-relative family care 3=relative home care 4=non-relative home care 5=center care 6=non-relative group home 8=relative group home		
case_num	Child Data Parent Data		case nmbr	family identifier	[e.g., AF9726]		
casenumx	newORscc			case number of the next observation—used to determine if it is the same child	see case_num		also used to determine spell length and reentry
ccbg	Parent Data		cc beg	child care subsidy program beginning date	YYYYMM [e.g., 199904]		
ccbgfm	newORscc			child care subsidy program beginning date at first observed subsidy month	see ccbg		
clm_pay	Child Data		clm pay	amount paid to provider by state	(dollars)	Divided by 100	
clmp2	CAnalysis			amount paid to 2nd provider by the state	(dollars)		
copamtfm	newORscc			family copayment at first observed subsidy month	see newfamcp		
copay	newORscc			family copayment dummy	1=copayment amount is greater than 0; 0=otherwise		

SAS Variable Name	Reference to SAS Code	Dropped —Not Needed	Raw Data Variable Name	Description	Value or Units	Recoded	Comments
copayamt	Child Data		copay amt	child care payment paid by parents to 1st provider	(dollars)	Divided by 100	
copay2	CAnalysis			child care payment paid by parents to 2nd provider	(dollars)		
copayfm	newORscc			family copayment (greater than zero) at first observed subsidy month	see copay		
copymiss	newORscc			family copayment (newfamcp) is missing	1=family copayment is missing; 0=otherwise		
count	Cfallrc3			number of subsidized children in each family	(number)	numc	
cp2inc	newORscc			copayment as a percent of household income (newfamcp/hhincom2)	(percentage)		
cumtime	ProvStab			counts all months of subsidy receipt for the child	(months)		
dob	Child Data		dob	date of birth	(YYYYMMDD)	Formatted in SAS as yymmdd8. [e.g., 19990401 for April 1, 1999]	
dur1	Duration3			duration of subsidy receipt	(months)		
durppm	ProvStab			duration with primary provider	(months)		
eeth	Child Data	X	eeth	*Ethnicity*: Hispanic Data not included in 1999 (Oct. 1998-Sept. 1999) or 2000 (Oct. 1999-Sept. 2000)	1=Hispanic, otherwise blank		used ' race' instead because separate race indicators were frequently missing
emp1fm	newORscc			employment status at first observed subsidy month	see employed		
employed	Parent Data		em	Type of income: earnings	1=case head is employed, 0=otherwise		
employx	newORscc			employment status in the previous observation within the same spell (used to create transe1 and transe2)	see employed		
exit	Re-entry			a month of subsidy receipt followed by a month of no subsidy receipt	1= exit ~ child received a subsidy, but they did not receive one in the next month, 0=otherwise		

SAS Variable Name	Reference to SAS Code	Dropped —Not Needed	Raw Data Variable Name	Description	Value or Units	Recoded	Comments
fac	Child Data		fac type	type of care at 1st provider	FAM=family; CNT=center; GRP=group, HOM=home; **NQC** =centers exempt from CCD registration; **QFM**=family providers who receive enhanced rate	caretype	**BOLD** = new coding in 2000 (Oct. 1999-Sept. 2000)
fac2	CAnalysis			type of care at 2nd provider	FAM=family; CNT=center; GRP=group, HOM=home		
fip	Parent Data		fips	01 Baker 02 Benton 03 Clackamas 04 Clatsop 05 Columbia 06 Coos 07 Crook 08 Curry 09 Deschutes 10 Douglas 11 Gilliam 12 Grant 13 Harney 14 Hood River 15 Jackson 16 Jefferson 17 Josephine 18 Klamath 19 Lake 20 Lane 21 Lincoln 22 Linn 23 Malheur 24 Marion 25 Morrow 26 Multnomah 27 Polk 28 Sherman 29 Tillamook 30 Umatilla 31 Union 32 Wallowa 33 Wasco 34 Washington 35 Wheeler 36 Yamhill	(number)		
fipfm	newORscc			county code at first observed subsidy month	see fip		
fs	Parent Data		fs	Type of income: food stamps	1=case head receiving food stamps, 0=otherwise		
fsfm	newORscc			food stamp participation at first observed subsidy month	see fs		
fsubvfm	newORscc			family subsidy value at first observed subsidy month	see newfamsv		
hawaii	Child Data	X	hawaii	*Ethnicity:* Native Hawaiian or Other Pacific Islander Data not included in 1999 (Oct. 1998-Sept. 1999) or 2000 (Oct. 1999-Sept. 2000)	1=Native Hawaiian or other Pacific Islander, otherwise blank		used ' race' instead because separate race indicators were frequently missing

SAS Variable Name	Reference to SAS Code	Dropped—Not Needed	Raw Data Variable Name	Description	Value or Units	Recoded	Comments
hhincfm	newORscc			household income (including TANF grant) at first observed subsidy month	see hhincom2		
hhincom	Parent Data		net income	household income	(dollars)	Divided by 100; added comma	
hhincom2	newORscc			household income including TANF grant if tanf status=1 and numc=1 then add $395 to hhincom numc=2 then add $460 to hhincom numc=3 then add $565 to hhincom numc=4 then add $660 to hhincom numc>=5 then add $765 to hhincom if tanf status=0 then hhincom2=hhincom	(dollars)		
home	Child Data		in home	home care	Y=care in own home; N=otherwise		
home2	CAnalysis			2nd provider home care	Y=care in own home; N=otherwise		
hr_bill	Child Data		hr bill	hours of care at 1st provider	(hours)		
hr_b2	CAnalysis			hours of care at 2nd provider	(hours)		
hsg	Parent Data		hsg	Type of income: housing assistance	1=case head receives housing assistance, 0=otherwise		
inccat0 inccat1 inccat2 inccat3 inccat4 inccat5	newORscc			Dummies for household income including TANF grant	1=0; 0=otherwise 1=$1-$500; 0=otherwise 1=$501-$1000; 0=otherwise 1=$1,001-$1,500; 0=otherwise 1=$1,501-$2,000; 0=otherwise 1=>$2,000; 0=otherwise		
ind	Child Data	X	ind	*Ethnicity:* American Indian or Alaskan Native Data not included in 1999 (Oct. 1998-Sept. 1999) or 2000 (Oct. 1999-Sept. 2000)	1=American Indian or Alaskan Native, otherwise blank		used ' race' instead because separate race indicators were frequently missing
iv	Parent Data		iv	Type of income: educational income	1=case head receives educational income, 0=otherwise		

SAS Variable Name	Reference to SAS Code	Dropped—Not Needed	Raw Data Variable Name	Description	Value or Units	Recoded	Comments
lastdata	newORscc			used to indicate end of observation period	24 (equals the last month of the 2 year sample) 36 (equals the last month of the 3 year sample)		
leftcen	newORscc			left censored cases	1=left censored (cases in October 1997); 0=otherwise		
mage	Nagerc3			variable used to construct mage1	(months)		
mage1	Nagerc3			child's age in current month	(months)	agemm	
md	newORscc			month of date of birth (dob)	(months)		
mm4ob	Re-entry			number of months observed = number of months between the person's exit and the end of the observation period	(months)		If someone exits in the next to last month of our sample period, we don't have enough months of observation to include them in the 3-month reentry rate (or any other). mm4ob must be at least 3 to be included in the reentry3 calculation, at least 6 to be included in the reentry6 calculation, etc.
mm2re	Re-entry			number of months it took to re-enter the child care subsidy program	(months)		
mojan mofeb momar moapr momay mojun mojul moaug mosep mooct monov modec	newORscc			seasonal/time controls	1=month of January; 0=otherwise 1=month of February; 0=otherwise 1=month of March; 0=otherwise 1=month of April; 0=otherwise 1=month of May; 0=otherwise 1=month of June; 0=otherwise 1=month of July; 0=otherwise 1=month of August; 0=otherwise 1=month of September; 0=otherwise 1=month of October; 0=otherwise 1=month of November; 0=otherwise 1=month of December; 0=otherwise		
newfamcp	CAnalysis			family copayment: for up to 2 providers, across all children within a family	(dollars)		
newfamsv	CAnalysis			family subsidy value: for up to 2 providers, across all children within a family	(dollars)		

SAS Variable Name	Reference to SAS Code	Dropped —Not Needed	Raw Data Variable Name	Description	Value or Units	Recoded	Comments
nocopay	newORscc			no family copayment dummy variable	1=copayment amount is equal to 0; 0=otherwise		
non	Parent Data		non	Type of income: non public assistance/other	1=case head receives some income from a source that is not public assistance; 0=otherwise		
nsv	newORscc			net subsidy value [(family subsidy value) – (family copayment)]	(dollars)		
nsv2inc	newORscc			net subsidy value as a percent of household income (nsv/hhincom2)	(percentage)		
nulprov	Child Data	X	nulprov	no description available Data not included in 1999 (Oct. 1998-Sept. 1999) or 2000 (Oct. 1999-Sept. 2000)	no definition available		
numc	Cfallrc3			number of subsidized children in each family	(number)		
numcfm	newORscc			number of children in family in subsidized care at first observed subsidy month	see numc		
numprov	ProvStab			count of how many providers the child had in all months on subsidy	(number)		
numprov	CAnalysis			number of providers for each child in that month	(number)		
parstat1 parstat2 parstat3 parstat4	newORscc			parent TANF/employment status	1=employed, no TANF; 0=otherwise 2=employed, TANF; 0=otherwise 3=not employed, TANF; 0=otherwise 4=not employed, no TANF; 0=otherwise		
paym	Child Data	X	paym	payment month Data not included in 1999 (Oct. 1998-Sept. 1999) or 2000 (Oct. 1999-Sept. 2000)	no definition available		

SAS Variable Name	Reference to SAS Code	Dropped—Not Needed	Raw Data Variable Name	Description	Value or Units	Recoded	Comments
period	newORscc			monthly periods	1=Oct 1997 2=Nov 1997 3=Dec 1997 4=Jan 1998 5=Feb 1998 6=Mar 1998 7=April 1998 8=May 1998 9=June 1998 10=July 1998 11=Aug 1998 12=Sept 1998 13=Oct 1998 14=Nov 1998 15=Dec 1998 16=Jan 1999 17=Feb 1999 18=Mar 1999 19=April 1999 20=May 1999 21=June 1999 22=July 1999 23=Aug 1999 24=Sept 1999 25=Oct 1999 26=Nov 1999 27=Dec 1999 28=Jan 2000 29=Feb 2000 30=Mar 2000 31=April 2000 32=May 2000 33=June 2000 34=July 2000 35=Aug 2000 36=Sept 2000		
periodno	newORscc			serial number indicating the month (created because the other duration study states have a data set that begins 3 months earlier than Oregon, so this variable is comparable to their period number)	period + 3		
periodsq	newORscc			period squared	see period		
periodx	newORscc			period status in the next observation (used to determine spell length and reentry)	see period		
prim_id	Child Data		prim id	child identifier	(number)		
prov	Child Data		prov nmbr	AFS provider identifier (who AFS are paying money to)	e.g., IHY00008		
prov2	Child Data	X	prov	license/regulated	# is present if they are licensed or regulated 999999 also indicates licensed or regulated;	regulate	
prov2	CAnalysis			2nd provider identification	(number)		
provmm	ProvStab			number of months each provider ID shows up for each child	(months)		
psmiss	newORscc			parent's social security number is missing	1=parent's ssn is missing; 0=otherwise		

SAS Variable Name	Reference to SAS Code	Dropped —Not Needed	Raw Data Variable Name	Description	Value or Units	Recoded	Comments
pssn	Parent Data		ssn	recoded parent's social security number	(number)		
race	Child Data		race	ethnicity	A=Asian; H=Hispanic; I=Indian; B=Black; W=White; O=Other-no equivalent; U=unknown		
race1 race2 race3 race4 race5 race6	newORscc			ethnicity dummies	1=black or African American; 0=otherwise 1=white; 0=otherwise 1=Hispanic; 0=otherwise 1=Asian; 0=otherwise 1=Native American; 0=otherwise 1=Native Hawaiian; 0=otherwise		
racefm	newORscc			ethnicity at first observed subsidy month	see race		
racemiss	newORscc			ethnicity is missing	1=ethnicity is missing; 0=otherwise		
rcumnpv	ProvStab			cumtime/numprov: ratio of months of subsidy to number of providers	(percentage)		
rdppcum	ProvStab			durppm/cumtime: ratio of months with primary provider to months of subsidy	(percentage)		
reas	Child Data		pay reas	reason for receiving subsidy	24 child care payment to JOBS Plus (employer-state work/training program) clients for activities other than care while working 94 child care for working JOBS Plus clients 94 dependent care needed to perform job search or other activities of OFSET (food stamp program employment and training) 94 child care for JOBS activities (job readiness, training or job search) 94 use for all ERDC-BAS benefits 94 use only for ERDC-SBG benefits; employment related day care: vocational training or education 94 ERDC-BAS payments charged to CCDBG for child care; employment related day care 94 payment in lieu of deduction charged to CCDBG for cc; pre-TANF work support 94 use only for payments in lieu of the deduction		

SAS Variable Name	Reference to SAS Code	Dropped —Not Needed	Raw Data Variable Name	Description	Value or Units	Recoded	Comments
reas29 reas62 reas70 reas92 reas93	newORscc			reason for receiving subsidy dummies	1= if reas code=29; 0=otherwise 1= if reas code=62; 0=otherwise 1= if reas code=70; 0=otherwise 1=if reas code=92; 0=otherwise 1=if reas code=93; 0=otherwise		In some analyses, the following reason codes were combined because of small sample sizes: • 89 and 92 • 93 and 94
reasfm	newORscc			reason for receiving subsidy at first observed subsidy month	see reas		
reasmiss	newORscc			reason code is missing	1=missing if reas=97 or reas=99; 0=otherwise		
reason1 reason2 reason3	newORscc			reason for receiving subsidy merged categories	1=reason 62, 89, 90, 92, 93, or 94; 0=otherwise 1=reason 24; 0=otherwise 1=reason 29 or 70; 0=otherwise “ . ”=reason 91, 97, 99		
reentry	Re-entry			indicates re-entry into the child care subsidy program after exit	1=re-entered the child care subsidy program; 0=otherwise		
reentry3	Re-entry			indicates re-entry into the child care subsidy program within 3 months after exit	1=re-entered the child care subsidy program within 3 months of exit; 0=otherwise		
reentry6	Re-entry			indicates re-entry into the child care subsidy program within 6 months after exit	1=re-entered the child care subsidy program within 6 months of exit; 0=otherwise		
reentry9	Re-entry			indicates re-entry into the child care subsidy program within 9 months after exit	1=re-entered the child care subsidy program within 9 months of exit; 0=otherwise		
reentry12	Re-entry			indicates re-entry into the child care subsidy program within 12 months after exit	1=re-entered the child care subsidy program within 12 months of exit; 0=otherwise		
refed	Child Data	X	refed	federal reason code for receiving subsidized child care Data not included in 1999 (Oct. 1998-Sept. 1999) or 2000 (Oct. 1999-Sept. 2000)	1= employment, including on-the-job training 2=training/education 3=both employment and training/education 4=protective services (Note: None in OR data) 5=other		
regfm	newORscc			licensed status at first observed subsidy month	see regulate		
regulate	CAnalysis			provider is licensed/regulated	1=regulated; 0=not regulated		
rel	Child Data		rel	relative care	Y=care by a relative; N=care by a non-relative		

SAS Variable Name	Reference to SAS Code	Dropped—Not Needed	Raw Data Variable Name	Description	Value or Units	Recoded	Comments
rel2	CAnalysis			2nd provider relative care	Y=care by a relative; N=care by a non-relative		
rightcen	Duration3			right censored cases	1=right censored (Sept. 2000); 0=otherwise		pertains to the first spell only
rnpvcum	ProvStab			numprov/cumtime ~ ratio of number of providers to months of subsidy	(percentage)		
sccexit	newORscc			exit from subsidy (here this month, gone next)	1=the child exited after this month, that is, we don't see them in the next month; 0=no exit; . =not currently receiving subsidized child care		
scchist	newORscc			subsidized child care history	(months)		starts at 0
scchist1	newORscc			counter for number of total months so far in the spell (scchist1 = scchist + 1)	(months)		starts at 1
sex	Child Data		sex	gender	M=male; F=female		
single	newORscc			single status	1=single; 0=otherwise		
singpar	Parent Data		sngl prnt	single parent	0= >1 adults over the age of 18 in the home; 1=otherwise From 10/97 – 2/98 the following values exist for singpar: 0, 1, 2, 3, 4, 5, 6, 7, 9. They represent the number of adults over the age of 18 in the home.		
singpfm	newORscc			single parent at first observed subsidy month	see singpar		
sm	Child Data Parent Data		ben mo	service month	YYYYMM	Formatted yymmn6. [e.g., 199710 for October 1997]	

SAS Variable Name	Reference to SAS Code	Dropped—Not Needed	Raw Data Variable Name	Description	Value or Units	Recoded	Comments
sm9710 sm 9711 sm9712 sm9801 sm9802 sm9803 sm9804 sm9805 sm9806 sm9807 sm9808 sm9809 sm9810 sm9811 sm9812 sm9901 sm9902 sm9903 sm9904 sm9905 sm9906 sm9907 sm9908 sm9909 sm9910 sm9911 sm9912 sm0001 sm0002 sm0003 sm0004 sm0005 sm0006 sm0007 sm0008 sm0009	newORscc			dummy for whether subsidy service was received in that particular month	1=’received subsidy in’ Oct 1997; 0=otherwise 1=Nov 1997; 0=otherwise 1=Dec 1997; 0=otherwise 1=Jan 1998; 0=otherwise 1=Feb 1998; 0=otherwise 1=Mar 1998; 0=otherwise 1=April 1998; 0=otherwise 1=May 1998; 0=otherwise 1=June 1998; 0=otherwise 1=July 1998; 0=otherwise 1=Aug 1998; 0=otherwise 1=Sept 1998; 0=otherwise 1=Oct 1998; 0=otherwise 1=Nov 1998; 0=otherwise 1=Dec 1998; 0=otherwise 1=Jan 1999; 0=otherwise 1=Feb 1999; 0=otherwise 1=Mar 1999; 0=otherwise 1=April 1999; 0=otherwise 1=May 1999; 0=otherwise 1=June 1999; 0=otherwise 1=July 1999; 0=otherwise 1=Aug 1999; 0=otherwise 1=Sept 1999; 0=otherwise 1=Oct 1999; 0=otherwise 1=Nov 1999; 0=otherwise 1=Dec 1999; 0=otherwise 1=Jan 2000; 0=otherwise 1=Feb 2000 0=otherwise 1=Mar 2000; 0=otherwise 1=April 2000; 0=otherwise 1=May 2000; 0=otherwise 1=June 2000; 0=otherwise 1=July 2000; 0=otherwise 1=Aug 2000; 0=otherwise 1=Sept 2000; 0=otherwise		
smx	newORscc			service month in the next observation (used to determine spell length and reentry)	YYYYMM		
spelnmbr	newORscc			indicator of observed spells ~ indicates whether it is the first, second, third, and so on observed spell for the child	(number)		
spmiss	newORscc			single parent status is missing	1=single parent status is missing; 0=otherwise		

SAS Variable Name	Reference to SAS Code	Dropped —Not Needed	Raw Data Variable Name	Description	Value or Units	Recoded	Comments
ssn	Child Data		ssn	recoded child's social security number	(number)		AFS recoded the SSN to create a unique family identifier without releasing the actual SSN
startm	newORscc			start month of first observed spell	YYYYMM		
status				signals the end of a spell of subsidy	0=exiting from SCC did not occur, or the observation is censored; 1=otherwise		
sv2inc	newORscc			percent of household income saved because of family subsidy value	(percentage)		
tanf	PAnalysis			tanf status	1=yes, 0=no, .=missing PSSN (obtained from separate TANF data file)	replaces tanf in Parent Data	
tanffm	newORscc			tanf status at first observed subsidy month	1=yes, 0=no, .=missing PSSN		
tanfx	newORscc			tanf status in the previous observation within the same spell	1=yes, 0=no, .=missing PSSN		Used to create transt1 and transt2
transe1	newORscc			transition into employment from current month to next month	1=transition into employment from current month to next month; 0=otherwise		indicates information about observed transitions, not transitions after the person leaves the child care subsidy program
transe2	newORscc			transition out of employment from current month to next month	1=transition out of employment from current month to next month; 0=otherwise		
transt1	newORscc			transition into TANF from current month to next month	1=transition into TANF from current month to next month; 0=otherwise		
transt2	newORscc			transition out of TANF from current month to next month	1=transition out of TANF from current month to next month; 0=otherwise		
type	newORscc			type of care	CENTER=center care FAM-NOTREL=non-relative family care FAM-REL=relative family care GRP-NOTREL=non-relative group home care GRP-REL=relative group home care HOME-REL=relative home care HOME-NOTREL=non-relative home care		
white	Child Data	X	white	*Ethnicity:* White Data not included in '99 (Oct. '98-Sept. '99) or '00 (Oct. '99-Sept. '00)	1=white, otherwise blank		Used ' race' instead because separate race indicators were frequently missing
yd	Nagerc3			year of date of birth (dob)	(year)		

D.1 Child-level

Objective: Read the raw child data file and create a new child data set that contains only one line per child with information on up to 2 providers.

♦ ♦

```
*Input monthly child-level raw data;
data c;
infile 'c:\my documents\Duration Study\Three Data Years\2000 data\C0900.csv' dlm=',' lrecl=1000
truncover;
input CASE_NUM $ PRIM_ID $ PROV $ SM yymmn6. PROV2 $ FAC $ REL $ HOME $
SSN $ RACE $ SEX $ DOB yymmdd8. REAS CLM_PAY HR_BILL COPAYAMT;
run;

data s0 (drop=PROV2);
set c;
CLM_PAY = CLM_PAY/100;
COPAYAMT = COPAYAMT/100;
format CLM_PAY COPAYAMT COMMA8.2;

* Get Provider licence number---PROV2
  if Provider has a license number than REGULATE=1, else REGULATE=0;
if PROV2=' ' or PROV2 = 'EXEMPT' then REGULATE=0; else REGULATE=1;
run;

* Delete duplicated lines for all variables;
proc sort data=s0 out=fixed1 NODUPLICATE;
by CASE_NUM;
run;

* Because we need to match using a unique child identifier, we must divide the sample into those
   observations that have the variable PRIM_ID and those missing it, but have SSN and/or DOB;
data c1 c2;
set fixed1;
if PRIM_ID ~=' ' then output c1; else output c2;
run;
data c3 c4;
set c2;
if SSN ~=' ' and SSN ~='X' and SSN ~='XXXXXXXXX' then output c3; else output c4;
run;
data c5 c6;
set c4;
if DOB ~=. then output c5; else output c6; run;

*Data set c1—has PRIM_ID
  Data set c2—doesn't have PRIM_ID
  Data set c3—from c2, not missing SSN
  Data set c4—from c2, missing SSN
  Data set c5—from c4, not missing DOB
```

```
  Data set c6—from c4, missing DOB;
* The following program is repeated for 3 groups of observations
     1)  those with PRIM_ID (data set c1)
     2)  those without PRIM_ID sorted by SSN and DOB (data set c3)
     3)  those without PRIM_ID sorted by CASE_NUM and DOB (data set c5)


* BY PRIM_ID;
proc sort data=c1;
by descending CLM_PAY descending HR_BILL;  *sort by provider with highest payment and hours;
run;
proc sort data=c1 out=numc1 NODUPKEY;  *delete cases with same PRIM_ID and PROV;
by PRIM_ID DOB PROV;
run;

* Count the number of providers for each child in the month and create NUMPROV;
proc means data=numc1 noprint;
var COPAYAMT;
by PRIM_ID DOB;
output out=numpv mean=cm;
run;
data nump (keep=PRIM_ID DOB NUMPROV);
set numpv;
NUMPROV = _FREQ_;
run;

* Merging data that has PRIM_ID with new variable NUMPROV;
proc sort data = nump;
by PRIM_ID DOB;
run;
data numc11;
merge numc1 nump;
by PRIM_ID DOB;
run;

* Sorting the sample, and then dividing the sample into the first providers and 2nd providers;
proc sort data=numc11;
by PRIM_ID DOB descending HR_BILL;
run;

data duprov noduprov;
set numc11;
by PRIM_ID DOB;
if first.PRIM_ID and first.DOB then output noduprov; else output duprov;
run;

data duprov2 noduprv2;
set duprov;
by PRIM_ID DOB;
if first.PRIM_ID and first.DOB then output noduprv2; else output duprov2;
run;
```

```
*Use the data set with second providers to create the second provider information variables;
data nodupv2 (drop=PROV FAC REL HOME CLM_PAY HR_BILL COPAYAMT);
set noduprv2;
PROV2 = PROV;
FAC2 = FAC;
REL2 = REL;
HOME2 = HOME;
CLMP2 = CLM_PAY;
HR_B2 = HR_BILL;
COPAY2 = COPAYAMT;
run;

proc sort data=nodupv2;
by PRIM_ID DOB;
run;

proc sort data=noduprov;
by PRIM_ID DOB;
run;

* merge the data sets back together with one child per line with up to two providers by PRIM_ID ;
data nodupvc;
merge noduprov nodupv2;
by PRIM_ID DOB;
run;

* delete cases where first provider is not the main provider. Check the log file to make sure the
   observation difference between NODUPVC and SBYPROV is small;
data sbyprov;
set nodupvc;
if HR_B2 > HR_BILL then delete;
run;

~ ~
* Same steps by SSN ;
proc sort data=c3;
by descending CLM_PAY descending HR_BILL;
run;

proc sort data=c3 out=numc2 NODUPKEY;
by SSN DOB PROV;  * Delete cases with same SSN DOB and PROV;
run;

* Count the number of providers for each child in the month and create NUMPROV;
proc means data=numc2 noprint;
var COPAYAMT;
by SSN DOB;
output out=snumpv mean=cm;
run;
data snump (keep=SSN DOB NUMPROV);
```

```
set snumpv;
NUMPROV = _FREQ_;
run;

* merging data that has SSN with new variable NUMPROV;
proc sort data = snump;
by SSN DOB;
run;
data numc22;
merge numc2 snump;
by SSN DOB;
run;

* Sorting the sample, and then dividing the sample into the first providers and 2nd providers;
proc sort data=numc22;
by SSN DOB descending HR_BILL;
run;

data dupssn nodupssn;
set numc22;
by SSN DOB;
if first.SSN and first.DOB then output nodupssn; else output dupssn;
run;

data dupssn2 nodupsn2;
set dupssn;
by SSN DOB;
if first.SSN and first.DOB then output nodupsn2; else output dupssn2;
run;

*Use the data set with second providers to create the second provider information variables;
data nodups2 (drop=PROV FAC REL HOME CLM_PAY HR_BILL COPAYAMT);
set nodupsn2;
PROV2 = PROV;
FAC2 = FAC;
REL2 = REL;
HOME2 = HOME;
CLMP2 = CLM_PAY;
HR_B2 = HR_BILL;
COPAY2 = COPAYAMT;
run;

proc sort data=nodups2;
by SSN DOB;
run;

proc sort data=nodupssn;
by SSN DOB;
run;
```

```
* merge the data sets back together with one child per line with up to two providers by PRIM_ID ;
data nodupsnc;
merge nodupssn nodups2;
by SSN DOB;
run;

* delete cases where first provider is not the main provider. Check the log file to make sure the
  observation difference between NODUPVC and SBYPROV is small;
data sbyssn;
set nodupsnc;
if HR_B2 > HR_BILL then delete;
run;

~ ~
* Same steps by DOB;
proc sort data=c5;
by descending CLM_PAY descending HR_BILL;
run;

proc sort data=c5 out=numc3 NODUPKEY;
by CASE_NUM DOB PROV;  * Delete cases with same CASE_NUM DOB and PROV;
run;

* Count the number of providers for each child in the month and create NUMPROV;
proc means data=numc3 noprint;
var COPAYAMT;
by CASE_NUM DOB;
output out=dnumpv mean=cm;
run;
data dnump (keep=CASE_NUM DOB NUMPROV);
set dnumpv;
NUMPROV = _FREQ_;
run;

* Merging data that has DOB with new variable NUMPROV;
proc sort data = dnump;
by CASE_NUM DOB;
run;
data numc33;
merge numc3 dnump;
by CASE_NUM DOB;
run;

* Dividing the sample into the first providers and 2nd providers;
data p1 p2;
set numc33;
if numprov=1 then output p1; else if numprov=2 then output p2;
run;

proc sort data=p2;
by CASE_NUM DOB descending HR_BILL; run;
```

```
* IF numprov <=2 then you can use CASE_NUM to get 'nodupdc' because different children may
have the same DOB;
data dupdob nodupdob;
set p2;
by CASE_NUM;
if first.CASE_NUM then output nodupdob; else output dupdob;
run;

*Use the data set with second providers to create the second provider information variables;
data nodupd2 (drop=PROV FAC REL HOME CLM_PAY HR_BILL COPAYAMT);
set dupdob;
PROV2 = PROV;
FAC2 = FAC;
REL2 = REL;
HOME2 = HOME;
CLMP2 = CLM_PAY;
HR_B2 = HR_BILL;
COPAY2 = COPAYAMT;
run;

proc sort data=p1;
by CASE_NUM DOB;
run;

proc sort data=nodupdob;
by CASE_NUM DOB;
run;

proc sort data=nodupd2;
by CASE_NUM DOB;
run;

* Merge data sets back together with one child per line with up to two providers by PRIM_ID;
data sbydob;
merge p1 nodupdob nodupd2;
by CASE_NUM DOB;
run;

* Merge together the 3 data sets to create data set 'childdat';
data childat;
set sbyprov sbyssn sbydob;
run;

proc sort data=childat;
by CASE_NUM;
run;

data cpamt (keep=CASE_NUM COPAYAMT COPAY2 clm_pay clmp2 sm);
set childat;
```

```
if copayamt eq 499.95 or copayamt eq 999.9 or copayamt ge 2000 then copayamt=.;
if copay2 eq 499.95 or copay2 eq 999.9 or copay2 ge 2000 then copay2=.;
run;

* Sum up co-payments for first and second providers, across all children in a family to create the
   variable NEWFAMCP (which will be the same for all children in a family). Note: Because only a
   few children had more than 2 providers in a month, we only kept data on the first and second
   providers;
proc report nowd headline headskip out=cpout; * nowd headline, and headskip are formatting
    functions. Nowd – runs without the report window and sends its output to the SAS procedure
    output, headline – underlines all column headers, headskip – writes a blank line beneath all
    column headers at the top of each page of the report;
column case_num copayamt copay2 newfamcp; * variable specification;
define case_num / group; * group case_num;
define copayamt / sum format=dollar10.2 noprint; * sum copayamt by case_num and put in dollar
format;
define copay2 / sum format=dollar10.2 noprint;  * sum copay2 by case_num and put in dollar format;
define newfamcp / computed format=dollar10.2; * compute 'newfamcp' ;

compute newfamcp; *compute new variable 'newfamcp';
newfamcp=sum(copayamt.sum, copay2.sum); *sum clm_pay and clmp2;
endcomp; * end compute;
run;

proc sort data=cpout;
by CASE_NUM;
run;

* Sum up subsidies for first and second providers, across all children in a family to create the
   variable NEWFAMSV (which will be the same for all children in a family). Note: Because only a
   few children had more than 2 providers in a month, we only kept data on the first and second
   providers;
proc report nowd headline headskip data=cpamt out=svout; * see above comments;
column case_num clm_pay clmp2 newfamsv sm;
define case_num / group;
define sm / mean noprint;
define clm_pay / sum format=dollar10.2 noprint;
define clmp2 / sum format=dollar10.2 noprint;
define newfamsv / computed format=dollar10.2;

compute newfamsv;
newfamsv=sum(clm_pay.sum, clmp2.sum);
endcomp;
run;

proc sort data=svout;
by CASE_NUM;
run;
* Create SAS child-month data set by merging 'childdat' and the data sets with NEWFAMCP
(cpout) and NEWFAMSV (svout). This data set has one line per child, info on up to 2 providers;
```

```
data save.c0900d;
 merge childat cpout svout;
 by case_num; run;
```

D.2 Family-level

Objective: Read the raw family data file and create new family data.

♦♦

```
libname save 'c:\My Documents\Duration Study\Three Data Years\1999 Data\';

* Reading the original text or excel (.csv) file and create a SAS data file;
data save.f0300d;
infile 'c:\my documents\Duration Study\Three Data Years\2000 data\f0300.csv'
input CASE_NUM $1-10 @11 SM yymmn6. PSSN $21-30 @ 31 CCBG yymmn6.
FIP 41-50 SINGPAR 51-60 HHINCOM 61-70 EMPLOYED 71-80 IV 81-90
TANF 91-100 HSG 101-110 FS 111-120 NON 121;
run;

* Decimal is added to household income variable (HHINCOM);
HHINCOM=HHINCOM/100;
format HHINCOM comma8.2;

proc sort; by CASE_NUM;
run;
```

D.3 Child-Family merge

Objective: For each month, merge the child and family data sets.

♦ ♦

```
libname save 'c:/my documents/Duration Study/Three Data Years/2000 Data/';
libname out 'c:/My Documents/Duration Study/Three Data Years/Child Family Merge/';

* Reading in the SAS child-level data set created above;
data c0;
set save.c0900d;
run;

* Count the number of families in the child data by counting the number of unique case numbers;
proc sort data=c0 out=Numcf NODUPKEY;
by CASE_NUM;
run;

* Reading in the SAS family-level data set created above;
data f0 (drop=SM);
set save.f0900d;
run;

* Count the number of families in the family data by counting the number of unique case numbers;
proc sort data=f0;
by CASE_NUM HHINCOM;
run;
proc sort data=f0 out=Numuf NODUPKEY;
by CASE_NUM;
run;

data mcf;
merge Numcf Numuf;
by CASE_NUM;
run;

* Check for cases without matching observations from the other data set. That is, are there cases in
   the family data but not in the child data, and are there cases in the child data but not in the family
   data. The few that were typically found usually resulted from the fact that a few of the child cases
   were deleted in each month due to errors. Thus, you would expect the number to be small (<10).
   The approach was to select a variable only found in each of the respective child or family data
   set.;
data cnof fnoc;
set mcf;
if IV=. then output cnof; * √ whether there are cases in the child data but not in the family data;
if SM=. then output fnoc; * √ whether there are cases in the family data but not in the child data;
run;

proc sort data=c0;
```

```
by CASE_NUM; run;
* Merge by case number the child and family data sets for the month and drop cases that are
  missing the matching observation. Note: If you have more than 25 unmatched observations for
  either the child or family data, we suggest you explore the raw data further as to the cause of
  these unmatched observations;
data out.cf0900d;
merge c0 Numuf;
by CASE_NUM;

if sm=. then delete;  * delete cases in family data but not found in the child data;
if IV=. then delete;  * delete cases in child data but not found in the family data;
run;
```

D.4 Longitudinal-RC file

Objective: Put child-family merge data sets into a longitudinal data set and then randomly select one child in each family for subsequent analysis.

♦♦

```
libname out 'c:/my documents/Duration Study/Three Data Years/Child Family Merge';
libname save 'c:/my documents/Duration Study/Three Data Years/Analysis';

* Put monthly child-family data sets (October 97 – September 00) into a longitudinal data set;

data all;
set out.cf1097d out.cf1197d out.cf1297d
out.cf0198d out.cf0298d out.cf0398d
out.cf0498d out.cf0598d out.cf0698d
out.cf0798d out.cf0898d out.cf0998d
out.cf1098d out.cf1198d out.cf1298d
out.cf0199d out.cf0299d out.cf0399d
out.cf0499d out.cf0599d out.cf0699d
out.cf0799d out.cf0899d out.cf0999d
out.cf1099d out.cf1199d out.cf1299d
out.cf0100d out.cf0200d out.cf0300d
out.cf0400d out.cf0500d out.cf0600d
out.cf0700d out.cf0800d out.cf0900d;
run;

*Keep one line for EACH child;
proc sort data=all out=nodupc NODUPKEY;
by CASE_NUM DOB;
run;

COUNT the number of children in each family;
proc freq data=nodupc;
tables CASE_NUM/out=bycount noprint;
run;
* Note: bycount data set contains variables CASE_NUM COUNT PERCENT;

proc sort data=nodupc;
by CASE_NUM;
run;

proc sort data=bycount;
by CASE_NUM;
run;
```

```
* The following code uses the idea of conditional probability to randomly select one child in each
family to insure that each observation has the same chance of being selected;
data random;
merge nodupc bycount;
by CASE_NUM;
retain k; * The value of k will be retained within this iteration;
if first.CASE_NUM then k=1; * When the data step reads the first observation from a CASE_NUM
                              category, k is set to 1;
if ranuni(62882575)<=k/count then do; * Selects one observation from each CASE_NUM at random /
                                        ranuni generates a random number between 0 and 1
                                        following a uniform distribution. The number in ( )
                                        provides the initial seed value for the random number
                                        generator;
  output;
  k=k-1; * When an observation is selected, k is set to 0 which prevents any other observations from
           being selected from that CASE_NUM;
 end;
 count=count-1; * Reduce count of subsidized children in the family by 1;
 run;

 proc sort data=random;
 by CASE_NUM;
 run;

 data randomc;
 merge random(keep=CASE_NUM DOB k) bycount(keep=CASE_NUM COUNT);
 by CASE_NUM;
 NUMC=COUNT; * Renaming COUNT to NUMC;
 run;

 proc sort data=all;
 by CASE_NUMM DOB;
 run;

 proc sort data=randomc;
 by CASE_NUM DOB;
 run;

 data allrc;
 merge all randomc;
 by CASE_NUM DOB;
 if k=. then delete;
 run;

*Twins?;
 proc sort data=allrc out=save.cfallrc nodupkey;
 by CASE_NUM DOB SM;
 run;

 proc freq; tables newfamcp newfamsv;
 run;
```

```
* verifying the above code by viewing the data;

options obs=50;

proc print data=all;
title 'all months of family and child data';
format SM CCBG yymmn6.;
format DOB yymmdd8.;
run;

proc print data=nodupc;
title 'nodupc---first line for each child in ALL';
format SM CCBG yymmn6.;
format DOB yymmdd8.;
run;

proc print data=bycount;
title 'bycount---COUNT=number of children in each family';
run;

proc print data=random;
title 'Randomly selected one child per family from nodupc';
format SM CCBG yymmn6.;
format DOB yymmdd8.;
run;

proc print data=randomc;
title 'randomc with correct COUNT=number of children';
format DOB yymmdd8.;
run;

proc print data=save.cfallrc;
title 'out.cfallrc---the Randomly selected child per family in ALL';
format SM CCBG yymmn6.;
format DOB yymmdd8.;
run;
```

D.5 Age

Objective: Create age variables before creating other analysis variables

♦♦♦

```
libname save 'C:\My Documents\Duration Study\Three Data Years\Analysis';

data save.nagerc3;
set save.cfallrc3;

* Creating child's age in current month (mage1);

md=month(DOB);
yd=year(DOB);

if SM = 13788 then do;   * October 1997;
val=19; * Sets initial value;
do y=1978 to 1996;
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md; * mage = child's age in months;
   leave; * Once mage is determined for that observation, then drop out of the DO loop;
  end;
end;
if yd=1997 then mage= 12-md;
if yd ~= 1997 then mage1=mage+10;  * 10 mths are added to or 2 months are subtracted from mage
                                     to obtain child's age in the current month (e.g., 10/97);
  else mage1=mage-2;
end;

else if SM=13819 then do; * November 1997;
val=19;
do y=1978 to 1996;
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md;
  leave;
  end;
end;
if yd=1997 then mage=0*12 + 12-md;
if yd ~=1997 then mage1=mage+11;
else mage1=mage-1;
end;

else if SM=13849 then do; * December 1997;
val=19;
do y=1978 to 1996;
  val=val-1;
  if yd=y then do;
```

```
     mage=val*12+12-md;
  leave;
  end;
end;
if yd=1997 then mage=0*12 + 12-md;
if yd ~=1997 then mage1=mage+12;
else mage1=mage;
end;

else if SM =13880 then do;  * January 1998;
val=18;
do y=1980 to 1997;
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md;
  leave;
  end;
end;
if yd=1998 then mage=0*12 + 12-md;
if yd ~=1998 then mage1=mage+1;
else mage1=mage-11;
end;

else if SM=13911 then do;  * February 1998;
val=18;
do y=1980 to 1997;
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md;
   leave;
  end;
end;
if yd=1998 then mage=0*12 + 12-md;
if yd ~=1998 then mage1=mage+2;
else mage1=mage-10;
end;

else if SM=13939 then do;  * March 1998;
val=18;
do y=1980 to 1997;
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md;
   leave;
  end;
end;
if yd=1998 then mage=0*12 + 12-md;
if yd ~=1998 then mage1=mage+3;
else mage1=mage-9;
end;
```

```
else if SM=13970 then do;  * April 1998;
val=18;
do y=1980 to 1997;
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md;
   leave;
  end;
end;
if yd=1998 then mage=0*12 + 12-md;
if yd ~=1998 then mage1=mage+4;
else mage1=mage-8;
end;

else if SM=14000 then do;  * May 1998;
val=18;
do y=1980 to 1997;
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md;
   leave;
  end;
end;
if yd=1998 then mage=0*12 + 12-md;
if yd ~=1998 then mage1=mage+5;
else mage1=mage-7;
end;

else if SM=14031 then do;  * June 1998;
val=18;
do y=1980 to 1997;
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md;
   leave;
  end;
end;
if yd=1998 then mage=0*12 + 12-md;
if yd ~=1998 then mage1=mage+6;
else mage1=mage-6;
end;

else if SM=14061 then do;  * July 1998;
val=18;
do y=1980 to 1997;
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md;
   leave;
  end;
```

```
end;
if yd=1998 then mage=0*12 + 12-md;
if yd ~=1998 then mage1=mage+7;
else mage1=mage-5;
end;

else if SM=14092 then do;  * August 1998;
val=18;
do y=1980 to 1997;
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md;
   leave;
  end;
end;
if yd=1998 then mage=0*12 + 12-md;
if yd ~=1998 then mage1=mage+8;
else mage1=mage-4;
end;

else if SM=14123 then do;  * September 1998;
val=18;
do y=1980 to 1997;
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md;
    leave;
  end;
end;
if yd=1998 then mage=0*12 + 12-md;
if yd ~=1998 then mage1=mage+9;
else mage1=mage-3;
end;

else if SM=14153 then do; * October 1998;
val=18;
do y=1980 to 1997;
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md;
   leave;
  end;
end;
if yd=1998 then mage=0*12 + 12-md;
if yd ~=1998 then mage1=mage+10;
else mage1=mage-2;
end;

else if SM=14184 then do;  * November 1998;
val=18;
do y=1980 to 1997;
```

```
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md;
            leave;
  end;
end;
if yd=1998 then mage=0*12 + 12-md;
if yd ~=1998 then mage1=mage+11;
else mage1=mage-1;
end;

else if SM=14214 then do;  * December 1998;
val=18;
do y=1980 to 1997;
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md;
   leave;
  end;
end;
if yd=1998 then mage=0*12 + 12-md;
if yd ~=1998 then mage1=mage+12;
else mage1=mage;
end;

else if SM=14245 then do; * January 1999;
val=18;
do y=1981 to 1998;
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md;
   leave;
  end;
end;
if yd=1999 then mage=0*12 + 12-md;
if yd ~=1999 then mage1=mage+1;
else mage1=mage-11;
end;

else if SM=14276 then do;  * February 1999;
val=18;
do y=1981 to 1998;
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md;
   leave;
  end;
end;
if yd=1999 then mage=0*12 + 12-md;
if yd ~=1999 then mage1=mage+2;
else mage1=mage-10;
```

```
end;

else if SM=14304 then do;  * March 1999;
val=18;
do y=1981 to 1998;
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md;
   leave;
  end;
end;
if yd=1999 then mage=0*12 + 12-md;
if yd ~=1999 then mage1=mage+3;
else mage1=mage-9;
end;

else if SM=14335 then do;  * April 1999;
val=18;
do y=1981 to 1998;
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md;
    leave;
  end;
end;
if yd=1999 then mage=0*12 + 12-md;
if yd ~=1999 then mage1=mage+4;
else mage1=mage-8;
end;

else if SM=14365 then do;  * May 1999;
val=18;
do y=1981 to 1998;
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md;
   leave;
  end;
end;
if yd=1999 then mage=0*12 + 12-md;
if yd ~=1999 then mage1=mage+5;
else mage1=mage-7;
end;

else if SM=14396 then do;  * June 1999;
val=18;
do y=1981 to 1998;
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md;
  leave;
```

```
  end;
end;
if yd=1999 then mage=0*12 + 12-md;
if yd ~=1999 then mage1=mage+6;
else mage1=mage-6;
end;

else if SM=14426 then do;  * July 1999;
val=18;
do y=1981 to 1998;
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md;
  leave;
  end;
end;
if yd=1999 then mage=0*12 + 12-md;
if yd ~=1999 then mage1=mage+7;
else mage1=mage-5;
end;

else if SM=14457 then do; * August 1999;
val=18;
do y=1981 to 1998;
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md;
   leave;
  end;
end;
if yd=1999 then mage=0*12 + 12-md;
if yd ~=1999 then mage1=mage+8;
else mage1=mage-4;
end;

else if SM=14488 then do; * September 1999;
val=18;
do y=1981 to 1998;
  val=val-1;
  if yd=y then do;
     mage=val*12+12-md;
   leave;
  end;
end;
if yd=1999 then mage=0*12 + 12-md;
if yd ~=1999 then mage1=mage+9;
else mage1=mage-3;
end;
```

```
agemm = mage1;
age = mage1 / 12;  * Age in current month—unit is years;
format age COMMA5.2;

* Age categories;
if agemm ge 0 and agemm lt 12 then agecat0 = 1; else agecat0 = 0;
if agemm ge 12 and agemm lt 24 then agecat1 = 1; else agecat1 = 0;
if agemm ge 24 and agemm lt 36 then agecat2 = 1; else agecat2 = 0;
if agemm ge 36 and agemm lt 48 then agecat3 = 1; else agecat3 = 0;
if agemm ge 48 and agemm lt 60 then agecat4 = 1; else agecat4 = 0;
if agemm ge 60 and agemm lt 72 then agecat5 = 1; else agecat5 = 0;
if agemm ge 72 and agemm lt 84 then agecat6 = 1; else agecat6 = 0;
if agemm ge 84 and agemm lt 96 then agecat7 = 1; else agecat7 = 0;
if agemm ge 96 and agemm lt 108 then agecat8 = 1; else agecat8 = 0;
if agemm ge 108 and agemm lt 120 then agecat9 = 1; else agecat9 = 0;
if agemm ge 120   then agecat10 = 1; else agecat10 = 0;
run;

proc freq data=save.nagerc3;
tables agecat0 agecat1 agecat2 agecat3 agecat4 agecat5
agecat6 agecat7 agecat8 agecat9 agecat10;
run;

proc means data=save.nagerc3  n nmiss min max mean;
run;
```

D.6 Analysis data file

Objective: Primary analysis that creates the analysis data set 'newORscc', includes all months of all spells excluding all spells for any child who received subsidy in Oct 97 and therefore was a left-censored case

♦ ♦

```
libname  save 'C:\My Documents\Duration Study\Three Data Years\Analysis';
libname new 'C:\My Documents\Duration Study\Three Data Years\Analysis';

proc sort data=save.nagerc3;
by case_num sm;
run;

data all ;
set save.nagerc3;

if agemm lt 0 then agemm = '.';
if agemm eq '.' then agemiss = 1;
else agemiss = 0;

* Creating new variables;
* sm=service month;
if sm = 13788 then period = 1;
if sm = 13819 then period = 2;
if sm = 13849 then period = 3;
if sm = 13880 then period = 4;
if sm = 13911 then period = 5;
if sm = 13939 then period = 6;
if sm = 13970 then period = 7;
if sm = 14000 then period = 8;
if sm = 14031 then period = 9;
if sm = 14061 then period = 10;
if sm = 14092 then period = 11;
if sm = 14123 then period = 12;
if sm = 14153 then period = 13;
if sm = 14184 then period = 14;
if sm = 14214 then period = 15;
if sm = 14245 then period = 16;
if sm = 14276 then period = 17;
if sm = 14304 then period = 18;
if sm = 14335 then period = 19;
if sm = 14365 then period = 20;
if sm = 14396 then period = 21;
if sm = 14426 then period = 22;
if sm = 14457 then period = 23;
if sm = 14488 then period = 24;

periodno = period + 3;
```

```
if SM=13788 then sm9710=1; else sm9710=0;
if SM=13819 then sm9711=1; else sm9711=0;
if SM=13849 then sm9712=1; else sm9712=0;
if SM=13880 then sm9801=1; else sm9801=0;
if SM=13911 then sm9802=1; else sm9802=0;
if SM=13939 then sm9803=1; else sm9803=0;
if SM=13970 then sm9804=1; else sm9804=0;
if SM=14000 then sm9805=1; else sm9805=0;
if SM=14031 then sm9806=1; else sm9806=0;
if SM=14061 then sm9807=1; else sm9807=0;
if SM=14092 then sm9808=1; else sm9808=0;
if SM=14123 then sm9809=1; else sm9809=0;
if SM=14153 then sm9810=1; else sm9810=0;
if SM=14184 then sm9811=1; else sm9811=0;
if SM=14214 then sm9812=1; else sm9812=0;
if SM=14245 then sm9901=1; else sm9901=0;
if SM=14276 then sm9902=1; else sm9902=0;
if SM=14304 then sm9903=1; else sm9903=0;
if SM=14335 then sm9904=1; else sm9904=0;
if SM=14365 then sm9905=1; else sm9905=0;
if SM=14396 then sm9906=1; else sm9906=0;
if SM=14426 then sm9907=1; else sm9907=0;
if SM=14457 then sm9908=1; else sm9908=0;
if SM=14488 then sm9909=1; else sm9909=0;

if FAC='CNT' then                TYPE='CENTER     ';
if FAC='FAM' and REL='N' then TYPE='FAM-NOTREL';
if FAC='FAM' and REL='Y' then TYPE='FAM-REL';
if FAC='GRP' and REL='N' then TYPE='GRP-NOTREL';
if FAC='GRP' and REL='Y' then TYPE='GRP-REL';
if FAC='HOM' and REL='Y' then TYPE='HOME-REL';
if FAC='HOM' and REL='N' then TYPE='HOME-NOTREL';

if FAC='CNT' then caretype = 5;
if FAC='FAM' and REL='N' then caretype = 2;
if FAC='FAM' and REL='Y' then caretype = 1;
if FAC='GRP' and REL='N' then caretype = 6;
if FAC='GRP' and REL='Y' then caretype = 8;
if FAC='HOM' and REL='Y' then caretype = 3;
if FAC='HOM' and REL='N' then caretype = 4;

if caretype = 1 then caretyp1 = 1; else caretyp1 = 0;
if caretype = 2 then caretyp2 = 1; else caretyp2 = 0;
if caretype = 3 then caretyp3 = 1; else caretyp3 = 0;
if caretype = 4 then caretyp4 = 1; else caretyp4 = 0;
if caretype = 5 then caretyp5 = 1; else caretyp5 = 0;
if caretype = 6 then caretyp6 = 1; else caretyp6 = 0;
if caretype = 8 then caretyp8 = 1; else caretyp8 = 0;

*Group all relative care (in-home, family and group home);
```

```
if caretype = 1 or caretype = 3 or caretype = 8 then
caretyp7 = 1; else caretyp7 = 0;
if caretype = '.' Then caremiss = 1; else caremiss = 0;

* Add group homes to family providers;
if caretype = 1 then carecat = 7;
else if caretype = 2 then carecat = 2;
else if caretype = 3 then carecat = 7;
else if caretype = 4 then carecat = 4;
else if caretype = 5 then carecat = 5;
else if caretype = 6 then carecat = 2;
else if caretype = 7 then carecat = 7;
else if caretype = 8 then carecat = 7;

* ethnicity;
if race = 'B' then race1 = 1; else race1 = 0;
if race = 'W' then race2 = 1; else race2 = 0;
if race = 'H' then race3 = 1; else race3 = 0;
if race = 'A' then race4 = 1; else race4 = 0;
if race = 'I' then race5 = 1; else race5 = 0;
if race = 'O' then race6 = 1; else race6 = 0;
if race = '.' or race = 'U' then racemiss = 1;
else racemiss = 0;

if singpar = 1 then single = 1;
else single = 0;
if singpar gt 1 or singpar = '.' then spmiss = 1;
else spmiss = 0;

if reas = 24 then reason2 = 1; else reason2 = 0;
if reas = 29 or reas = 70 then reason3 = 1; else reason3 = 0;
if reas = 62 or reas = 89 or reas = 90 or reas = 92 or reas = 93
or reas = 94 then reason1 = 1;
else reason1 = 0;

if reas = 62 then reas62 = 1; else reas62 = 0;
if reas = 70 then reas70 = 1; else reas70 = 0;
if reas = 92 then reas92 = 1; else reas92 = 0;
if reas = 93 then reas93 = 1; else reas93 = 0;
if reas = 29 then reas29 = 1; else reas29 = 0;
if reas = 91 or reas = 97 or reas = 99 then reasmiss = 1;
else reasmiss = 0;

*Fix TANF variable -- set to missing if PSSN is missing;
if PSSN = ' ' then do;
TANF = .;
psmiss = 1;
end;

*Parent's employment and TANF status;
if tanf ne '.' then do;
```

```
if tanf = 0 and employed = 1 then parstat1=1; else parstat1 = 0;
if tanf = 1 and employed = 1 then parstat2=1; else parstat2 = 0;
if tanf = 1 and employed = 0 then parstat3=1; else parstat3 = 0;
if tanf = 0 and employed = 0 then parstat4=1; else parstat4 = 0;
psmiss = 0;
end;

* Use new family copay and subsidy value variables and fix family copay amounts;
if newfamcp = 999.90 or newfamcp ge 9998
then newfamcp = .;

if newfamcp = . then copymiss = 1;
else copymiss = 0;

if newfamcp eq 0 then nocopay = 1;
else nocopay = 0;

if newfamcp gt 0 then copay=1;
else copay = 0;

* Adding in TANF grant to household income;
if tanf = 1 then do;
if numc = 1 then hhincom2 = hhincom + 395;
else if numc = 2 then hhincom2 = hhincom + 460;
else if numc = 3 then hhincom2 = hhincom + 565;
else if numc = 4 then hhincom2 = hhincom + 660;
else if numc ge 5 then hhincom2 = hhincom + 765;
end;
else if tanf = 0 then hhincom2=hhincom;

* Creating income categories;
if hhincom2 = 0 then inccat0=1; else inccat0 = 0;
if 0 < hhincom2 le 500 then inccat1=1; else inccat1=0;
if 500 < hhincom2 le 1000 then inccat2=1; else inccat2=0;
if 1000 < hhincom2 le 1500 then inccat3=1; else inccat3=0;
if 1500 < hhincom2 le 2000 then inccat4=1; else inccat4=0;
if hhincom2 gt 2000 then inccat5=1; else inccat5 = 0;

* nsv = net subsidy value;
nsv = newfamsv-newfamcp;

if hhincom2 gt 0 then do;
sv2inc = newfamsv/hhincom2;
cp2inc = newfamcp/hhincom2;
nsv2inc = nsv/hhincom2;
end;

periodsq = period**2;

if period = 4 or period = 16 or period = 28 then mojan = 1; else mojan = 0;
if period = 5 or period = 17 or period = 29 then mofeb = 1; else mofeb = 0;
```

```
if period = 6 or period = 18 or period = 30 then momar = 1; else momar = 0;
if period = 7 or period = 19 or period = 31 then moapr = 1; else moapr = 0;
if period = 8 or period = 20 or period = 32 then momay = 1; else momay = 0;
if period = 9 or period = 21 or period = 33 then mojun = 1; else mojun = 0;
if period = 10 or period = 22 or period = 34 then mojul = 1; else mojul = 0;
if period = 11 or period = 23 or period = 35 then moaug = 1; else moaug = 0;
if period = 12 or period = 24 or period = 36 then mosep = 1; else mosep = 0;
if period = 13 or period = 1 or period = 25 then mooct = 1; else mooct = 0;
if period = 14 or period = 2 or period = 26 then monov = 1; else monov = 0;
if period = 15 or period = 3 or period = 27 then modec = 1; else modec = 0;

* create a subset of data for the first observed subsidy month by case_num and rename select variables;
data first (keep=case_num startm tanffm reasfm singpfm
  racefm emplfm carefm carefm2 agemmfm agefm ccbgfm
  copayfm fsubvfm hhincfm fipfm numcfm fsfm regfm leftcen);
  set all;
  by case_num;


if first.case_num;
startm = sm;
tanffm = tanf;
reasfm = reas;
singpfm = singpar;
racefm = race;
emplfm = employed;
carefm = caretype;
carefm2 = caretyp2 ;
agemmfm = agemm;
agefm = age;
ccbgfm = ccbg;
copayfm = copay;
cpamtfm = newfamcp;
fsubvfm = newfamsv;
hhincfm = hhincom2;
fipfm = fip;
numcfm = numc;
fsfm = fs;
regfm = regulate;

* If service month is October 1997 then left censored=1;
if sm = 13788 then leftcen = 1;
else leftcen = 0;
run;

data newall;
  merge all first;
  by case_num;
```

```
*Keep all cases but those in the first month (October 1997);
if leftcen = 0;
run;

data allmos;
  retain t1 spelnmbr ;  * The value of t1 and spelnmbr will be retained within this iteration;
  retain lastdata 24;    *Creates the variable lastdata to contain the value 24 and to retain this value
                          throughout this iteration—used to test for censoring;
  set newall nobs=nobs; *newall is transformed into a dataset named allmos, nobs=nobs is equal to
                          the number of obs in the input dataset;
    by case_num;

if first.case_num then do;
      t1=0; spelnmbr=1;   *t1 and t2 are start and stop time variables of the obs ;
        transe1 = 0; transe2 = 0;
        transt1 = 0; transt2 = 0;
        scchist = 0;
      end;
else t1=t1+1;
t2=t1+1;
scchist = t1;
obplus=_n_ + 1;  *obplus is the position in the dataset of the next obs,
                  _n_ is a SAS automatic variable that starts at 1 for the first obs in the dataset and
                  increases by 1 each time the program loops;


*Designed to look ahead 1 obs from the current obs to see if the current obs is the end of a spell;
if _n_< nobs then  *only read the next obs if _n_< nobs, nobs is equal to the number of obs in the
                    input dataset;
set newall (keep=case_num sm period employed tanf
      rename=(case_num=casenumx sm=smx period=periodx
          employed = employx tanf=tanfx))  *For example, casenumx = case number for next obs;
      point=obplus;  *Access the dataset at the obs pointed to by obplus (see above);


*This code is designed to decipher if current obs is the end of a spell;
if case_num=casenumx and periodx=period+1 and _n_<= nobs then do; *The next observation must
                              satisfy these 3 conditions to satisfy the requirements of being a
                              continuation of the current spell;
*Next obs is in the same spell;
 status = 0; sccexit = 0; *Status is equal to 0 if the event of interest (exiting from SCC) did not occur
                          or if the obs is censored, 1 otherwise;
if case_num=casenumx then do;
      if employed=0 and employx = 1 then transe1 = 1;  *Transition into employment from current
                                                         month to next month;
         else transe1= 0;
        if employed=1 and employx=0 then transe2 = 1;  *Transition out of employment from current
                                                         month to next month;
         else transe2 = 0;
        if tanf=0 and tanfx=1 then transt1 = 1; * Transition into TANF from current mth to next mth;
         else transt1 = 0;
```

```
        if tanf=1 and tanfx=0 then transt2 = 1; * Transition out of TANF from current mth to next mth;
          else transt2 = 0;
 output;
 end;
 end;

else do;

*This obs is end of spell or censor;
if period=lastdata then do;
status = 0; sccexit=0;
transe1=0; transe2=0;
transt1=0; transt2=0;
end;
else do;
status = 1; sccexit=1;
end;
  output;

*Set up for next spell;
spelnmbr=spelnmbr+1;
t1=-1;
end;
run;

*Creates the permanent SAS data set 'newORscc';
data new.newORscc;
  set allmos;
scchist1 = t2;
run;

* Procedures administered to check the data;
proc contents data=new.newORscc;
run;

proc means data= new.newORscc (where=(spelnmbr=1)) n nmiss min max mean;
title 'First spell only';
run;

proc means data= new.newORscc  n nmiss min max mean;
title 'All spells';
run;

options obs=100;
proc print data=new.newORscc;
  var case_num sm spelnmbr  employed transe1 transe2 tanf transt1 transt2
   t2 scchist scchist1 status;
format sm yymmn6.;
run;
```

```
options obs = 50;

proc print data= new.newORscc;
  var case_num spelnmbr sm period t1 t2 scchist scchist1
sm9710 sm9711 sm9712 sm9801 sm9802 sm9803 sm9804 sm9805 sm9806
sm9807 sm9808 sm9809 sm9810 sm9811 sm9812 sm9901 sm9902 sm9903
sm9904 sm9905 sm9906 sm9907 sm9908 sm9909;
format sm yymmn6.;
run;

options obs = 50;
proc print data=new.newORscc;
run;

proc print data= new.newORscc;
  var case_num dob spelnmbr sm startm tanf tanffm reas reasfm
singpar singpfm race racefm employed emplfm;
format sm yymmn6.;
format dob yymmdd8.;
run;

proc means data=new.newORscc (where=(spelnmbr=1));
  var sccexit agecat0 agecat1 agecat2 agecat3
    agecat4 agecat5 agecat6 agecat7 agecat8 agecat9
    agecat10
    numc newfamsv newfamcp fcopaych nocopay
    inccat0 inccat1 inccat2 inccat3 inccat4 inccat5
    caretyp2 caretyp4 caretyp6 caretyp7 caretyp5
    regulate
    parstat1 parstat2 parstat3 parstat4
    scchist1
    mojan mofeb momar moapr momay mojun mojul moaug
    mosep mooct monov modec
    period
    periodno;
 run;

data fix;
 set new.newORscc (where=(spelnmbr=1));

if parstat1 eq '.' then parstat1 =0;
if parstat2 eq '.' then parstat2 =0;
if parstat3 eq '.' then parstat3 =0;
if parstat4 eq '.' then parstat4 =0;

proc means data=fix;
  var sccexit parstat1 parstat2 parstat3 parstat4 ;
run;

proc univariate data= new.newORscc (where=(spelnmbr=1));
  var newfamsv newfamcp nsv nsv2inc sv2inc cp2inc;
```

```
run;

data withcp;
  set new.newORscc (where=(spelnmbr=1));

 if copay = 1;

proc univariate data= withcp;
  var newfamsv newfamcp nsv nsv2inc sv2inc cp2inc;
run;

proc univariate data= new.newORscc (where=(spelnmbr=1));
  var hhincom2 hhincom;
run;

data firstmo;
  set new.newORscc (where=(spelnmbr=1));
by case_num;

if first.case_num;

proc univariate  data=firstmo;
var newfamsv newfamcp nsv nsv2inc sv2inc cp2inc hhincom2 hhincom;
title 'First month characteristics';
run;

data withcp;
  set firstmo;

if copay eq 1;

proc univariate  data=withcp;
var newfamsv newfamcp nsv nsv2inc sv2inc cp2inc hhincom2 hhincom;
title 'First month characteristics of families with copay';
run;

proc univariate data= new.newORscc (where=(spelnmbr=1));
  var hhincfm fsubvfm copayfm ;
title 'First month characteristics';
run;

data check ;
  set new.newORscc (where=(spelnmbr=1));

if nsv lt 0;

proc print data=check;
  var case_num sm newfamsv newfamcp nsv hhincom2 numc;
run;
```

D.7 Spell-Length

Objective: Estimates durations with both the lifetest method and KaplanMeier.

♦♦

```
libname  save  'C:\My Documents\Duration Study\Wave 3\Child Family Merge';
 libname new 'C:\My Documents\Duration Study\Wave 3\Child Family Merge';

 data new.lastmo;
    set new.newORscc (where=(spelnmbr=1));  *First observed spell;
    by case_num;

if last.case_num;
dur1=scchist1;  *Counter for number of total months so far in the spell;
if spelnmbr eq 1 and sm eq 14488 then rightcen = 1;
else rightcen = 0; * If first observed spell and service month is equal to September 1999 then right
censored=1;
run;

proc freq data=new.lastmo;
  tables rightcen sm dur1 scchist1 scchist ;
  format sm yymmn6.;
run;

* Duration using newORscc data ~ last observation of first spell using lifetest method;
proc lifetest data=new.lastmo method=life
  intervals=1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
17 18 19 20 21 22 23 24;
 time scchist1*rightcen(1);
title 'Duration using newORscc data -- last obs of first spell';
run;

* Duration using newORscc data ~ last observation of first spell using KaplanMeier method;
proc lifetest data=new.lastmo
  intervals=1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
17 18 19 20 21 22 23 24;
 time scchist1*rightcen(1);
title 'Duration using newORscc data -- last obs of first spell';
run;
```

D.8 Reentry

Objective: Determines the percent of children exiting a subsidy spell who return within 3, 6, 9, and 12 months.

♦ ♦

```
libname new 'C:\My Documents\CHILD CARE DURATION study ';

data reentry;
set new.newORsc3;
by case_num;
length exit reentry mm2re mm4ob 3;  *Length is a SAS variable that specifies the number of bytes
                                     [can range from 3 to 8] the SAS system uses to store numeric
                                     variables. Exit, reentry, mm2re and mm4ob are variables that
                                     will be created in this program, and 3 is the number of bytes
                                     'length' is set to;

*This code determines whether the obs represents an exit;
if (period<24) then do; *Using 2 years worth of data;
  exit=last.case_num; *The last case_num is equal to exit;
  if not exit then do;
  pt=_n_+1; *_n_ = the iteration number that the data step is on. If the period is less than 24 and the
             observation that was read in by the first set statement is not the last observation of that
             particular value of case_num, then SAS will retrieve from new.newORsc3 the next
             observation;
  set new.newORsc3 (keep=period rename=(period=periodx)) *periodx is the period status in the next
                                                          obs;
   point=pt; *point= specifies pt (whose value is the observation that is retrieved);
 if (periodx>(period+1)) then exit=1; *If the period status of the next obs is greater than the period of
                                      the current observation + 1, then there must be a break in the
                                      spell resulting in an exit (e.g., periodx=12 > (period=10 + 1 =
                                      11));
 end;
 end;

*Specifying 2 new variables: mm4ob and mm2re;
 if exit then do;
   mm4ob=24-period-1; *mm4ob is the number of months observed ~ number of months between
                       the child's exit and the end of the observation period;
* The next 2 lines sets reentry to 0 so that it isn't considered to be missing;
     if (mm4ob ge 1) then do;
        reentry = 0;
        if not last.case_num then do;
        pt = _n_+1;
     set new.newORsc3 (keep=case_num sm period
       rename=(case_num=casenumx sm=smx period=periodx)) *Renamed variables represent variable
                                                          status in the next obs;
       point=pt;
       reentry = 1;
```

```
      mm2re = (periodx-period); *mm2re represents 'how many months did it take to re-enter' the
                                  subsidy program (e.g., periodx=12 - period=6  = 6);
      end;
     end;
 end;
 run;
 proc freq data=reentry;
  tables reentry exit;
  run;

  data next;
   set reentry;
    by case_num;
    length reentry3 reentry6 reentry9 reentr12 3; *See above explanation of 'length', variables
                                                   reentry3-12 are created below;

if exit then do;
  if (mm4ob ge 3) then reentry3=reentry and (mm2re le 3);

*The 'and (mm2re le 3)' portion of this code is referred to as an Boolean expression. SAS first
evaluates whether number of months observed is greater to or equal to 3, and then the Boolean
expression is evaluated and will return either a 1 or 0 for reentry3 depending on whether the
expression is true. This code is NOT equivalent to: if (mm4ob ge 3) and (mm2re le 3) then
reentry3=reentry, BUT is equivalent to:
        if exit then do
         if (mm4ob ge 3) then do
          if (reentry ne 0) and (mm2re le 3) then reentry3=1
          else reentry3=0
         end
        end;

   if (mm4ob ge 6) then reentry6=reentry and (mm2re le 6);
     if (mm4ob ge 9) then reentry9=reentry and (mm2re le 9);
      if (mm4ob ge 12) then reentr12=reentry and (mm2re le 12);
end;
run;

*Checking results and possible problems with reentry variables;
proc freq data=next;
 tables sm exit reentry reentry3 reentry6 reentry9 reentr12;
 title 'Reentry - whole sample - NewORsc3 data';
 run;

data last;
 set next;
 if (reentry3 = 1 and reentry6 = 0) or
  (reentry6 = 1 and reentr12 = 0);

  proc print data=last;
   var case_num sm period exit mm4ob reentry mm2re
     reentry3 reentry6 reentry9 reentr12;
```

```
title 'Possible problems with reentry variables';
format sm yymmn6.;
run;
```

D.9 Stability

Objective: Examines the proportion of the total time on subsidy that children remained with their primary provider. The primary provider ratio measures the number of subsidized months during which the child continued to receive care from this provider.

♦♦♦

```
libname save 'c:/my documents/Duration Study/Three Data Years/Analysis';

* Create the following variables
        total months of subsidy (cumtime)
        duration (number of months) with primary provider (longest)
        ratio of months with primary provider to months of subsidy
        number of providers

NOTE that if the person had 2 spells but the same provider, count all
the months with the same provider, even if there is a break in time;

*This data set includes all the monthly observations for one random child per family-not just the
first spell;
data s(keep=CASE_NUM DOB SM PROV TANF);
set save.cfallrc3;
run;

proc sort data=s;
by CASE_NUM DOB PROV;
run;

*This step creates a data set with one observation for each child-provider combination. The proc
means procedure automatically creates a variable _freq_ which counts the number of observations
of each child-provider combination because of the by statement. This variable is then used to count
the number of months with the provider. Any variable can be used in place of "TANF", but it must
be numeric;
proc means sum data=s noprint;
var TANF;
by CASE_NUM DOB PROV;
output out=primd mean=TANFm sum=TANFs;
run;

*This step renames the _freq_ variable to provmm. It is a count of the number of months with that
provider, ignoring any breaks in time;
data primpv(keep=CASE_NUM DOB PROV PROVMM);
set primd;
PROVMM=_FREQ_;
run;

*Sort the data so that the provider with the most months is listed first for each child;
proc sort data=primpv;
```

```
by CASE_NUM DOB descending PROVMM; run;
*Keep the number of months for the primary provider, that is the one with the most months;
data primpvd(keep=CASE_NUM DOB  DURPPM);
set primpv;
by CASE_NUM;
if first.CASE_NUM;
DURPPM=PROVMM;  *Duration with primary provider;
run;

*Mean number of providers;
proc means data=primpv noprint;
var PROVMM;
by CASE_NUM DOB;
output out=numpv mean=PROVMMm;
run;

*Count how many providers the child had in all months on subsidy;
data npv(keep=CASE_NUM DOB NUMPROV);
set numpv;
NUMPROV=_FREQ_;
run;

*Creating a variable (CUMTIME) to count all months of subsidy receipt for the child, not just the
first spell;
proc sort data=s;
by CASE_NUM DOB SM;
run;

proc means sum data=s noprint;
var SM;
by CASE_NUM DOB;
output out=smd mean=SMm sum=SMs;
run;

data cumd(keep=CASE_NUM DOB CUMTIME);
set smd;
CUMTIME=_FREQ_;
run;

*Sorting data sets for merge;
proc sort data=cumd;
by CASE_NUM DOB;
run;

proc sort data=npv;
by CASE_NUM DOB;
run;

proc sort data=primpvd;
by CASE_NUM DOB;
run;
```

```
data sall;
merge cumd npv primpvd;
by CASE_NUM DOB;

*NOTE: Need to drop all monthly observations for a child if that child was receiving subsidy in
Oct. 97 and therefore was a left-censored case. The leftcen variable is created in a database of first
observations for each child, and merged back in so that if the child was receiving subsidy in Oct. 97,
then leftcen = 1 for each of its monthly observations, no matter which month. See pages 48-49 for
this code;

*Ratio of months with primary provider to months of subsidy;
RDPPCUM=DURPPM/CUMTIME;

*Ratio of number of providers to months of subsidy;
RNPVCUM=NUMPROV/CUMTIME;

*Ratio of months of subsidy to number of providers;
RCUMNPV=CUMTIME/NUMPROV;

format RDPPCUM RNPVCUM RCUMNPV comma4.2;

if RDPPCUM <=0.25 then RDPPC=     '0.25 or less       ';
if 0.25<RDPPCUM <=0.50 then RDPPC='0.25<RDPPCUM <=0.50';
if 0.50<RDPPCUM <=0.75 then RDPPC='0.50<RDPPCUM <=0.75';
if 0.75 <RDPPCUM <1 then RDPPC='0.75 <RDPPCUM <1';
if RDPPCUM=1 then RDPPC='RDPPCUM=1';

if RNPVCUM <=0.25 then RNPVC='0.25 or less       ';
if 0.25<RNPVCUM<=0.50 then RNPVC='0.25<RDPPCUM <=0.50';
if 0.50<RNPVCUM<=0.75 then RNPVC='0.50<RDPPCUM <=0.75';
if 0.75 <RNPVCUM<1 then RNPVC='0.75 <RNPVCUM<1';
if RNPVCUM=1 then RNPVC='RNPVCUM=1';
run;

* Checking results;
proc sort data=sall;
by CASE_NUM DOB;
run;

proc freq;
tables cumtime durppm numprov rdppc rnpvc;
run;

proc freq;
tables durppm*cumtime rdppc*cumtime
/norow nocol nopercent;
run;

proc freq;
tables cumtime*rdppc /norow nocol nopercent;
```

```
run;

proc freq;
tables cumtime*rdppc;
run;

proc univariate data=sall;
var cumtime durppm numprov rdppcum rnpvcum;
run;

proc sort data= sall;
by CUMTIME;
run;

proc means ;
var RDPPCUM;
by CUMTIME ;
run;

proc freq;
tables cumtime*numprov;
run;
```