



Child Care and Development Fund Issues Related to Data Collection and Analysis

Technical Bulletin #2r

(Reviewed: September 2011)

I. INTRODUCTION

This technical bulletin addresses whether States should develop systems for the submission of full populations versus samples and the various advantages for analysis of subsidized child care data. Your State's subsidized child care data set has the potential to become a valuable resource for researchers, policy makers, and practitioners as the population of children and families in need continues to grow. The development and implementation of the national child care data set in conjunction with State data sets, can inform progressive change and continuous improvement through a better understanding of the characteristics of the subsidized child care population.

The Personal Responsibility and Work Opportunity Act of 1996 requires States and Territories to collect information on all family units receiving assistance through the Child Care Development Fund (CCDF). Section 611(C) requires States to submit monthly case level data on a quarterly basis. The Child Care Bureau has determined that States will be permitted to submit case level data for the entire population or a sample of the population under approved sampling guidelines.

The Child Care Bureau will be developing a Federal Child Care Information System which allows States to submit either a full population or small (approximately 200 families) monthly samples of subsidized child care recipients each quarter (Form ACF-801) and population values for all families and children in care annually (Form ACF-800), for federal reporting purposes.¹

States are encouraged to consider developing quality child care reporting systems that provide sufficient flexibility to produce both required reports to Congress and a variety of analyses needed to continually improve child care policy and practice at local, statewide, and national levels.

¹ States are expected to submit monthly *simple random samples*. However, at least one State that does not have a statewide integrated computer systems for all CCDF program is considering a stratified random sample to ensure proportionate representation of all families receiving some percentage of CCDF funding. The process of creating a stratified random sample entails organizing/sorting a population into homogenous subgroups based on one or more (stratification) variables, and based on each subgroup's relative proportion in the population, randomly selecting the number of cases from each subgroup which constitutes the same proportion of the desired sample size.

II. SHOULD STATES SUBMIT SAMPLE OR POPULATION CASE-LEVEL DATA?

The answer depends upon how your State's subsidized child care data will be used. If the only purpose for the data collection is to produce annual reports on the general characteristics of families receiving assistance, analyses based on simple random samples will suffice. If the subsidized child care data set is to have other uses, however, sampled data should not be used because it limits analysis to descriptive cross-sectional research both at the State and national levels. If entire monthly populations are included in State and national data sets, longitudinal research will be possible in addition to descriptive cross-sectional research.

Advantages and Disadvantages of Submitting Full Population Data

Many States are now redesigning their systems to better meet more complex information needs. States may want to develop information systems which will eventually allow them to submit population data because:

- Population reporting is *less burdensome* for child care workers.

If your State collects all mandated data elements in an automated child care information system, the least burdensome manner of reporting on the monthly subsidized child care caseload is to write a single program to extract required data elements on all program participants. This program creates an electronic file that can then be relayed electronically to the ACF computer which houses the national child care database. Except for the initial writing of the extraction program, this scenario for full population reporting is accomplished without any caseworker intervention and with only a minimal amount of system staff intervention. For States in this situation, sampling creates an additional step in the data submission process.

- By submitting population data, States contribute to a growing National Child Care Data Set of subsidized child care information.

When States submit population data instead of sampled data, they enhance the capacity of a National Child Care Data Set to inform important issues for child care planning and policy. Many more types of analysis are available through the use of population as opposed to sampled data, including 1) the availability of controlled longitudinal research and 2) the availability of monthly and quarterly whole number counts of all children and families receiving the child care subsidy. Sample data restricts analyses to cross-sectional research based estimates of the number of children in care. Table 1 lists some of the analyses that will be available for sample and population data.

The primary drawback of submitting full populations is the cost to develop and maintain a comprehensive information system.

Advantages and Disadvantages of Submitting Small Samples

States may want to develop information systems which will eventually allow them to submit sample data because:

- Sampled data can be easier to provide if the State is not developing a unified data system.
- It may be easier to ensure the quality of sampled data through a quality control reporting system already in place in some States.
- Sampling could be a simple way to meet minimum Federal requirements.

Disadvantages for States which have information systems submitting small samples:

- Sampling creates an additional step in the data submission process for States which have developed comprehensive information systems.
- Longitudinal research questions cannot be addressed through small samples.

III. DATA ANALYSIS POSSIBILITIES

States must determine what typical research possibilities would be useful for programmatic, budget, and legislative planning. Table 1 shows the types of analyses that can be developed from full population or sampled data.

Table 1
State Data Set Structure and Available Analyses²

Data Set Structure	Available Cross-Sectional Designs		Available Longitudinal Designs
Monthly Population Data Submissions	All subgroup analyses Monthly, quarterly, or annual analyses at the national, regional, state, and local levels	All descriptive statistics including caseload counts	Repeated cross-sectional Controlled panel study
200 Cross-Sectional Monthly Samples	Limited subgroup analysis Annual or quarterly analyses at the national and regional levels	All descriptive statistics excluding caseload counts	Repeated cross-sectional (cohort)

Longitudinal Analysis

To answer many critical child care policy questions, the experiences of families over time must be understood. Assessing families' success in maintaining stable child care placements or stable employment, their changing service needs, and/or changing patterns of service utilization all require the longitudinal study of families throughout the course of their participation in CCDF programs. Small cross-sectional samples will not provide information on the same subsidized child care recipients each month, and therefore controlled longitudinal studies will not be possible for States that collect and store sampled data.

² See *Data Analysis Possibilities for the Mandated National Child Care Data Set Required by PRWORA "White-Paper"* (To be released Fall, 1997) for more information.

Controlled longitudinal studies measure change over time while controlling for important exogenous variables that may affect the behavior of subsidized child care recipients. For example, in an analysis of change over time, the use of multiple year birth cohorts (panels) would allow an examination of whether or not the age of a child in care has an effect on changing service needs. It may be that children of welfare families who entered subsidized child care immediately after the implementation of the TANF legislation experience less stability in their child care arrangements and have parents who have more difficulty attaching to the labor force than children of non-welfare families. These difficulties may be due to age effects such as the inaccessibility of subsidized child care services for 0-3 year olds, or period effects, which are due to the implementation of certain policies or social issues that impact cohort members' success in obtaining a job.

When records for the entire population are collected and stored in a State child care database, a family historical record can be built with information collected and submitted to ACF during each succeeding monthly report. This database structure would support conducting controlled longitudinal (retrospective panel) studies. If small cross-sectional samples are stored in State databases, different cases/families will appear in each monthly submission and historical case records cannot be built for each subsidized child care family. By storing population data you will have the flexibility to conduct both longitudinal and cross-sectional research. In addition, by submitting population data to ACF, national longitudinal profiles at the county, State, *and* national levels will be available.

For example, with sample data it will be possible to track how general client characteristics such as race or reason for participation change over time by comparing a series of annual "snapshots". However, these techniques are not a substitute for controlled longitudinal studies because they neglect the dynamic nature of social behavior and the impact of maturation and periodicity on social change. Use of monthly population data supports all studies that could be undertaken with sampled data. The reverse, however, is not true.³

³ Case-linking can be used to create panels for controlled longitudinal research. However, it is not recommended because maintenance of small panel groups is costly and time consuming due to attrition, sample replacement, and follow-up.

Caseload counts

Although many statistical tests and multivariate analyses can be conducted using a sample or population of cases, a significant limitation of cross-sectional samples is the inability to produce *monthly* and *quarterly* whole number counts, since the aggregate report will provide only biannual total population values. For example, using sampled data, if 30% of the families who receive subsidized child care in a given month are former TANF recipients, the *exact* number of families within this target group is indeterminable because the target group's actual monthly population size is unknown.⁴ The samples will provide sufficient information to conduct inferential hypothesis tests,⁵ but monthly and quarterly caseload counts of the total number of children in the subsidized child care population will only be available if population instead of sampled data is used to conduct statistical analyses.

Thus, maximizing the capacity of your child care data set to inform policy makers entails support for longitudinal analysis and monthly and quarterly whole number counts of the number of children and families receiving care.

IV. CONCLUSION

It is in the State's interest to participate in a national data set for many reasons, including the ability to move increasingly toward data-based decision making, form useful partnerships with other States and researchers, and take advantage of technical innovations in data management. For example a national data set could help States:

- Update and streamline administrative data systems to obtain more useful and comparable data on their entire client population;
- Develop complex profiles to better understand local, regional and statewide patterns of supply, utilization, and unmet need;
- Draw a wide variety of subsamples with known characteristics and precise error estimates to obtain specific data needed to answer specific questions;
- Track changing interrelationships among demographic and subsidy patterns over time;
- Obtain estimates for State populations in relation to regional or national estimates;
- Work with other States to better understand how different policies and service delivery systems affect child care supply, demand, and outcomes in the context of broader population demographics and welfare-to-work environments; and
- Participate in a coordinated effort to track trends over time and develop complex profiles across State, regions, population groups, and geographic areas.

It is recommended that States with adequate information collection systems collect and report population data instead of sample data so that county, State, and national level, cross-sectional

⁴ Research on AFDC participants has shown that caseloads fluctuate monthly (Pavetti 1993).

⁵ Inferential statistical methods are procedures for making generalizations about a population based on information contained in a sample.

and longitudinal analyses can be conducted.⁶ If State data collection is limited to the federal minimum of small (n = approximately 200) monthly samples reported each quarter, only quarterly and annual cross-sectional research will be available for a national data set. It is vitally important to the quality of the subsidized child care data and the research that evolves from this initiative that State and national level data sets have the flexibility to not only report annually to Congress, but also to produce statistically valid longitudinal research and subgroup analyses to inform child care policy and practice. The Child Care Bureau will be supporting States in the development of information collection systems reporting on full populations by providing additional technical assistance over the next 3-5 years to States interested in this service.

⁶ If your State does not currently have an automated child care system, a statistically valid/representative sample should be used and disseminated to local researchers and policy analysts who wish to conduct their own analyses.