

AG-3198-C-13-0012

CACFP Sponsor and Provider Characteristics

Clearance Version Final Report

Volume III Study Methodology

August 2018

Frederic B. Glantz
Amy A. Germuth
Karen Della Torre, Westat

Submitted to:

Ashley Chaifetz
Social Science Research Analyst
Special Nutrition Evaluation Branch
Office of Policy Support
Food and Nutrition Service, USDA
3101 Park Center Dr.
Alexandria, VA 22302

Submitted by:



USDA is an equal opportunity provider, employer, and lender

Table of Contents

Introduction	1-1
Chapter 1: Sample Selection.....	1-1
Sampling Frame For The Study.....	1-1
Sampling Design.....	1-3
Main Study Component.....	1-3
At-Risk Component.....	1-3
Sample Selection And Characteristics Of The Initial Sample	1-5
Chapter 2: Data Collection	2-1
Letter To Regional Offices And State Cacfp Administering Agencies.....	2-1
Data Collection Package	2-2
Follow-Up Survey	2-3
Mixed At-Risk Centers.....	2-3
Regular Child Care Centers	2-4
Data Receipt And For Surveys Completed By Mail Data Collection.....	2-5
Data Collection Issues	2-6
Final Distribution Of The Sample.....	2-6
Initial Sample.....	2-6
Chapter 3: Analytic Methodology	3-1
Calculating Standard Errors.....	3-1
Weighting And Treatment Of Missing Values.....	3-1
Base Weights.....	3-1
Nonresponse Adjusted Weights.....	3-2
Post-Stratified Final Weights	3-3
Analytic Sample.....	3-4
Item Nonresponse	3-4
Treatment Of Missing Values	3-4
Nonresponse Bias Analysis.....	3-6

List of Exhibits

Volume III

Chapter 1

Exhibit 1.1: Solutions to Outstanding Issues at the Time of Sample Selection	1-2
Exhibit 1.2: Overview of Sampling Design	1-4
Exhibit 1.3: Distribution of the Initial CACFP Sponsor Sample.....	1-5
Exhibit 1.4: Distribution of CACFP Provider Sample and Completed Surveys by Sampled State	1-6

Chapter 2

Exhibit 2.1: Results of the Follow-Up Survey of Mixed At-Risk Centers.....	2-4
Exhibit 2.2: Results of the Follow-Up Survey of Child Care Centers	2-5
Exhibit 2.3: Sponsor Analytic Sample: Distribution of Completed CACFP Sponsor Surveys by Sample State.....	2-7
Exhibit 2.4: Analytic Sample: Distribution of Completed CACFP Provider Surveys By Sampled State.....	2-8

Chapter 3

Exhibit 3.1: Post-Stratification Domains and Control Totals for the Sponsor Data	3-3
Exhibit 3.2: Post-Stratification Domains and Control Totals for the Provide.....	3-4
Exhibit 3.3: Distribution of Survey Items by Item Response Rate: Sponsor and Provider Surveys.....	3-5
Exhibit 3.4: Auxiliary Data Availability on the Sponsor and Provider Frame Files	3-6
Exhibit 3.5: Comparison of Respondents and Nonrespondents for Continuous Auxiliary Variables by Stratum	3-7
Exhibit 3.6: Comparison of Respondents and Nonrespondents for Categorical Auxiliary Variables by Type Stratum.....	3-8

List of Abbreviations

Abbreviation	Description
AR	At-Risk
ADA	Average Daily Attendance
CACFP	Child and Adult Care Food Program
CCC	Child care center
FDCH	Family day care home
FNS	Food and Nutrition Service
F/RP	Free/reduced-price
FY	Fiscal year
HSC	Head Start center
ICCC	Independent child care center (independent center)
MOS	Measure of Size
OSHC	Outside-school-hours center
PSU	Primary Sampling Unit
PPS	Probability Proportional to Sampling
PY	Program year
SCCC	Sponsored child care center (sponsored center)
SFA	School Food Authority
SFSP	Summer Food Service Program
SNAP	Special Nutrition Assistance Program
USDA	United States Department of Agriculture
WIC	Special Supplemental Nutrition Program for Women, Infants, and Children

INTRODUCTION

This volume describes the methodology used to conduct the CACFP Sponsor and Provider Characteristics Study. It provides additional technical details on the methodology and is intended for technical readers. Chapter 1 presents the details of the multi-stage sampling design, including the desired levels of precision for study estimates and the determination of the necessary sample sizes; Chapter 2 describes the steps used to collect and process the data; and Chapter 3 describes the analytic approach.

Chapter 1: Sample Selection

The CACFP Sponsor and Provider Characteristics Study includes two major objectives: 1. produce national estimates of the characteristics of all CACFP sponsors and providers that serve children; and 2. produce national estimates of the sponsors and centers that participate in the CACFP At-Risk Afterschool Center Program.

SAMPLING FRAME FOR THE STUDY

No list was available that could be used as a sampling frame for the universe of all CACFP sponsors and providers. A cluster sample design was implemented, using States and the District of Columbia (DC, hereafter included as a State) as the primary sampling unit (PSU). A probability sample of 23 States was selected and, in cooperation with State CACFP Agencies in those States, a complete enumeration of CACFP sponsors and providers was constructed for each State.

The sample States were asked to provide a list of all CACFP sponsors by the types of CACFP providers they support, such as family day care homes. The reference month for these lists was October 2014, the first month of the 2014-15 program year. Each State provided the requested lists. Four issues were encountered during the construction of these lists.

1. Disparate terminology across States. While every effort was made to use terminology typically used in the “CACFP community” in the request for these data, nevertheless, some States used idiosyncratic terminology for some requested data items, causing some initial confusion when some study staff thought were unable to provide the requested data. This problem was resolved by providing each State with an information sheet containing the study’s definition of each requested data items.
2. Classification of CACFP sponsors. For sponsors with a single type of provider, the study used the same definitions to classify types of sponsors as is used on Form FNS-44.¹ Some States do not use the Form FNS-44 definitions for sponsors with more than one type of provider. For example, the this form classifies any CACFP sponsor with at least one family day care home participating under its auspices as an family day care home sponsor, regardless of the number and other types of providers sponsored, leading to a misclassification. Therefore, a priority rule was used to classify mixed sponsors using the information available in the data files

¹ Form FNS-44 can be found at: <https://fns-prod.azureedge.net/sites/default/files/FNS-44.pdf>

provided by the States and ensured that there was at least a 10 percent chance of selecting a family day care home provider from a family day care home sponsor. The priority rule is hierarchical as follows: (1) any sponsor with at least one At-Risk center was classified as an At-Risk sponsor, and it was further classified as a Mixed At-Risk sponsor if it had at least one non At-Risk center; (2) if a sponsor was not classified as At-Risk and administered family day care homes that accounted for 10 percent or more of all providers under its aegis, it was classified as an family day care home sponsor; (3) among the sponsors not classified as At-Risk or family day care home, if a sponsor had 10 percent or more Head Start centers, it was classified as a Head Start sponsor; (5) all other sponsors were treated as center sponsors²family day care home; (6) independent child care centers (independent centers) are classified into two strata, regular independent centers and At-Risk independent centers.³

3. Unavailable supplementary data. In addition to the minimum information needed to select representative samples of CACFP sponsors and providers in each State, the initial data request to the States included several additional pieces of information about each sponsor and provider in the State (e.g., for CACFP providers, this included: number of children enrolled, licensure status, racial and ethnic mix of children enrolled, and tiering status of family day care homes). These supplementary data were intended to be used to conduct a more robust non-response bias analysis than would otherwise be possible. However, nearly all States reported that they did not have one or more of requested variables in their database. Some States reported that requested variables might be available in other State databases and, in some cases, data sets were merged. Unfortunately, these data were not consistently available across the sample States.
4. State file issues. The data files received from the States to constitute the sponsor and provider frames varied widely in quality and structure. The number of files received per State varied, but the States generally provided separate files for sponsor- and provider-level (center and home) information. Many States provided more than one file for each level (sponsor or provider), and the files were combined to obtain sufficient information to create the frame. However, approximately half of the initial files submitted did not have sufficient information to create the frames. At the time of selecting the sponsors' samples, many issues remained unresolved. Exhibit 1.1 provides the final determinations for the outstanding issues at the time of sponsor selection.

Exhibit 1.1: Solutions to Outstanding Issues at the Time of Sample Selection	
Issue	Solution
Four States did not provide a complete provider list at the time of sponsor selection.	The information on the sponsor file (e.g. number of providers by type) was used to select the sponsor sample. Then, the provider lists were requested for the selected sponsors.
Two sponsors in one State did not provide the type of their centers for the sponsors with mixed center types.	Based on the sponsor type, all the centers from one sponsor were classified as At-Risk and the centers from the second sponsor were classified as Head Start.
Seven States did not provide a designation of their independent or self-sponsored centers.	The sponsoring organizations were classified with only one center as independent or self-sponsored.

² This group may include some family day care homes or Head Start centers.

³ In Volume I, the At-Risk centers are not classified as independent and sponsored, but as public and private. This classification was used for sampling purposes only.

Exhibit 1.1: Solutions to Outstanding Issues at the Time of Sample Selection	
Issue	Solution
In one State, some centers were defined as Early Head Start.	These centers were classified as Head Start for sampling purposes.
Only ten States provided a reliable Center ID.	The combination of center name and sponsor ID was used for the 13 other States as the center identifier.
Some States had duplicate centers.	When the Center ID was not reliable, the address and the phone number were used as criteria to identify duplicates.

SAMPLING DESIGN

MAIN STUDY COMPONENT

The main study component used a three-stage sampling design for selected sponsored providers and their sponsors. The first stage included a nationally representative sample of States. The second stage sample is a sample of sponsors from each of the 23 sample States. The State sponsor sample is representative of all sponsors in the State. Similarly, the third stage sample of providers is a sample of providers representative of all providers in the second stage sponsor sample. Because each stage is nested within the previous stage, the total resultant sample at each stage is nationally representative.

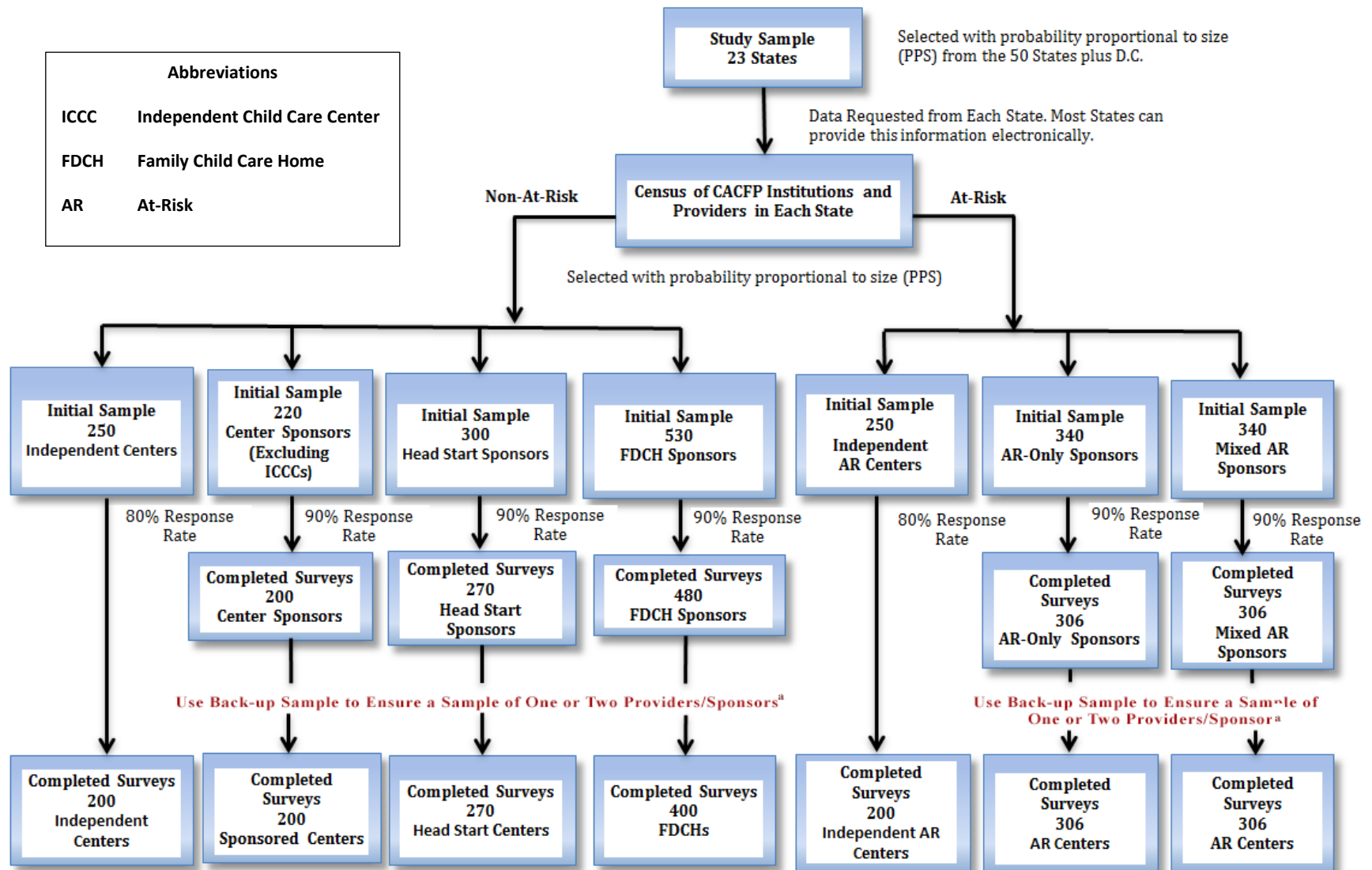
A two-stage sampling design was used to select the sample of independent centers. The first stage was the same sample of States as the sponsored child care centers (sponsored centers) sample. However, since independent centers do not have a separate sponsoring organization by definition, a representative sample of independent centers was selected from within each sampled State in the second stage instead of sponsors.

AT-RISK COMPONENT

A three-stage sampling design was also used to select a nationally representative sample of after-school At-Risk centers and their sponsors. The first stage included the same first-stage sample of States as the main study component. However, because the At-Risk component focuses only on questions related to sponsoring and operating At-Risk centers, separate second- and third-stage samples were selected for this component. The second-stage sample frame included only CACFP sponsors identified on State lists as sponsoring At-Risk centers. For the third-stage sample, At-Risk centers from the second-stage sample of At-Risk center sponsors were selected.

The initial sample sizes were determined by the final sample sizes (after non-response) needed to meet the requirements for levels of precision and confidence intervals ($\pm 5\%$ for estimates for the total population of sponsors and providers; and $\pm 10\%$ for sub-groups estimates, i.e., estimates for each type of sponsor and each type of provider). Exhibit 1.2 summarizes the sample design for the study including the initial sample sizes, expected response rates, and the expected sizes of the final analytic sample.

Exhibit 1.2: Overview of Sampling Design



^a Two family day care homes were selected for each of the 20 largest family day care home sponsors in the sample.

SAMPLE SELECTION AND CHARACTERISTICS OF THE INITIAL SAMPLE

A stratified probability sample ensures that the study could meet the precision requirements for each type of CACFP sponsor and CACFP provider. Within each first-stage sample State, sponsor lists were stratified by sponsor type, and second-stage probability proportional to size (PPS) samples of sponsors and independent centers were selected. At the third-stage, a PPS sample of providers was selected from within second-stage sample sponsor. The total number of meals served by family day care homes and centers was used as the measure of size. Exhibit 1.3 presents the characteristics of the initial sponsor sample and Exhibit 1.4 presents the characteristics of the initial provider sample.

Exhibit 1.3: Distribution of the Initial CACFP Sponsor^a Sample					
State	Initial Sample Size				
	Sponsor Type				
	Sponsored centers	Head Start	Family day care home	At-Risk	Total Sponsors
Arkansas	10	11	12	43	76
Arizona	9	12	16	11	48
California	10	6	25	84	125
Florida	8	10	21	25	64
Georgia	9	10	22	37	78
Iowa	10	14	22	8	54
Illinois	9	10	12	37	68
Indiana	8	14	9	47	78
Kansas	9	11	18	27	65
Kentucky	10	11	8	28	57
Massachusetts	10	8	19	49	86
Missouri	9	14	10	37	70
North Carolina	9	13	7	31	60
North Dakota	10	5	5	14	34
New Jersey	18	14	12	32	76
Nevada	3	9	3	15	30
New York	10	9	79	52	150
Ohio	9	7	0	58	74
Pennsylvania	10	4	8	54	76
South Dakota	7	0	3	18	28
Texas	10	12	0	43	108
Utah	9	9	6	7	31
Washington	10	12	11	27	60
Total	216	225	328	827	1,596

^a Independent child care centers are self-sponsored centers. That is, the sponsor and the center are the same. Independent centers are included with other center sponsors in this exhibit.

Exhibit 1.4: Distribution of CACFP Provider Sample and Completed Surveys by Sampled State						
State	Initial Sample Size					
	Provider Type					
	Independent center	Sponsored child care center	Head Start Center	Family day care home	At-Risk Center	Total Providers
Arkansas	6	10	14	12	53	95
Arizona	8	9	14	22	11	64
California	9	10	15	57	71	162
Florida	31	9	14	21	20	95
Georgia	9	9	14	23	57	112
Iowa	12	10	14	23	11	70
Illinois	11	9	14	23	42	99
Indiana	7	8	14	16	51	96
Kansas	7	9	14	25	43	98
Kentucky	20	10	14	9	34	87
Massachusetts	5	10	14	32	57	118
Missouri	20	9	14	11	48	102
North Carolina	9	9	14	13	26	71
North Dakota	8	10	5	21	27	71
New Jersey	7	18	13	12	42	92
Nevada	2	2	9	6	23	42
New York	11	10	15	82	110	228
Ohio	13	9	14	12	44	92
Pennsylvania	12	10	14	13	60	109
South Dakota	8	10	8	6	10	42
Texas	14	10	14	45	51	134
Utah	7	9	10	17	6	49
Washington	14	10	14	15	29	82
Total	250	219	299	516	926	2210

Chapter 2: Data Collection

LETTER TO REGIONAL OFFICES AND STATE CACFP ADMINISTERING AGENCIES

Prior to the start of data collection, FNS sent letters to Regional Child Nutrition Program Directors. The letter included information about the objectives of the study, the data collection procedures, and the role of the regional offices in the study. In addition to FNS contact information, the letters included an email address and telephone number for a Kokopelli Associates staff contact whom the regional offices could call or email if they require clarification or additional information. These letters were signed by the director of the FNS Child Nutrition program.

Letters to the State Child Nutrition Program Directors were sent by Kokopelli Associates. This letter requested submission of the following information:

1. For each CACFP sponsor in each State as of November 1, 2014:
 - Organization name, Identification Number, address, telephone number, contact person, email address, and fax number;
 - Sponsor type (sponsors of child care centers, Head Start centers, independent centers, and sponsors of family day care homes); and
 - Sponsor size (number of centers and family day care homes sponsored, total number of meals claimed for March 2014).
2. For each CACFP provider in each State as of November 1, 2014:
 - Organization name, address, telephone number, contact person, email address, and fax number);
 - Indicator of center type (child care, independent child care center, Head Start/Early Head Start, family day care home);
 - family day care home type (family day care home, group day care home, or License Exempt Provider);
 - Approval type (by licensing agency);
 - For each center percentage of meals reimbursed at free, reduced-price, and paid rates;
 - For family day care homes, tiering status (Tier I, Tier II, Tier II-mixed);
 - Provider size (meals claimed and number of participants);
 - Percentage of enrollment⁴ that is: Hispanic or Latino, American Indian or Alaskan Native, Asian, Black or African American, native Hawaiian or Other Pacific Islander, White; and
 - CACFP sponsor name and Identification Number to link each child care site to the sponsor.

⁴ This demographic data was not used in the analysis.

DATA COLLECTION PACKAGE

The study team developed questionnaires targeting sponsors and providers. To reduce respondent burden, individual questionnaires were designed for sponsors of child care centers, Head Start centers, family day care homes, and At-Risk centers. The study team also designed a questionnaire for organizations that sponsor more than one type of provider. Similarly, individual provider questionnaires were designed for independent child care centers, sponsored child care centers, Head Start centers, family day care homes, and At-Risk after school programs. Hardcopy, web, and computer-assisted telephone interviewing (CATI) versions were created for each instrument. The study used a single integrated database for all sponsor and provider surveys, regardless of the mode used to complete the survey.

Data collection began on May 21, 2015. All sampled sponsors and providers received a data collection package via Federal Express with their initial invitation to participate in the study.

The data collection package contained:

- An introductory letter from Kokopelli Associates;
- A customized brochure that:
 - cited the FNS auspices for the study,
 - provided an introduction to the study, including objectives,
 - described the importance of their participation,
 - provided instructions for completing the survey, including a URL and personalized, password for Internet access,
 - assured confidentiality of responses, and
 - provided a toll-free help line number and email address.
- Endorsement letters from the National CACFP Sponsors Association and the CACFP National Forum, and
- A targeted sponsor or provider questionnaire with return envelope.

Approximately two weeks after the initial mailing, a second data collection package was mailed to non-respondents. All sponsors and providers with an email address on the sample file were emailed a reminder with a link to the survey URL. Two weeks after the follow-up mailing, bilingual telephone data collectors began to call non-respondents in an attempt to complete the survey over the phone. After 20 unsuccessful call attempts, a replacement provider or sponsor was selected, and the survey process was initiated for the replacement sponsor or provider.

This multi-mode approach enabled respondents to complete the survey using their preferred method. The initial use of self-administered modes (web and mail) provided cost efficiencies by eliminating the need for a telephone interviewer to administer the survey for all respondents. A total of 1,097 respondents completed the survey through the web-link; 1,512 completed the mail; survey; and 212 completed a telephone interview.

To obtain several additional data elements that were not included in the sponsor and provider survey instruments, survey participants were re-contacted for a Follow-Up Survey. These data elements allowed FNS to answer the following research questions (RQs):

1. What are the tiering rates for family day care homes? What is the category eligibility distribution by percentage?
2. For family day care homes: What percentage of children enrolled in CACFP Tier II family day care homes receive meals at the Tier I rates? and
3. What are the counts and percentages of participation in the CACFP of centers by meal reimbursement category (free, reduced-price, and paid)?

FOLLOW-UP SURVEY

This Follow-Up Survey⁵ was conducted using two attempts, or rounds, to obtain information on the proportion of children in child care centers and in the regular CACFP component of Mixed At-Risk centers whose meals are reimbursed at each of the three income-eligibility categories: free, reduced-price, and paid. The first round of the Follow-Up Survey, conducted from February 6, 2017 through March 1, 2017, attempted to contact participating sponsors and providers directly to obtain the additional data. As the response rates for the first round of the Follow-up Survey were quite low, an additional round of data collection was conducted to obtain data from centers that did not respond to the first round of the Follow-Up Survey. The second round of the Follow-Up Survey was conducted from July 10, 2017 through August 28, 2017. Rather than simply repeat the first round of the Follow-Up Survey, survey procedures were modified for the second round in two important respects:

1. With assistance from the FNS Regional Offices, the second data retrieval effort enlisted the support and cooperation of the State CACFP administering agencies in each of the 23 States.
2. Study team senior staff, including former State CACFP administrators, were responsible for all contacts with the State agencies.

These two procedural changes greatly increased the cooperation received and the total results for the Follow-Up Survey.

MIXED AT-RISK CENTERS

Eleven of the 23 study states had at least one Mixed At-Risk center for which income eligibility information was needed. A total of 103 Mixed At-Risk centers were included in the Follow-Up Survey. Three of the 11 States (New York, Ohio and New Jersey) indicated that they could not provide us with the requested income eligibility data for any of their Mixed At-Risk centers. At the outset of the second round of the Follow-Up Survey, income eligibility data was obtained directly from the 13 Mixed At-Risk centers in New York (2), Ohio (5), and New Jersey (6). No responses could be obtained for the Mixed At-Risk Centers in Ohio. Researchers then followed up with the CACFP administering

⁵ The Follow-Up Survey instrument can be found in Volume IV.

agencies in the remaining 8 States. Exhibit 2.1 presents the results of the Follow-Up Survey and the additional data retrieval effort for Mixed At-Risk centers.

Exhibit 2.1: Results of the Follow-Up Survey of Mixed At-Risk Centers			
Sample Disposition	First Round of Follow-Up Survey	Second Round of Follow-Up Survey	Total Follow-Up Survey Data Collection Effort
Initial Sample Size	201	104 ^b	201
Ineligible for the Survey	0	11	11
Eligible	201	93	190
Refusal	25	0	0
No Response	79	12	12
Completed Survey	97	81	178
Response Rate^a	48.2%	87.1% ^b	93.7%

^a The response rate was calculated using the formula: $R / \{R + N + e(U)\}$, where R is the number of respondents, N is the number of eligible nonrespondents, and $e(U)$ is the estimated number of eligible nonrespondents among the eligible unknown cases. In this study, there were no eligible unknown cases. Thus, $e(U) = 0$. Note that N in the formula represents the three terms, R , NC , and O in the OMB formula. Therefore, this formula is essentially the same as the formula provided on page 14 of the *OMB Standards and Guidelines for Statistical Surveys* (September 2006).

^b Both refusal and no response cases were included in the second round of the Follow-Up Survey.

At the conclusion of the first round of the Follow-Up Survey, no data had been obtained for 104 of the 201 eligible Mixed At-Risk centers. The sample for the second round of the Follow-Up Survey was comprised of the 104 non-responsive Mixed At-Risk centers, as well as those that refused. Six States informed the study team that one or more of their centers were no longer Mixed At-Risk centers, leaving 93 eligible and 11 ineligible Mixed At-Risk centers. For 12 centers in the second round of the Follow-Up Survey, no information could be obtained from the State or from directly contacting the center, due to inaccurate or unavailable contact information. Responses were obtained for 81 of the 93 Mixed At-Risk centers, a response rate of 87.1 percent. When the results from the two Follow-Up Surveys are combined, the study team obtained data for 178 of the 190 eligible Mixed At-Risk centers, a response rate of 93.7 percent.

REGULAR CHILD CARE CENTERS

Each State had at least 1 regular child care center in the first round of the Follow-Up Survey, for a total of 344 child care centers. Exhibit 2.2 shows that at the time of the first round of the Follow-Up Survey, 3 child care centers were no longer participating in the CACFP and were ineligible for the survey, leaving 341 eligible centers. Data were obtained for 99 centers; 2 centers refused to participate in the survey, and the study team was unable to contact 240 centers. The response rate for the first round of the Follow-Up Survey was 29.0 percent.

Exhibit 2.2: Results of the Follow-Up Survey of Child Care Centers			
Sample Disposition	First Round of the Follow-Up Survey	Second Round of the Follow-Up Survey	Total Follow-Up Survey Data Collection Effort
Initial Sample Size	344	240 ^b	344
Ineligible for the Survey	3	24	27
Eligible	341	216	317
Refusal	2	0	2
No Response	240	23	23
Completed Survey	99	193	292
Response Rate^a	29.0%	89.4% ^b	92.1%

^a The response rate was calculated using the formula: $R / \{R + N + e(U)\}$, where R is the number of respondents, N is the number of eligible nonrespondents, and $e(U)$ is the estimated number of eligible nonrespondents among the eligible unknown cases. In this study, there were no eligible unknown cases. Thus, $e(U) = 0$. Note that N in the formula represents the three terms, R , NC , and O in the OMB formula. Therefore, this formula is essentially the same as the formula provided on page 14 of the *OMB Standards and Guidelines for Statistical Surveys* (September 2006).

^b Both refusal and no response cases were included in the second round of the Follow-Up Survey.

The 240 child care centers that did not respond to the first round of the Follow-Up Survey formed the initial sample for the second round of data collection. The original protocol for the second round of the Follow-Up Survey called for staff to directly contact these 240 child care centers to try to obtain the requisite income eligibility data. As soon as the State notified the child care centers of their inclusion in the Follow-Up Survey, the study team attempted to contact the child care centers to collect the data.

Many States informed the study team did not need to contact the child care centers directly as the State could provide the data for the child care centers. The study team received the income eligibility data for the Follow-Up Survey child care centers in all States that indicated they could do so. Of the initial 240 child care centers, 24 were found to be ineligible for the Follow-Up Survey,⁶ leaving 216 eligible child care centers. Data were obtained for a total of 193 child care centers; 23 child care centers did not respond. The completion rate for this part of the additional survey data retrieval was 89.4 percent. By combining the results from the two rounds of the Follow-Up Survey, the response rate is calculated at 92.1 percent (292 of the 317 eligible child care centers).

DATA RECEIPT AND FOR SURVEYS COMPLETED BY MAIL DATA COLLECTION

As described, participants in the study were offered three options for completing the survey: mail, web, or telephone. For the forms received by mail, data management staff reviewed each completed form to assess the quality and completeness of the survey responses and to organize the materials for electronic receipt. Receipt staff electronically recorded the receipt status, and forwarded the forms to scanning operations for data capture. TeleForm software system was used to process the paper instruments. This software system scanned and extracted responses, validated data capture

⁶ These centers were no longer in operation.

by identifying items which were not scan-readable and needed to be verified and entered by data entry staff, and stored data for each survey in its own central data depository.

Staff incorporated the data edits within the automated survey administration application (web and phone) as the survey was completed by the participant. The study team conducted a manual and computerized review of surveys completed via these modes to ensure that all responses were being captured in the sponsor and provider databases. Manual review of the data entered into the automated system was identical to that of the mail data. Once data collection was complete, the mail data was merged the web and telephone data.

DATA COLLECTION ISSUES

To obtain the minimum information needed for sampling, the study team contacted the States who were unable to provide complete information to obtain the requisite information. The information requested from the State Child Nutrition Program Directors proved to be burdensome for some State officials to provide.

Data management staff reviewed variable frequencies, conducted range and logic checks, tracked and resolved problem cases, and verified data edits. They completed a manual review of the frequency distribution of collected data items to identify errant skip pattern and other anomalies. In instances where the participant answered a question that should have been skipped, the answer was deleted and coded as missing. If a respondent entered an out of range value, the response was recoded to '999' (the maximum number of digits allowed) and entered into the decision log.

In multiple instances, participants replied using two modes. For those cases, the record with the most complete data was maintained and the duplicate survey data removed. A log was maintained to document all problems and resolutions. Respondent issues were resolved either through participant contact by project staff or by data collectors, depending upon the nature of the problem.

FINAL DISTRIBUTION OF THE SAMPLE

INITIAL SAMPLE

Exhibits 2.3 and 2.4 show the distributions of the samples of completed sponsor and provider surveys. These constitute the samples that will be used in the analysis.

Exhibit 2.3: Sponsor Analytic Sample: Distribution of Completed CACFP Sponsor¹ Surveys by Sampled State					
State	Number of Completed Sponsor Surveys				
	Sponsor Type (by provider type)				
	Sponsored child care center	Head Start center	Family day care home	At-Risk center	TOTAL SPONSORS
Arkansas	10	8	12	25	55
Arizona	8	11	12	11	42
California	9	5	24	78	116
Florida	7	9	19	19	54
Georgia	8	10	19	34	71
Iowa	10	14	22	8	54
Illinois	7	8	11	29	55
Indiana	7	12	8	43	70
Kansas	9	10	18	26	63
Kentucky	9	11	5	26	51
Massachusetts	8	7	19	42	76
Missouri	6	13	9	31	59
North Carolina	8	12	7	28	55
North Dakota	7	5	5	12	29
New Jersey	11	11	10	26	58
Nevada	2	8	3	14	27
New York	7	9	77	47	140
Ohio	5	4	0	31	40
Pennsylvania	8	4	6	46	64
South Dakota	6	0	3	17	26
Texas	7	12	0	66	85
Utah	7	7	6	7	27
Washington	9	12	11	26	58
Total	175	202	306	692	1,375

¹ Independent child care centers are self-sponsored centers. That is, the sponsor and the center are one and the same. Independent centers are included with providers in Exhibit 3.7.

Exhibit 2.4: Provider Analytic Sample: Distribution of Completed CACFP Provider Surveys by Sampled State						
State	Number of Completed Provider Surveys					
	Provider Type					
	Independent center¹	Sponsored center	Head Start center	Family day care home	At-Risk center	TOTAL PROVIDERS
Arkansas	6	8	4	9	30	57
Arizona	2	5	8	9	4	28
California	5	9	10	32	41	97
Florida	24	4	20	12	8	58
Georgia	7	6	11	18	42	84
Iowa	8	3	4	12	1	28
Illinois	10	5	11	10	27	63
Indiana	6	8	8	5	43	70
Kansas	7	8	11	10	38	74
Kentucky	16	8	13	2	25	63
Massachusetts	5	9	12	15	44	85
Missouri	17	7	13	3	38	78
North Carolina	8	9	13	7	14	51
North Dakota	8	8	5	13	19	53
New Jersey	7	4	3	9	20	43
Nevada	2	1	8	4	17	32
New York	10	9	15	64	85	183
Ohio	5	3	5	6	8	27
Pennsylvania	9	8	9	12	44	82
South Dakota	8	7	2	10	7	34
Texas	13	5	8	25	28	79
Utah	6	3	3	12	2	26
Washington	12	5	6	9	19	51
Total	201	142	191	308	604	1,446

¹ Independent child care centers are self-sponsored centers. That is, the sponsor and the center are one and the same.

Chapter 3: Analytic Methodology

The analyses conducted to address the study's research questions included national estimates of means, medians, and standard deviations for continuous variables, and frequency distributions and cross-tabulations for categorical variables. This chapter discusses calculated standard errors; weighting and treatment of missing values; item nonresponse; and the nonresponse bias analysis.

CALCULATING STANDARD ERRORS

The procedure for calculating standard errors used by standard analysis software assumes the sample was selected using simple random sampling (SRS) in which individual sampling units are selected at random with equal probability and without replacement (WOR) directly from the entire population. Thus, statistical packages, like SPSS, may be used to calculate different measures of central tendency and to obtain values related to statistical. By contrast, the complex sample for this study utilized multiple stages and had unequal selection probabilities using probability-proportional-to-size (PPS) sampling. Because of the multi-stage design and PPS sampling, these estimates would have been biased had the team had coded SPSS to calculate them. Instead, the research team used SPSS Complex Samples 22.0 to calculate all measures, including standard errors. SPSS Complex Samples is a module that accounts for complex (stratified/clustered) sampling designs, correctly calculating standard errors with weighted data. This option allowed the data analyst to select a sample according to its design and incorporate the design specifications into the data analysis, ensuring valid results.

WEIGHTING AND TREATMENT OF MISSING VALUES BASE WEIGHTS

Within each State, the sponsors were classified into 7 strata: 1) family day care home sponsors; 2) Head Start center sponsors; 3) sponsors of child care centers; 4) sponsors of only At-Risk centers; 5) Mixed At-Risk sponsors; 6) independent centers; and 7) At-Risk independent centers.

In the first five strata, the sponsors were selected using the PPS method. For strata 1, 2, 3, and 4, the measure of size (MOS) was the square root of the number of providers. However, for stratum 5 the MOS was the square root of the number of At-Risk centers only. In strata 6 and 7 (independent centers and At-Risk independent centers), the MOS was equal to 1, so simple random sampling (SRS) was used. Given their size, some sponsors included in the sample with certainty. All of the 422 eligible family day care home sponsors in the frame were included in the sample.

Each provider was classified in the same stratum as its sponsor, resulting in the same 7 strata described above. In strata 2 to 5, one provider was selected from each selected sponsor except for some large sponsors. In those cases, more than one provider was selected to reflect the expected number of times they would be sampled (i.e., multiple hits during PPS sampling). Given the large

variation among sponsors in the family day care homes sponsored, two providers were selected from large sponsors to reduce the variation in the sampling weights. The ineligible providers were replaced following the randomized list of providers to ensure that the number of eligible units remained as close as possible to the provider sample size.

The provider base weight was obtained by multiplying the sponsor base weight to the conditional provider base weight, which was equal to the number of providers divided by the number of providers for a given sponsor. In the At-Risk strata 4 and 5, some provider base weights were extreme within the weight distribution. In a few small States, all available At-Risk sponsors were included in the sample with certainty and, despite the large difference in their sizes, one provider was selected for each of them. Extreme weights can inflate drastically the variance of the estimators and, therefore, reduce the analytical capacity of the data. Given the effect of the extreme weights on the estimators' precision, the weights in strata 4 and 5 were trimmed. All weights larger than the 95th percentile were reduced (top coded or capped) to its corresponding value. A total of 23 eligible cases of 587 providers (3.9 percent) were trimmed. Twelve providers from the trimmed group responded to the study.

Eleven sponsors were selected but could not be reached. These sponsors were included in the weighting and treated as non-respondents.

NONRESPONSE ADJUSTED WEIGHTS

Weights were adjusted by creating nonresponse cells (or groups) and re-weighting the respondents within each cell. Nonresponse cells were formed within each State by crossing the stratum by the self-representing status of the sponsors. These adjustment cells ensured that the weight from the nonrespondent certainty sponsors was redistributed to other certainty sponsors from the same State. In three instances, cells with an adjustment factor larger than 2.5 were combined. Another instance required combining cells from two different non-certainty States. The replicate sponsor weights were adjusted in the same way as the full sample weight.

For the provider sample, the adjustments for nonresponse were performed independently because the provider response propensity was different from the corresponding sponsor response propensity. Because some providers responded when sponsors did not do so, the sponsor nonresponse adjusted weight was not used, as it would give a weight of zero to some responding providers. Hence, the provider nonresponse adjustment began with the provider base weight. By forming nine groups (eight certainty States and a combined group of the non-certainty States), the study team was able to create the provider nonresponse cells. Those nine cells were further subdivided by the sponsor type stratum and the self-representing status of the sponsors. To adjust for sponsor nonresponse, cells were combined in three instances.

POST-STRATIFIED FINAL WEIGHTS

For the nonresponse adjusted weights, the post-stratification was conducted independently for the sponsors and the providers. The control totals were obtained from the FNS Data Bank for October and December 2014. The FNS Data Bank calculates the population counts for the family day care home sponsors and homes and child care center sponsors and child care centers including independent centers (without breakdowns by type) on a quarterly basis (March, June, September, and December). The FNS Data Bank includes the detailed population counts for child care sponsors and providers by type on a semi-annual basis (March and October). However, the sponsor counts are not mutually exclusive; a sponsor with multiple types of providers is included in multiple categories. Independent centers are included in both sponsor and provider counts.

The sponsors were classified into two post-stratification domains: family day care homes and sponsored centers. The family day care home domain was formed of all the organizations that sponsored at least one family day care home. All the other sponsors including the independent (both At-Risk and not At-Risk) were classified as sponsored centers for post-stratification. Detailed center counts by sponsor type were unavailable because the defined sponsor types are not mutually exclusive; sponsors with mixed providers are double-counted. Furthermore, all independent centers were included in the sponsor and provider counts in the FNS Data Bank.

Exhibit 3.1 provides the control totals used for the sponsor data. The control totals are the number of sponsors in a given type in the entire population based on the FNS Data Bank for September and December 2014. The family day care home control total was solely based on the December 2014 data but the sponsored center control total used the prorated September and December 2014 counts to calculate an estimated count for the survey reference date October 2014.⁷

Exhibit 3.1: Post-Stratification Domains and Control Totals for the Sponsor Data			
Sponsor Post-Stratification Domain	Data	Control Totals	Post-Stratification Adjustment Factors
Family day care home	December 2014	812	1.10
Sponsored center	September & December 2014	20,251	1.06
Total		21,063	

The post-stratification consisted of adjusting the sponsor nonresponse adjusted weights so that they sum to the control totals shown in Exhibit 3.1. To post-stratify the provider data, the study team used the following four domains: family day care homes, At-Risk centers, Head Start centers, and sponsored centers. The family day care home domain included all family day care homes. The At-Risk

⁷ The study team did not use proration for the FDCH count because the September 2014 count was unusually higher than other quarterly counts, and the December 2014 count was much closer to the average count over time. Specifically, the average count during October 2011 through September 2012 was 848. The average count during October 2012 to September 2013 was 855. Similarly, the December 2014 count was 812 excluding FDCHs in territories, whereas the September 2014 count was 1,065.

category⁸ included all providers identified as At-Risk including the independent ones. The Head Start category was formed by the Head Start centers. The remaining providers including independent centers were classified as sponsored centers. The post-stratification adjustment factors were moderate for the provider data, ranging from 1.17 to 1.38, as illustrated in Exhibit 3.2.⁹

Exhibit 3.2: Post-Stratification Domains and Control Totals for the Provider Data			
Provider Post-Stratification Domain	Data	Control Totals^a	Post-Stratification Adjustment Factors
Family day care home	December 2014	115,708	1.28
At-Risk center	October 2014	16,936	1.18
Head Start center	October 2014	11,497	1.18
Sponsored center	October & December 2014	32,915	1.38
Total		177,056	

^aSource: FNS National Data Bank, March 2015.

ANALYTIC SAMPLE

In Chapter 2, Exhibits 2.3 and 2.4 show the distributions of the samples of completed sponsor and provider surveys. These constitute the samples used in the analysis.

ITEM NONRESPONSE TREATMENT OF MISSING VALUES

Response rates for all items were calculated as part of the assessment of the survey data's quality. To make the response rate calculations accurate, "-1" is coded to indicate a skip, thus removing all legitimate item nonresponse. For "check all that apply" variables, a summary value was created so as not to count non-selected items within a series of response choices. Therefore, the research team assigned a "1" if a response was selected and a "0" otherwise. Then, the non-negative values were summed within each question and calculated the response rate on the summed value.

An item response rate of 60 percent was used as the benchmark for assessing the potential for nonresponse bias. Exhibit 3.3 shows the frequency distributions of the sponsor and provider survey items by item response rates. The clear majority of sponsor survey item response rates were quite high: 139 out of 170 survey items had response rates of at least 90 percent. Most items in the provider survey had response rates of at least 60 percent and 132 out of 169 items had response rates of at least 90 percent.

⁸ As discussed under Sample Selection, the sample for the main study component includes a representative sample of all centers including At-Risk centers. However, the number of At-Risk centers in this sample was too small to develop reliable estimates of the characteristics of At-Risk centers. Because of this, a separate nationally-representative sample of At-Risk centers was selected for the At-Risk study component.

⁹ Due to lack of separate counts of independent centers, the population size of independent centers may not be reliably estimated even with post-stratification. This would also affect the population size estimate for sponsored child care centers.

Exhibit 3.3: Distribution of Survey Items by Item Response Rate: Sponsor and Provider Surveys		
Item Response Rate	Number of Survey Items	
	Sponsor Survey	Provider Survey
0-49%	0	2
50-59%	0	0
60-69%	2	3
70-79%	7	10
80-89%	22	23
90-99%	130	125
100%	9	7
Total	170	169

There are two items in the provider survey items with a response rate below 60 percent.

- 1. What is the usual number of children per adult for groups of 3 to 5-year-olds served during weekends or evenings at this site?** At 44.8 percent, this was the only variable when looking at “all providers” that fell below the threshold. An earlier branch question asked:

Is the number of children per adult different during weekends or evenings that your child care site is in operation?

- If the answer is “Yes,” then participants were asked: What is the number of children per adult different during weekends or evenings that your child care site is in operation?
- If the answer is “No,” then participants were not asked the above follow-up question. Sixteen participants skipped over the follow-up question, leaving the answer to the follow-up question blank; those blank responses were coded as missing.

However, those missing cases are still included in the denominator for response rate calculations (response rate = [(13 responses) / (13 responses + 16 missing cases)] = 13/29 = 44.8%). Similar situations occurred with this variable when looking at only independent centers (response rate = 50.0%) and at child care centers (response rate = 40.0%).

- 2. On what basis does your <provider type> offer these discounts?** This is a “mark all that apply” question. There is a prior branch question which asks:

Do you offer a child care discount to families who pay for care? If the answer is “Yes,” then participants were directed to **(2): On what basis does your <provider type> offer these discounts?** When looking at all providers, the item response rate for **(2)** is 70.4 percent. However, when reviewing the item response rate for the sub-group family day care home providers, the item response rate for **(2)** is 56.9 percent. The explanation is that 71 family day care home providers had missing data for the screening question about offering discounts; those responses were coded as missing, and those participants were not asked question **(2)**. However, those missing cases are still included in the denominator for response rate calculations.

NONRESPONSE BIAS ANALYSIS

A nonresponse bias analysis was conducted for all provider types with response rates less than 80 percent. If the respondents and non-respondents were similar with respect to the survey variables, the magnitude of any potential nonresponse bias is likely immaterial. However, data were not present for the survey variables for the non-respondents. As a result, the study team chose to use auxiliary variables from other sources. A finding of no significant differences in the auxiliary is taken as an indication that the potential nonresponse bias is not serious. If the auxiliary variables are different between the respondents and non-respondents, it is an indication of potential bias. Exhibit 3.4 shows the auxiliary data that were available for each sampled State, though the data was not available for all States.

Exhibit 3.4: Auxiliary Data Availability on the Sponsor and Provider Frame Files						
State	From Sponsor file		From Provider file			
	Number of Centers and Family Day Care Homes	Total Number of Meals Claimed for March, 2014	Number of Children Enrolled	Licensure Status	Percent of Enrollment by Race/ Ethnicity	Family Day Care Home Tiering Status
AR	x	x			x	x
AZ	x	x				
CA	x			x		x
FL	x		x			x
GA	x	x			x	
IA	x					x
IL	x	x				
IN	x	x	x	x		x
KS	x	x		x		x
KY	x					
MA	x	x	x	x		x
MO	x					x
NC	x	x				
ND	x	x	x			x
NJ	x		x			x
NV	x				x	
NY	x		x	x	x	x
OH	x					
PA	x					x
SD	x	x	x			
TX	x	x				
UT	x	x			x	x
WA	x	x				

The License Status variable has two statuses: licensed and license-exempt. The race/ethnicity variables are given in percent of enrollment in each race or ethnicity. Small categories were combined, and we used the following scheme: Percent Hispanic, Percent Black, Percent White, and Percent Other Race(s). The latter three variables sum to 100 percent.

A small stratum was collapsed within a similar stratum. When necessary, Kokopelli collapsed Head Start and child care centers together and At-Risk with Mixed At-Risk.¹⁰ Exhibit 3.5 displays the results for continuous variables

Exhibit 3.5: Comparison of Respondents and Nonrespondents for Continuous Auxiliary Variables by Stratum									
Auxiliary Variable	Stratum¹	Sample Size	Respondent Mean	Respondent Mean Standard Error	Non-Respondent Mean	Non-Respondent Mean Standard Error	Estimated Difference	P-Value	Bonferroni Alpha
Number of Providers	1	455	491.8	72.6	413.5	58.6	78.3	0.3456	0.0100
	2	265	19.0	3.5	22.1	4.4	-3.1	0.5763	0.0100
	3	211	24.1	12.4	34.4	19.4	-10.2	0.6510	0.0100
	4	291	52.3	13.2	42.5	12.4	9.8	0.6129	0.0100
	5	296	225.2	84.0	338.7	107.7	-113.4	0.1579	0.0100
Meal Claim	1	228	74,397	27,127	45,727	13,780	28,670	0.2255	0.0100
	2	137	20,505	3,721	21,709	6,091	-1,204	0.8379	0.0100
	3	114	24,022	8,167	7,157	2,903	16,865	0.0468	0.0100
	4	144	49,624	27,233	20,078	5,576	29,546	0.3398	0.0100
	5	155	369,462	258,621	108,998	30,723	260,464	0.3145	0.0100
Number of Children	1	173	8.2	0.9	10.8	5.8	-2.6	0.6051	0.0167
	2,3	152	98.3	14.8	38.9	12.0	59.4	<0.0001	0.0167
	4,5	195	82.8	20.7	41.0	18.6	41.8	0.0759	0.0167
Percent Black	1	131	5.0	16.0	0.0	0.0	5.0	0.7555	0.0167
	2,3	91	45.7	15.5	17.9	11.3	27.8	0.1658	0.0167
	4,5	118	34.8	27.4	37.2	23.3	-2.4	0.8415	0.0167
Percent Hispanic	1	131	0.5	0.3	0.0	0.0	0.5	0.1676	0.0167
	2,3	91	12.8	4.1	13.7	8.6	-0.9	0.9054	0.0167
	4,5	118	10.3	6.9	10.1	9.7	0.2	0.9829	0.0167
Percent White	1	131	17.3	61.5	7.6	89.6	9.7	0.7585	0.0167
	2,3	91	45.0	14.3	55.2	11.6	-10.2	0.0723	0.0167
	4,5	118	23.8	14.7	17.8	12.0	6.0	0.6015	0.0167
Percent Other Race(s)	1	131	0.6	0.6	1.3	1.4	-0.7	0.4878	0.0167
	2	91	4.0	2.0	8.1	3.1	-4.0	0.0240	0.0167
	3	118	16.9	16.9	4.0	2.9	12.9	0.4442	0.0167

¹ Note: When two strata were collapsed, both stratum numbers are given.

and Exhibit 3.6 shows the results for categorical variables. Only one case showed a significant difference between respondents and non-respondents, for the auxiliary variable “number of children served by the provider” in the collapsed stratum. This significant difference was caused by a single

¹⁰ We used the threshold of 30 based on the rule of thumb, which is considered reasonable from our experience. So, if the number of respondents or non-respondents in a stratum was less than 30, the stratum was collapsed into another stratum following the collapsing pattern (strata 2 and 3, and 4 and 5). However, if the collapsed size of respondents or non-respondents was less than 30 but greater than 20, no further collapsing was done.

outlier in the respondent data.¹¹ When the outlier was excluded, the difference became barely non-significant with a p-value of 0.0178, which is greater than the Bonferroni alpha of 0.0167, indicating that the potential nonresponse bias was not serious.

Family day care homes are reimbursed based on their tiering status.¹² The tiering status variable has two categories: Tier I and Tier II. The base weights for the providers were used to compare the weighted estimates between the respondents and non-respondents, using the t-test for continuous variables, or a Chi-squared test for categorical variables, with a significance level of five percent. When the nonresponse bias is small, these estimates are comparable at the subpopulation level represented by the sampled states with available data (Exhibit 3.6) for a particular auxiliary variable.

Auxiliary Variable	Type Strat	Sample Size	Respondent Mean	Respondent Mean Standard Error	Non-respondent Mean	Non-Respondent Mean Standard Error	Chi-square	P-Value
Tier Status	1	328	76.0	3.7	72.3	10.4	0.3265	0.7829
	2	163	6.9	1.1	6.2	2.4		
	3	139	8.2	1.7	10.7	6.6		
	4	194	4.8	1.0	4.8	1.4		
	5	204	4.1	0.8	5.9	2.6		
License Status	1	190	79.1	2.8	85.1	7.6	1.2772	0.3648
	2,3	114	12.2	1.8	5.3	2.3		
	4,5	199	8.7	1.6	9.5	6.0		

The t-test was performed for each continuous auxiliary variable by provider type as a group, so the Bonferroni adjusted alpha value¹³ was used. In the one case where the auxiliary variable was a categorical variable, a Chi-square test was performed to test the association between the response status and the type strata for the categorical auxiliary variable in a two-way table. A non-significant test result meant that the distribution of the respondents over the type strata is similar with the distribution of the non-respondents over the type strata (Type Stratum: family day care home, Head Start center, child care center, At-Risk center, and Mixed At-Risk center).

¹¹ The outlier value was 1,048, which was the largest among the respondent values, and the second largest value was 276. Without the outlier, the mean estimate changed from 98.3 to 84.6.

¹² Tier II reimbursement rates are lower than Tier I reimbursement rates. Tier I homes are those located in low-income areas or homes in which the provider's household income is less than or equal to 185% of the federal poverty level.

¹³ It is necessary to adjust the significance criterion (α) when simultaneously testing multiple hypotheses. The Bonferroni correction sets the significance cut-off at α/n . For example, in the example above, with 20 tests and $\alpha = 0.05$, you'd only reject a null hypothesis if the p-value is less than 0.0025. It should be noted that the Bonferroni correction tends to be somewhat conservative.