Appendix A
Sampling and Weighting Procedures

The *Family Child Care Homes Legislative Changes Study* involved several surveys, including surveys of sponsors, current CACFP providers, former CACFP providers, and parents of children currently served by CACFP providers. For current Tier 2 and former CACFP providers, the study included an *operations survey* and a *menu survey.* A *meal observation* data collection was conducted in a random subsample of current Tier 2 homes. Most of the analyses presented in this report focus on nutritional aspects of meals offered as measured in the menu survey, with portion sizes imputed from the more limited meal observations. The sample design for these surveys and the weighting procedures used in the analysis are described below. The sampling and weighting for other surveys are discussed in other reports in this series. Appendix F, which examines the menu survey of former CACFP providers, describes the sample for that survey.

Sample

The sample universe for the *Family Child Care Homes Legislative Changes Study* consisted of family child care sponsors, family child care homes, and families of children participating in the CACFP. A nationally representative sample of 20 States was selected, with probability proportional to the size of each State's share of CACFP family child care home reimbursements.[1] All selected State agencies agreed to participate in the study and provided lists of the CACFP sponsors in their State. Sponsors were also selected within States with probability proportional to size, based on the number of homes sponsored.[2]

Each selected sponsor was asked for a list of the family child care homes sponsored, including three groups of homes: Tier 1 homes active (i.e., receiving CACFP reimbursement) in January 1998; Tier 2 homes active in January 1998; and all homes active in January 1997.[3] Sample frames for current Tier 1 and Tier 2 providers were defined to include all homes active in January 1998. Within each sponsor's list of homes in each tier, a random sample was drawn. The base number of providers to be selected from each sponsor's list was constant across sponsors within each tier (four for Tier 1, six for Tier 2); if the total on the sponsor's list was equal to or less than the base number, all were selected.[4]

---

[1] Four States were included with certainty (California, Michigan, Minnesota, and Texas).

[2] Sponsors were sampled with replacement, meaning that a sponsor could be selected more than once.

[3] Homes received tier designations only when tiering was implemented in July 1997.

[4] The number selected depended on the number of times the sponsor was selected (i.e., if the sponsor was selected twice, double the base number would be selected from the sponsor's list).

A sample of 300 sponsors was selected within the 20 States, comprising a representative sample of the 1,165 sponsors active in the country.[5] Of the selected sponsors, 289 supplied lists of current and former providers, for a response rate of 96.3 percent. From these lists, 1,134 Tier 2 providers were selected for menu survey participation, of whom 393 were found to be ineligible for study. The primary reasons for ineligibility were that the provider left the CACFP or changed tier status; a few providers were deceased by the time of the survey. The menu survey was completed by 542 Tier 2 providers, for an estimated response rate among eligibles of 73.1 percent.[6]

Of the 387 providers selected for meal observation (one-third of the Tier 2 sample), 159 were found to be ineligible. Providers were ineligible if they no longer participated in the CACFP, changed tier, were deceased, or cared only for infants less than 1 year old. The 97 participating providers who allowed meal observations represented an estimated response rate of 42.5 percent.

In multistage sampling, it is sometimes useful to consider the compound response rate, which is the product of the response rates at each sampling stage. In the present instance, the compound response rates are 70.4 percent for providers completing the menu survey and 40.9 percent for completed meal observations.

Weighting

For producing population-based estimates of means and proportions of characteristics relating to providers, each respondent provider received a sampling weight. These weights combined the basic weight reflecting the probability of selection of the provider with an adjustment for unit nonresponse. The resulting weighted data yield estimates for all providers in the population.

The overall provider weight was obtained as the product of the State weight, the conditional sponsor weight (adjusted for nonresponse), and the conditional provider weight (adjusted for nonresponse), which is based on the conditional probability of selecting a provider given that the sponsor and the State have been selected.

Basic Sponsor Weights

A preliminary first step in determining provider weights was calculation of *sponsor weights*. As described above, a sample of sponsors was selected in each of the 20 States selected in the first stage. Therefore, the overall probability of inclusion of a sponsor is the inclusion probability of the State in which the sponsor is located multiplied by the probability of including the sponsor in the sample, given that the State was selected.
Sponsor weights were computed as follows:

1. Let W_i represent the weight for the ith selected State. $i = 1, 2, 3, 4, ...19, 20$. $W_i = 1$ for States selected with certainty. For the 16 noncertainty States, W_i is the inverse of the State's probability of selection. The probability of selection was proportional to the size of the State, with total CACFP reimbursements for family child care homes being the measure of size.

---

[5] A total of 311 were selected, but 11 were not eligible because they had left the CACFP.

[6] This calculation assumes that all nonrespondents not known to be ineligible for participation in the study were eligible. If five providers who supplied fewer than 3 menus and were dropped from the analysis are included, the response rate increases to 73.3 percent.

2. Let W_{ij} be the weight for the *j*th selected sponsor in the *i*th State. We have

$$W_{ij} = W_i \; W_{j/i}$$

where $W_{j/i}$ is the conditional weight of the *j*th sponsor given that the *i*th State has been selected.

We now determine $W_{j/i}$. Let the number of sponsors in the *i*th State be S_i. Let the number selected in the sample be s_i. Let the number of providers belonging to the *j*th sponsor in the *i*th State be P_{ij}.

- In 12 States, all sponsors in the State were included in the sample with certainty. In these States, we have

$$W_{j/i} = 1.$$

Therefore, the overall sponsor weight in these States is $W_{ij} = W_i$.

- The sponsors in the other eight States were selected with probability proportional to the number of providers and **with replacement**. Therefore, the same sponsor can get selected more than once. Let r_{ij} be the number of times ("hits") the *j*th sponsor gets selected in the *i*th State. The conditional weight for these sponsors is

$$W_{j/i} = \frac{r_{ij} \; P_i}{n_i \; P_{ij}}$$

where n_i is the total number of sponsor hits in the *i*th State and $P_i = \sum_{j=1}^{S_i} P_{ij}$ is the total number of providers.

The overall basic sampling weight for the *j*th sponsor in the *i*th State is given by:

$$W_{ij} = W_i \; W_{j/i}.$$

Adjustment for Nonresponse at the State and Sponsor Levels

There is no nonresponse at the State level.

For sponsor nonresponse adjustment, assume that s^*_i sponsors respond to the survey out of the s_i sponsors selected in the *i*th State. Then the nonresponse adjustment to the weights of the responding sponsors is

$$A_i = \frac{\sum_{j=1}^{s_i} W_{ij}}{\sum_{j=1}^{s^*_i} W_{ij}}.$$

The nonresponse adjusted conditional weight is given by

$$W^a_{j/i} = W_{j/i}\ A_i.$$

The overall nonresponse adjusted basic sampling weight is given by

$$W^a_{ij} = W_i\ W^a_{j/i}.$$

This weight was used in sponsor tabulations.

Basic Provider Weights

In calculating provider weights, two changes were made to the conditional sponsor weight that was determined above for sponsor tabulations. Since we selected a sample of providers for each "hit" of the sponsor, for computing the conditional weight of the sponsor for getting the provider weights, we did not include r_{ij} the number of hits. Also, the adjustment for nonresponse of the sponsor was different than done for the sponsor weights used for sponsor characteristics. This was because the number of sponsors giving the list of providers for selection was slightly different from the number of sponsors responding to the survey. The number of providers in the responding group and the number in the nonresponding group were also different.

We first describe the nonresponse adjustment to the sponsor weight.

The conditional sponsor weight for provider tabulations is

$$W^p_{j/i} = \frac{P_i}{P_{ij}}.$$

Let the number of sponsors submitting provider lists be s^{**}_i out of the s_i selected. Then the nonresponse adjustment to the sponsor weight is

$$A^{*}_{i} = \frac{\sum_{j=1}^{s_i} W_{ij} P_{ij}}{\sum_{j=1}^{s^{**}_i} W_{ij} P_{ij}}$$

and the adjusted sponsor weight is

$$W^{b}_{j/i} = W^{p}_{j/i} A^{*}_{i}.$$

The overall sponsor weight is given by

$$W^{b}_{ij} = W_i W^{b}_{j/i}.$$

This sponsor weight was used for all provider tabulations.

For the selection of providers from a selected sponsor, we stratified the providers by Tier 1, Tier 2, and dropout (former providers). Let P_{ijk} denote the number of providers in the *k*th stratum (k= 1,2,3).

Let p_{ijk} be the number of providers selected. Then the basic conditional weight for the *l*th selected provider in the *k*th stratum belonging to the *j*th sponsor in the *i*th State is

$$W_{l/ijk} = \frac{P_{ijk}}{p_{ijk}}.$$

Adjustment for Provider Nonresponse

If out of p_{ijk} providers in the sample, only p^{*}_{ijk} respond, the nonresponse-adjusted conditional provider sampling weight is

$$W^{a}_{l/ijk} = \frac{p_{ijk}}{p^{*}_{ijk}} W_{l/ijk}.$$

The overall provider weight is

$$W^{a}_{ijkl} = W_i \; W^{b}_{j/i} \; W^{a}_{l/ijk}.$$

This weight was used for all provider tabulations in the report. To take account of the complex sampling structure, weighted regressions and all variances were estimated using SUDAAN software. The analysis of portion sizes, which controlled for a large number of provider characteristics, used unweighted regressions.

Other multivariate analyses reported here use weighted (linear or nonlinear) regressions to estimate tiering effects, weighting each observation in inverse proportion to its probability of being included in the sample. Unweighted regressions use sample variances and covariances to estimate the regression parameters for the sample (and for the hypothetical population for which it is a random sample). In sampling-weighted regression, the weights are used to estimate the population values of these variances and covariances, and the population parameter estimates are derived from these. Because sampling weights normally increase the error of estimate (unlike weighting associated with generalized least squares), unweighted estimates are preferred when they can be assumed to be unbiased. For example, if the population regression is correctly specified and the sampling probabilities are completely determined by the included variables, then the unweighted regression will yield unbiased estimates of the regression coefficients. When these conditions cannot be satisfied, as is the present case, sampling weights are commonly used to correct for differences in sampling rates, despite the associated increase in errors of estimate. Sometimes, for example, sampling rates are defined in terms of sparsely sampled categories, with category samples too small to allow them to be represented by dummy variables. In other cases, sampling rates are functions of measured characteristics, which may be added to the regression; however, the estimates then depend on correct specification of the functional form for these added characteristics. Finally, the requirement concerning the correctness of the original specification is quite stringent. In our case, for example, a regression may be misspecified in ways that make it quite sensitive to differences in sampling rates but still offer adequate controls for characteristics associated with tier when applied to a common population.

Nonresponse Bias

The possibility of nonresponse bias—that is, important differences between sample members who respond to the survey and those who do not—deserves consideration in any sample survey. With compound response rates in the range of 41 to 70 percent, the potential for bias is significant. A series of analyses was therefore performed to assess the extent of any bias.

The analyses are necessarily based on those few items of information that are known for the nonresponding as well as the responding providers. These are the number of children enrolled in the home (as reported on the sponsor list) and the provider's location. The latter is represented in the analysis as the percent of homes in each census region (Northeast, South, Midwest, and West).

The analysis compared the mean or percent for all selected sample members and the mean or percent for those responding to the survey. The difference can be viewed as the extent to which the respondents over- or under-represent the specified characteristics of the original sample. As a guide to the importance of the difference, we use a one-sample t-test; that is, we compare the mean of the respondents with the mean of the total sample, taking into account the standard error of the mean of the respondents (but treating the full-sample mean as a constant). The data are unweighted in this

analysis because sampling weights were not computed for nonrespondents. The analysis was carried out separately for the menu survey and meal observation samples. Both samples include only Tier 2 providers.

The results, shown in Exhibits A.1 and A.2 , generally show small differences between the responding providers and the sample frame from which they were drawn. In both samples, the responding Tier 2 providers had somewhat more enrolled children, on average, than the sample selected, although the difference approaches statistical significance only for the menu survey ($p < 0.10$). There is also some tendency for the Northeast region to be over-represented, the West region to be under-represented among menu survey respondents, and the Midwest to be under-represented among meal observation respondents. None of these regional differences are statistically significant, however.

Exhibit A.1
Comparison of Tier 2 Providers Responding to the Menu Survey to Sample Selected

	Respondents	Original Sample	Difference Respondent-Original	Respondent Standard Error	p-value
Mean number of children enrolled	8.9	8.5	0.4	0.24	0.08
Percent of providers that are in region:					
Northeast	25.4%	22.8%	2.6	1.91	0.16
South	21.4	21.8	-0.4	1.8	0.84
Midwest	25.2	24.6	0.6	1.9	0.73
West	27.9	30.9	-3.0	1.96	0.13

Exhibit A.2
Comparison of Tier 2 Providers with Completed Meal Observations to Sample Selected

	Respondents	Original Sample	Difference Respondent-Original	Respondent Standard Error	p-value
Mean number of children enrolled	9.3	8.5	0.8	0.62	0.19
Percent of providers that are in region:					
Northeast	28.1%	22.7%	5.4	4.61	0.24
South	21.9	21.4	0.5	2.42	0.91
Midwest	19.8	24.6	-4.8	4.09	0.24
West	30.2	31.3	-1.1	4.71	0.82