Appendix A
Sampling and Weighting Procedures

Sample

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 CACFP providers, the study included an Operations Survey, a Menu Survey, and a Meal Observation data collection. Most of the analyses presented in this report rely on the Current Provider Operations Survey. The sample design for this survey 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 C, which examines the Former Provider Operations Survey, describes the sample for that survey.)

The sample universe for the study consisted of family child care sponsors, family child care homes, and families 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 CACFP 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 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]

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.

---

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.

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

From these lists, 954 Tier 1 and 1,134 Tier 2 providers were selected for survey participation. Of these, 287 Tier 1 and 388 Tier 2 providers were found to be ineligible (usually because they had left the program between sample selection and survey implementation), leaving samples of 667 and 746, respectively. The Operations Survey questionnaire was completed by 576 Tier 1 providers and 595 Tier 2 providers, for estimated response rates among eligibles of 86.4 and 79.8 percent, respectively. [6]

In multi-stage 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 80.6 and 74.4 percent for Tier 1 and Tier 2, respectively.

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 *i*th selected State. i = 1, 2, 3, 4, ...19, 20. W_i = 1 for States selected with certainty.

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

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

[6] This calculation assumes that all nonrespondents not known to be ineligible were eligible.

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 weight 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

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 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 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 and the number in the nonresponding groups 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 responding to the 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

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

The overall sponsor weight is given by

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

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

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

The overall provider weight is

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

This weight was used for all provider tabulations in the report. To take account of the complex sampling structure, SUDAAN software is used for variance estimation.

All multivariate analyses reported here use weighted linear regressions to estimate Tier 2, 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 weights, 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 mis-specified 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 75 to 80 percent, we would not necessarily expect nonresponse bias to be a problem, but the possibility cannot be ruled out. A series of analyses was therefore performed to assess the extent of any bias.

The analyses is 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 geographic 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 mean of the full sample 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 Tier 1 and Tier 2 providers (the two strata were weighted separately, which corrects for any potential nonresponse bias on this dimension). The results, shown in Exhibits A.1 and A.2, generally show very small differences between the responding providers and the sample frame from which they were drawn. However, the responding Tier 2 providers had significantly more enrolled children, on average, than the sample selected (9.0 *vs.* 8.5 enrolled children, according to sponsor data). There is also some tendency for the western geographic region to be underrepresented among Tier 1 providers (26.1 *vs.* 29.5 percent, $p < 0.10$). Neither of these differences appears sufficient to cause any substantive distortion of findings.

Exhibit A.1
Comparison of Responding Tier 1 Providers with Sample Selected

	Respondents	Original Sample	Respondent-Original Difference	Respondent Standard Error	p-value
Mean number of children enrolled	8.8	8.6	0.2	0.24	0.40
Percent of sponsors that are in region:					
Northeast	27.0	25.2	1.8	1.89	0.34
South	26.0	25.3	0.7	1.87	0.73
Midwest	20.9	20.0	0.9	1.73	0.61
West	26.1	29.5	-3.4	1.87	0.08

Exhibit A.2
Comparison of Responding Tier 2 Providers with Sample Selected

	Respondents	Original Sample	Respondent-Original Difference	Respondent Standard Error	p-value
Mean number of children enrolled	9.0	8.5	0.5	0.23	0.02
Percent of sponsors that are in region:					
Northeast	22.9	21.6	1.3	1.70	0.44
South	21.5	21.9	-0.4	1.66	0.80
Midwest	23.9	24.1	-0.2	1.72	0.91
West	31.7	32.3	-0.6	1.88	0.72