

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 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 survey of parents (the "household" 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.

Sample

The sample universe for the study consisted of family child care sponsors, family child care homes, and families of children cared for in CACFP homes. 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

Random 50-percent subsamples of the Tier 1 and Tier 2 provider samples were designated as the samples for the household survey. These providers were asked to obtain permission from the parents of children in their care for the parents to be surveyed, and to submit the list of children with consenting parents. The sample of children for the household sample was drawn from this list. All children on the list were sampled up to a maximum of eight for Tier 1 providers and 10 for Tier 2 providers. If the provider list included more than the maximum number of children, the maximum

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.

number was drawn randomly from the provider's list. If more than one child from a family was selected, one was designated as the "reference child" about whom most questions were asked.

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 the lists of providers, 465 Tier 1 providers and 447 Tier 2 providers were selected for the next stage of sampling for the household survey. Of these, 109 Tier 1 providers and 137 Tier 2 providers were determined to be ineligible, mainly because they had left the CACFP between the time the sample was selected and the time that lists of children were requested. Of the remainder, 160 Tier 1 providers and 156 Tier 2 providers sent usable lists of children whose parents agreed to be interviewed. This represents response rates of 44.9 and 50.3 percent, respectively, assuming that all providers who were not determined to be ineligible were actually eligible. The response rates at this stage were lower than in any other part of the survey. Some providers simply refused to give lists, some never responded to telephone calls or mailings after the request had been sent, and some reported that all of the parents of their children refused to be interviewed.

The submitted lists comprised 1,068 children in Tier 1, all of whom were selected, and 1,220 children in Tier 2, of whom 1,038 were selected. These children were from 739 households served by Tier 1 providers, and 786 households served by Tier 2 providers. Of these households, 104 served by Tier 1 providers and 123 served by Tier 2 providers were found to be ineligible because they no longer had children in care with the CACFP providers. Interviews were ultimately completed by 576 (Tier 1) and 624 (Tier 2) households, for response rates of 92.0 and 95.0 percent, respectively.⁶

It is useful in multi-stage samples to consider the compound response rate, which is the product of the response rate at each sampling stage – i.e., the sponsor response rate, the response rate among providers asked to submit lists of children, and the parent's response rate. The compound response rates for Tier 1 and Tier 2 households are 38.6 and 44.6 percent, respectively. The major factor contributing to these low response rates is the large proportion of selected providers who did not submit lists of consenting parents, as response rates at the other two stages exceeded 90 percent.

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

⁶ These response rates assume that nonresponding households included the same proportion of ineligible households as the households that were reached.

Weighting

For producing population-based estimates of means and proportions of characteristics relating to households and children, each respondent gets a sampling weight. These weights combine the inverse of the probabilities of selection and nonresponse adjustments.

The subsample of providers from whom lists of children was obtained was drawn from both Tier 1 and Tier 2 providers. A subsample was selected in each stratum using probability proportional to size sampling in which the number of children enrolled was used as the measure of size.

Households represented the fifth stage of sampling: States, sponsors within States, Tier 1 and Tier 2 providers within sponsors, subsamples of the Tier 1 and Tier 2 providers, and households within providers. The overall household weight was therefore obtained as the product of the State weight; the conditional sponsor weight (adjusted for nonresponse); the conditional provider weight (adjusted for nonresponse); and the conditional household weight (adjusted for nonresponse). The conditional household weight 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. $i = 1, 2, 3, 4, \dots, 19, 20$. $W_i = 1$ 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

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

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 for sponsor tabulations. Since we selected a sample of providers for each “hit” of the sponsor, we did not include r_{ij} number of hits in computing the conditional weight of the sponsor for computing the provider weights. Also, the adjustment for nonresponse of the sponsor differs. This was because the number of sponsors giving the list of providers for sampling was slightly different from the number of sponsors responding to the survey. The number of providers in the responding and the nonresponding groups was also different.

We first describe the nonresponse adjustment to the sponsor weight.

The conditional sponsor weight for provider tabulations is

$$W_{j/i}^p = \frac{P_i}{P_{ij}}$$

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

$$W_{j/i}^b = W_{j/i}^p A_i^*$$

The overall sponsor weight is given by

$$W_{ij}^b = W_i W_{j/i}^b$$

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

$$W^a_{ijkl} = W_i W^b_{j/i} W^a_{l/ijk}$$

Sampling Weights for the Subsamples of Tier 1 and Tier 2 Homes Selected for Interviews with Parents and for Children

We show below the derivation of the weights for the Tier 1 subsample of providers for household interviews and for the children served by these providers. The corresponding Tier 2 weights were generally derived similarly. Because children were subsampled from Tier 2 provider lists, however, while all children on Tier 1 provider lists were selected, an additional factor appears in the Tier 2 child weight, as discussed below.

Provider Subsample

Let the number of Tier 1 providers in the main sample in the i th State be m_{ij} . This is the number obtained by aggregating all the selected Tier 1 providers in the main sample from all selected sponsors in the i th State. Let c_{ilq} be the number of children belonging to the q th Tier 1 selected respondent provider in the i th State. Let the subsample of providers selected with probability proportional to the number of children with each provider in the State be u_{il} . The conditional weight for the subsample of Tier 1 providers in the i th State is

$$W^s_{il} = u_{il} \left(\frac{c_{ilq}}{c_{il}} \right)$$

where $c_i = \sum_{q=1}^{m_i} c_{iq}$ is the total number of children over all selected providers in the State. Out of

u_{il} subsample of Tier 1 providers, let the number of respondents be u_{i11} . Let the number of eligible Tier 1 providers in the subsample who are nonrespondents be u_{i12} . Let the number of ineligible Tier 1 providers be u_{i13} . The nonresponse adjustment for the subsample of providers is given by

$$A^s_{il} = \frac{\sum_{q=1}^{u_{i11}} W^s_{ilq} + \sum_{q=1}^{u_{i12}} W^s_{ilq}}{\sum_{q=1}^{u_{i11}} W^s_{ilq}}.$$

The nonresponse adjusted conditional subsampling weight is

$$W^{as}_{il} = W^s_{il} A^s_{il}.$$

Each provider in the subsample of providers received this weight. Next, we identified providers in the subsample by sponsor. A Tier 1 provider in the subsample belonging to the j th sponsor in the i th State received an overall weight of

$$W^s_{ij1l} = W^a_{ij1l} W^{as}_{il}.$$

Child Weight

Let the number of responding eligible children in Tier 1 be c^*_{ilq1} . Let the number of eligible children who are nonrespondents be c^*_{ilq2} . Let the number of ineligible children be c^*_{ilq3} .

The nonresponse adjusted child level weight is given by

$$c_{il} = \frac{(c^*_{ilq1} + c^*_{ilq2})}{c^*_{ilq1}}.$$

Children from Tier 2 providers were subsampled. Let the number of children sampled from the c_{i2q} children with the q th subsampled provider be c^*_{i2q} . The basic conditional sampling weight at the provider level for Tier 2 children is

$$\frac{c_{i2q}}{c^*_{i2q}}.$$

(For Tier 1 children, this factor is simply 1.) Multiplying this by the nonresponse adjustment for Tier 2 children we have a child-level weight of

$$c_{i2} = \frac{c_{i2q}}{c_{i2q}^*} \frac{(c_{i2q1}^* + c_{i2q2}^*)}{c_{i2q1}^*}.$$

The final child level weight for the child belonging to the l th Tier 1 provider and the j th sponsor in the i th State is given by

$$c_{ijll}^f = W_{ijll}^s c_{il}.$$

The corresponding weight for a child in Tier 2 is the same, with the subscripts changed from “1” to “2”. This weight was used for all child (household) tabulations in the report.

All multivariate analyses reported here use weighted linear regressions, 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 39 to 45 percent, the potential for nonresponse bias is very real. 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 sample members. At all sampling stages, the sample member's location is known and is coded as being in one of the four census geographic regions (Northeast, South, Midwest, and West). For sponsors, we also know the number of homes sponsored and the proportion of Tier 1 and Tier 2 homes, as reported by the State agency. For providers, information is available on the number of children enrolled in the home, as reported on the sponsor list. For households, the available information is the number of children in the household who are in the care of the sampled provider, as reported by the provider.

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 to the mean of the total sample, taking into account the standard error of the mean of the respondents (treating the full-sample mean as a universe mean, with no sampling error). The data are unweighted in this analysis because sampling weights were not computed for nonrespondents.

The analysis of sponsor nonresponse is reported elsewhere (Bernstein and Hamilton, E-FAN-02-003). It showed a slight over-representation of sponsors with larger numbers of homes. No pattern of geographic bias was found.

The analysis of providers 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. None of the differences are statistically significant for either Tier 1 or Tier 2. Thus there is no indication of important response bias at this sampling stage, even though this stage had relatively low response rates.

One would not expect to encounter significant nonresponse bias at the household sampling stage because of the high response rates at this stage. This expectation is borne out in Exhibits A.3 and A.4, which show no statistically significant differences between the responding household and the sample selected on those characteristics available for examination.

**Exhibit A.1
Comparison of Responding Tier 1 Providers to Sample Selected^a**

	Respondents	Original Sample	Difference Respondent-Original	Respondent Standard Error	p-value
Mean number of children enrolled	11.7	11.3	0.4	0.515	0.98
Percent of providers that are in region:					
Northeast	22.3%	23.3%	-1.0%	3.33	0.76
South	22.9	21.3	1.6	3.37	0.63
Midwest	22.9	22.9	0.0	3.37	0.98
West	31.9	35.2	-3.3	3.73	0.86

^a Responding providers are those who supplied lists of one or more parents willing to be interviewed.

**Exhibit A.2
Comparison of Responding Tier 2 Providers to Sample Selected^a**

	Respondents	Original Sample	Difference Respondent-Original	Respondent Standard Error	p-value
Mean number of children enrolled	11.2	10.9	0.3	0.48	0.51
Percent of providers that are in region:					
Northeast	22.8%	22.6%	0.2%	3.31	0.94
South	16.1	19.6	-3.5	2.89	0.23
Midwest	29.6	27.3	2.3	3.60	0.52
West	31.5	30.5	1.0	3.66	0.80

^a Responding providers are those who supplied lists of one or more parents willing to be interviewed.

Exhibit A.3
Comparison of Responding Tier 1 Households to Sample Selected

	Respondents	Original Sample	Difference Respondent-Original	Respondent Standard Error	p-value
Mean number of children in the provider's care	1.44	1.45	-0.01	0.03	0.71
Percent of households that are in region:					
Northeast	16.8%	19.1%	-2.3%	1.56	0.15
South	17.2	17.7	-0.5	1.57	0.73
Midwest	24.7	24.1	0.6	1.80	0.75
West	37.0	34.1	2.9	2.01	0.15

Exhibit A.4
Comparison of Responding Tier 2 Households to Sample Selected

	Respondents	Original Sample	Difference Respondent-Original	Respondent Standard Error	p-value
Mean number of children in the provider's care	1.37	1.38	-0.01	0.02	0.59
Percent of households that are in region:					
Northeast	24.7%	22.8%	1.9%	1.73	0.26
South	15.9	16.8	-0.9	1.46	0.54
Midwest	28.4	28.0	0.4	1.81	0.82
West	25.4	26.2	-0.8	1.74	0.62