# Appendix B
# Sample Weights

This appendix describes the sample weights that were used in analyses of the CSFII and ECCCS data.

## CSFII Sample Weights

The CSFII sampling weights compensate for the variable probabilities of selection, differential response rates, and possible sampling frame deficiencies, for both household- and person-level data. The paragraphs below describe how person-level data were weighted to achieve national representativeness, taking into account nonresponse and noncoverage. A composite estimation approach was used to combine the CSFII 1994 to 1996 and 1998 samples, covering four years in total. Although the 1998 supplement to the CSFII comprised children aged 0 to 9 only, the sample was designed to allow these data to be combined with the CSFII/DHKS data for 1994 to 1996, using appropriate weights provided by USDA. These four-year sampling weights were therefore used on the combined dataset.

**Base weights** equal to the reciprocal of the probability of selection were assigned to each sample person. The probability of selection is the product of the probabilities of selecting: (1) the primary sampling unit (PSU); (2) the segment within the PSU; (3) the household within the segment; and (4) the eligible sample person within the household.

The base weights were adjusted for nonresponse on two factors: screening nonresponse, and person nonresponse. **Screening nonresponse** adjustments were made within four classes: census region, MSA status, minority status, and quarter of the year of field operations. The base weight of each eligible sample person was increased by a factor corresponding to the screener nonresponse rate within each class. Screener nonresponse-adjusted weights were then adjusted further to account for **person nonresponse** on classes defined by income level, age, sex, census region, MSA status, quarter of the year of field operations, and minority status of the segment. This produced nonresponse-adjusted base weights for sample persons who responded.

Finally, to compensate for variation in sample counts and possible undercoverage of certain groups, the nonresponse-adjusted weights were ratio adjusted to population estimates from the March Current Population Survey (CPS) for each year of data collection (i.e., 1994-1996 and 1998). Nonresponse rates were calibrated using an iterative process of raking ratio or multiplicative weighting so that the sum of the final weights corresponded to CPS subpopulations defined according to sex, age, home ownership, and several other household and segment characteristics. In other words, the target percentage established by the CPS is the weighted percentage for the sample using the final calibrated weights.

Jackknife replicate weights for variance estimation are also provided for each set of sampling weights. The jackknife replication method was designed to reflect the stratification and clustering used in the CSFII/DHKS sample design, and to capture the effects of the raking ratio adjustments described above. The replicate weights provided for the combined four-year sample (1994 to 1996 and 1998) were used.

## ECCCS Sample Weights

The construction of weights for the ECCCS was conceptually similar to that for the CSFII.[57] Base weights were assigned as the inverse of the sampling probabilities, adjusted for for various special conditions. (For example, Massachusetts was used for the pretest, and this required adjusting the sampling probability somewhat.) Nonresponse adjustments were then made to these sampling probabilities. This was done by stratifying respondents and nonrespondents into homogenous cells and then inflating the inverse of the conditional sampling probabilities for respondents within each cell to account for missing observations from nonrespondents within that same cell. When the resulting weight was unreasonably large, the inflation factor was truncated and a proportional spreading procedure was used.

Sampling for this study followed a multistage, multiphase design. Six weights were constructed for the data analysis, corresponding to CACFP sponsors, child care providers, provider menus, on-site meal consumption, households, and 24-hour recall data. It is the last of these six weights that is used in the current study.

States were the primary sampling units. Eight states were selected with certainty, and another 12 states were selected with probability proportional to size.

Within states, sponsors were stratified by type of provider sponsored (family child care homes (FCCH), child care centers, and Head Start centers), and sponsors were then selected with probability proportional to size. When selecting the sponsor sample, independent child care centers (ICCCs) were treated as sponsors. This was necessary because the state lists of sponsors did not distinguish between "true" sponsors and ICCCs, but in fact ICCCs are child care providers, not sponsors. They entered the provider sample, where they were assigned appropriate weights.

From the sponsor sample, child care providers were sampled (the full provider sample) and asked to answer a provider survey. Providers were also asked to complete a menu survey and a food-preparer interview.

A subset of the full provider sample was selected for the on-site observations (the on-site provider sample). From that subset, children were selected and their CACFP meal consumption was observed. The children's parents were interviewed about their meal consumption while not in child care and to obtain information on household characteristics such as income and maternal employment.

The selection of children into the sample was based on logistic considerations. Although children in FCCHs are usually fed together, children in centers tend to be fed in small groups whose composition is homogenous with respect to age. To allow the observers at centers to watch the food consumption of the sampled children, a group of noninfant children was first chosen and then six children were selected from that group. (If the center served infants, one infant and five children from the chosen non-infant group were selected instead.) In FCCHs, the selected sample simply comprised six non-infant children (if the number of eligible children was as many as six), or five noninfants and one infant (if any eligible infants were enrolled). Children ineligible to be sampled included infants who were exclusively breastfed, children who were not enrolled for both of the scheduled observation

[57] For a full description of the construction of the ECCCS weights, see Fox *et al.*, 1997, Appendix E.

days, and siblings of sample members. The first two groups were deemed outside of scope. The siblings were represented by other children enrolled with the same provider, i.e., by increasing the child weights of the other children in the same group proportionally.

The analysis of meals consumed in care is intended to describe **children in care on a typical day**—not all children enrolled in care. Hence, children who were selected into the sample but absent on one or both observation days were not nonrespondents for purposes of constructing the corresponding weights, but rather outside of scope.

The overall response rate for the 24-hour recall sample was quite low, including as it did attrition at all of the intermediate stages. Nonetheless, the resulting sample of children is in principle nationally representative. The assigned weights take account of both sampling probabilities and nonresponse rates to the extent possible.