APPENDIX E
SAMPLING AND WEIGHTING

The emergency food assistance system (EFAS) study required complex sample design and weighting procedures. Censuses were made of some types of organizations—food banks, food rescue organizations, and emergency food organizations—while kitchens and pantries were sampled. This appendix describes the sampling and weighting procedures for the study, with special emphasis on the target population, sampling frame, sample design, and weighting plan for each component.

TARGET POPULATIONS

The target population for a survey is the set of entities for which inferences are to be made using study results. For the Emergency Food Assistance System Study, the target population is composed of food banks, food rescue programs, emergency food organizations, pantries, and kitchen facilities providing emergency food assistance and located within the contiguous United States.

The level of the organization used in defining the target population was the facility. By facility, we mean the individual locations or establishments at or from which EFAS services are provided. An EFAS agency may operate multiple food banks, pantries, or kitchens. The population units included in this study are the individual facilities rather than the parent organization. In addition, some EFAS organizations provide more than one type of EFAS service at the same location. A food bank might provide pantry services, for instance. Each EFAS activity at such multipurpose locations is treated as a separate facility.

Some EFAS services are provided in transient locations rather than, or in addition to, the fixed location where the organization owns or rents building space. Only permanent locations are included as separate entities in this study. Mobile services and other services provided at transient locations are considered a part of the fixed location or facility from which they operate. For instance, an emergency food kitchen might offer meals onsite at a fixed facility and also distribute box meals via mobile vans throughout the city. In this case, the fixed location was considered one kitchen facility. If selected for this study, that facility was asked to include both onsite and mobile services

Specific definitions for food banks, food rescue programs, emergency food organizations, food pantries, and emergency kitchens are provided next.

The food bank population

Food banks are umbrella organizations or clearinghouses that solicit marketable and surplus food and grocery products and distribute these products to local nonprofit charities or client agencies, which in turn distribute the food directly to needy individuals and families. Note that under our definition food banks supply food to other organizations, not directly to those in need.

Organizations functioning as independent food banks may also be loosely associated or allied with other food banks. For instance, America’s Second Harvest, the largest food bank network in the United States, has both “direct affiliate” food banks and Subsidiary Distribution Organizations (SDOs) linked to these food banks. These SDOs obtain significant amounts of food from their affiliated America’s Second Harvest food bank, which they then distribute to EFAS facilities. These SDOs typically operate autonomously and often have their own board of directors. To illustrate, a direct-affiliate food bank may formally represent America's Second Harvest for an entire state, and it may have direct operational responsibility for part of the state. In addition, two SDOs may function as primary distribution agencies for the remainder of the state. Under our target population definition, SDOs are considered separate food bank facilities. In terms of the target population of food banks, this state would be viewed as being served by three food banks.

The food rescue program population

Food rescue programs are organizations that obtain perishable foods from farmers or from food retail establishments, such as restaurants, and then distribute these products to local nonprofit charities or client agencies, which in turn distribute the food directly to needy individuals and families. Food rescue organizations differ from food banks in their emphasis on perishable food.

Emergency food organizations

Emergency Food Organizations are defined as organizations that are not principally EFAS agencies but that sometimes distribute commodities they receive from the Emergency Food Assistance Program (TEFAP), through the state directors of TEFAP. For the purpose of this study, the definition of emergency food organizations is limited to organizations that distribute these commodities to other agencies, such as pantries or kitchens (rather than distributing them directly to needy people or households).

The food pantry population

Food pantries are facilities that distribute groceries and other basic supplies directly to needy individuals, without charge, for offsite use. It is not uncommon for food pantries to be referred to as "food banks." For this study, the function, rather than the name, defines the nature of the EFAS facility.

Some EFAS facilities are targeted to specific clienteles. With the exceptions noted in the remainder of this section, we include such facilities in the relevant target population when they clearly provide EFAS services.

The emergency kitchen population

Emergency kitchens are defined as facilities that prepare or assemble meals for distribution either onsite or offsite to needy recipients who do not reside on the premises. The meals are provided at little or no cost. Facilities distributing food funded under Title IIIC of the Older Americans Act, the Child and Adult Care Food Program, or the National School Lunch and School Breakfast programs are excluded from the target population.[1]

For a food provider to be considered an EFAS facility, meals must be the primary service that it offers. Many EFAS facilities require modest participation in other activities at the time the individual receives food. For example, there may be mandatory service referral activities or a

[1]Each of these programs has been studied extensively by itself and is excluded from the present study as a way of clearly delineating the objectives and scope of the work.

religious service. For purposes of the study, such secondary activities are compatible with being labeled an EFAS provider. However, when food distribution is incidental to other activities, as in senior day care centers, the facility was not classified as an EFAS provider.

Some shelters provide meals to people not spending the night there. Such shelters were considered to be operating an emergency kitchen and were included in the study.

SAMPLING FRAMES

For this study, area and list frames were used. Construction of list frames began with the creation of lists of the nation's food banks and established food rescue programs. For kitchens and pantries, such lists were unavailable. Instead, an area frame was created of primary sampling units (PSUs), with PSUs defined as a county or a group of adjacent counties. List frames of pantries and kitchens were created for sampled PSUs only.

The food bank frame

Construction of the frame for food banks began with the list of food banks affiliated with Second Harvest, as are most food banks in the United States. To a substantial degree, Second Harvest also supplied information on food banks not formally associated with their organization but known to them via informal channels. To identify the remaining food banks, we consulted the *International Food Bank Directory*, USDA's National Hunger Clearinghouse, USDA's *Citizen's Guide to Food Recovery*, and national organizations involved with hunger, such as Foodchain, the Salvation Army, United Way, the American Red Cross, Catholic Charities of America, and others.

The food rescue frame

The frame for food rescue programs was constructed from lists of agencies associated with the largest organization of food rescue programs, Foodchain, and the Society of St. Andrew. Most large food rescue organizations are associated with these two organizations, so we believe their affiliates should account for a large majority of all food rescue activity currently taking place in the United States. Local, nonaffiliated food rescue agencies are a source of undercoverage for the frame.

Emergency food organization frames

Lists of emergency food organizations were obtained from state TEFAP directors.

The kitchen and pantry frames

Building a list frame for all EFA kitchens and pantries would have been cost prohibitive. Instead, our frame-building for kitchens and pantries involved dividing the land area of the contiguous United States into nonoverlapping geographic areas, selecting a sample of 360 areas, and then developing list frames of kitchens and pantries for the sampled areas only. The geographic areas that comprise the area frame are referred to as primary sampling units (PSUs). All locations within the contiguous United States were included in one, and no more than one, PSU.

To build the area frame, we began with county-equivalent records from the Area Resource File (ARF) maintained by the Department of Health and Human Services for the 48 Contiguous States and the District of Columbia.[2] The phrase "county-equivalent" is used because of the way ARF treats independent cities (Department of Health and Human Services, 1998). Generally, the ARF combines independent cities with their original counties. For instance, the city of Manassas, Virginia is combined with Prince William County. Some relatively large independent cities, however, are treated as county equivalents; Alexandria, Virginia is one such county equivalent.

PSU formation involved dividing the contiguous United States into land areas in which each land area would meet a prespecified minimum size constraint of 4,250 persons living in poverty.[3] Counties with smaller poverty populations were collapsed with adjacent counties to yield PSUs that met the minimum size constraint. Collapsing occurred within cells defined by Census region and metropolitan area, whenever possible.[4] Collapsing was done within region and further within the same MSA/PMSA, if possible. An automated procedure was created for this purpose, but some manual adjustment was needed afterward. The procedure was designed to create PSUs that were contiguous, similar with respect to their poverty counts, and had the minimal number of counties collapsed to form the PSU.

Each ARF record for a county contains variables listing the IDs of its adjacent counties. These variables were used to create the collapsing algorithm. We began the process by extracting those county records that met the minimum size measure of 4,250 persons in poverty.[5] These records were placed in an interim PSU file, where they were available for further collapsing if necessary. Then we sorted the remaining small counties by region, MSA/PMSA Code, and poverty count. For each county data record, we merged on the poverty count for each of the counties listed as adjacent to that county and added their region and MSA/PMSA characteristics. These adjacent counties were then reordered from smallest to largest in terms of poverty count. Beginning with the first small county record, we added the county's poverty count to that of its adjacent counties until a collapsed set of counties was obtained that met the minimum size constraint.

The completed area frame contained 1,895 PSUs. Region and metropolitan status variables were defined for each PSU; for collapsed counties these variables were defined based on majority population rules.

Multiple sources were used to obtain lists of kitchens and pantries located in each sampled PSU. Lists of Second Harvest facilities were extracted from the databases maintained by its affiliated

---

[2]U.S. Department of Health and Human Services. Bureau of Health Professions. Office of Data Analysis and Management. Area Resource File (ARF) as of February 1998. Washington, DC, February 1998.

[3]The ideal measure would have been based on the number of EFAS pantries and kitchens, but this information is unknown.

[4]Metropolitan status was defined as: rural area; metropolitan area of less than 100,000 population; metropolitan area of 100,000 to 249,999 population; metropolitan area of 250,000 to 999,999 population; and metropolitan area of 1,000,000 or more population.

[5]The number-in-poverty variable for the county was a constructed variable defined as the population size in 1996 (the date of the relevant data item in the ARF file) multiplied by the fraction of the population in poverty in 1989 (an ARF variable derived from the relevant data item in the ARF file).

food banks. As a part of our census of food banks and established food rescue programs, we requested lists of the food pantries and kitchens served by the remaining food banks and food rescue programs. State directors of the Emergency Food Assistance Program (TEFAP) were contacted to obtain lists of the EFAS facilities to which they distributed TEFAP commodities. Additional lists of EFAS facilities were derived by telephone contacts with approximately 15 social service agencies and other local informants in each sampled PSU. (See Appendix A for a description of this extensive local calling activity.)

The results of this process were multiple, overlapping, hardcopy lists for each sampled PSU that contained kitchens and pantries, as well as ineligible social support organizations like senior centers and after-school programs. Our next step was to review each list, labeling each entry as a kitchen, a pantry, or an out-of-scope organization. Duplicate listings for the same location were common. Our approach was to first deduplicate each individual list, marking out duplicate entries for the same facility. Kitchen and pantry services provided from the same location were treated as two different facilities. Listings that could not be classified were labeled as "unknowns" and treated as kitchens in sampling. The multiple lists for each PSU were then ordered and a unique linkage approach used in the manual deduplication across lists. Listings were considered ineligible when the same facility was found on previous lists, based on the ordering of the lists. After each list was internally deduplicated, the listings were compared with the listings on the previous lists and only new entries were retained. Once completed, the listings corresponding to kitchens, pantries, and unknowns were numbered sequentially within PSUs.

SAMPLE DESIGNS AND WEIGHTS

The Emergency Food Assistance System Survey consisted of multiple samples, one for each type of EFAS facility. Sampling for pantries and kitchens involved the selection of an area sample of 360 PSUs, list construction for the sampled PSUs, and then selection of pantry and kitchen listings. Only 402 food banks, 91 food rescue organizations, and 124 emergency food organizations were identified. Rather than sample them for interviewing, we attempted interviews with all food banks, food rescue organizations, and emergency food organizations. Little or no cost savings would have been attached to a sample, as opposed to a census. The remainder of this section discusses kitchen and pantry selection.

Area sample

For food pantries and kitchens, we began with the area frame already described and selected a probability sample of 360 PSUs, with probability proportional to size (PPS). The size measure was the number of persons living in poverty. The total number of EFAS kitchens and pantries in the universe was an unknown quantity. The sampling uncertainty associated with the absence of information on EFAS facilities led to the decision to select a large number of PSUs (360), each of which is fairly large (4,250 or more persons in poverty).

To select the PPS sample of 360 PSUs, we used Chromy's probability-minimum-replacement sequential sampling procedure.[6] To make n_1 PSU selections from a frame of N_1 PSUs in the frame,

[6]James R. Chromy. "Sequential Sample Selection Methods", Presented at the Annual Meeting of the American Statistical Association, Washington, DC, 1979.

Chromy's procedure partitions the PSUs into 360 zones of equal size, in terms of the size measure. (Individual PSUs may straddle zone boundaries.) Exactly one sample PSU is then selected from each zone. This zoned sequential selection makes possible a deep implicit stratification of the PSUs by a controlled ordering of the PSUs. Moreover, the zones can be defined so that all pairs of PSUs have a chance of appearing together in the sample,[7] a requirement for unbiased estimation of sampling variances. The probability-minimum-replacement feature refers to the treatment of PSUs for which the expected number of selections exceeds 1 (for example, self-representing PSUs). The actual number of times a PSU is selected for any specific sample differs from the expected number by less than 1, and the average number of selections over all possible implementations of Chromy's procedure equals the expected number given in the above computation.

Chromy's sequential sample selection procedure was used to select n_1=360 PSUs with probability proportional to the number in poverty. No explicit strata were used in sampling. Prior to sample selection, the PSUs were sorted in a serpentine manner by Census region, metropolitan status, percent minority, and total population (including poor and nonpoor). Percent minority was defined as the total black and Hispanic populations divided by the total population and then converted to these percentage-based categories: 0-9, 10-24, 25-49, and 50-100. Let $S(i)$ be the size measure associated with the ith PSU. Then the expected relative frequency $n_1(i)$ with which the ith PSU was selected is given by

$$(1) \qquad E[\, n_1(i)] = \frac{360\, S(i)}{S(+)}$$

where $S(+)$ is the sum of the size measures over all PSUs in the area frame and 360 is the total number of PSUs selected. Note that very large PSUs could be sampled multiple times (that is, $n_1(i)>1$ for some PSUs). The sample of 360 PSU selections resulted in a total of 294 unique PSUs. The sample size at the second stage was adjusted to reflect the number of times each unique PSU was selected. Thus, very large PSUs (such as Los Angeles) were selected multiple times and had multiple second-stage samples selected from them.

With this approach, the expected frequency of selection of the ith PSU is

$$(2) \qquad E[n_1(i)] = \frac{360\ S(i)}{S(+)}$$

where $n_1(i)$ is the number of times the ith PSU is selected, $S(i)$ is the total poor in the ith PSU, and $S(+)$ is the total poor across all PSUs. The associated sampling weight for each first-stage selection was calculated as the inverse of the probability of selection or

$$(3) \qquad PSUWT(i) = \frac{S(+)}{360\, S(i)} \ .$$

---

[7]This is done by introducing a random element in how the zones themselves are defined.

Selection of providers

The next stage of sampling was the selection of kitchens and pantries from the list frames developed for each sampled PSU. Separate samples of kitchens and pantries were randomly selected from each PSU with equal probability within PSUs.

We began the sampling process by selecting the initial samples of 2,384 kitchens and 2,410 pantries from the list frame. Stratified simple random sampling was used to select separate samples of kitchens and pantries where the PSUs formed the strata. A permanent random number was generated for each facility in the frame. Let j denote the type of facility (kitchen versus pantry). To select the sample of type j facilities from PSU i, we ordered the type j facilities within each PSU by the permanent random number and selected the first $n_2(ij)$ frame units for the second-stage sample.

With PPS selection of the PSUs, our intention had been to select a fixed sample size from each PSU. Had our size measure been well correlated with the number of EFAS facilities in each PSU, this approach would have led to a self-weighting sample (that is, one with equal sampling weights for kitchens and pantries across PSUs). However, our size measure, the number of people in poverty, proved to be related to, but imperfectly correlated with, the number of kitchens and pantries in each PSU. To correct for the unequal weighting this imperfect correlation could have produced, we adjusted the PSU sample sizes of kitchens and pantries to reflect the discrepancy between the expected number of kitchens and pantries and the actual number found of each type.[8] However, to ensure that some sample was selected from each PSU, minimum sample sizes of four pantries and eight kitchens were set for each PSU.[9] PSUs containing fewer than four pantries or fewer than eight kitchens had all facilities of that type included in the sample with certainty. We also set maximum sample sizes of 12 pantries and 20 kitchens for each PSU. Imposing minimum and maximum PSU sample sizes led to unequal weighting across PSUs but was deemed necessary to control for the potential deleterious effect that clustering can have on the variance of survey estimates.

The sampling frame for each PSU, a set of hardcopy lists, had been manually deduplicated. To ensure that undetected duplication did not compromise study results, we searched each PSU's frame after sampling for potential duplicates of each sampled facility. The unique linkage rule was again applied, with sampled records labeled as ineligible when they did not correspond to the first frame listing for the facility. A total of 79 sampled pantries and 176 sampled kitchens were declared ineligible because they were duplicates of a previous frame listing for the PSU.

The unit we sampled was the facility, which was defined as the cross between type of EFAS service provided (kitchen versus pantry) and location. Locations providing both types of EFAS services were treated as two different facilities. When the pantry facility was sampled for such locations, the pantry portion of the questionnaire was administered, when the kitchen facility was sampled, the kitchen portion was administered.

---

[8]This adjusted sample size reflected the number of times the PSU was selected for the sample.

[9]PSUs that were selected multiple times had the minimums (and maximums) adjusted by the number of times the PSU was selected for the sample.

To be eligible for the survey, the sampled facility had to be in current operation at the specified location and providing the specified EFAS service. Misclassified facilities, such as a sampled pantry that turned out to be a kitchen or vice versa, were labeled as ineligible for that portion of the survey, as were sampled pantries and kitchens that had changed locations. New frame records by service type entities were created for such misclassified locations, and these frame records were given an opportunity of selection *if not already listed for that type of EFAS service*. Due to the multiple sources used to generate lists for each PSU, we expected that most such misclassified entities would be found on the frame correctly classified. For clarity, we refer to the sample selected from the original frame as the *primary sample* and to the sample selected from frame records added during data collection as the *secondary sample*.

To correct for potential coverage problems in our frame, we asked each sampled facility to report other locations where their parent organization provided EFAS services. We also asked each sampled facility to list nearby EFAS facilities. Location information was collected for each facility mentioned by a primary selection. The information provided was used to determine whether the mentioned facilities were already listed in the original frame (most were). Previously unlisted facilities were added to the frame and given an opportunity of selection as a part of the secondary sample.

The multiple ways in which secondary facilities were identified made it impossible to determine the probability with which they had been identified. This led to a decision to treat the discovered facilities as just another frame unit and not attempt to account for the probability of their discovery. Separate samples were selected of primary and secondary facilities with equal selection probabilities within PSUs.

The sampling approach for secondary facilities differed somewhat for pantries and kitchens. Our original plan called for using stratified Bernoulli sampling to select secondary facilities, with PSUs as the strata and the same probability of selection as that used for the PSU's primary facilities of that type. This approach results in a variable sample size but allows for sample selection on a flow basis as secondaries are identified. This stratified Bernoulli sampling approach was used for secondary pantries. A random number was generated for each secondary pantry discovered during data collection. If the random number was less than the probability of selection for that PSU's primary pantries, the secondary pantry was included in the sample. A total of 508 secondary pantries were identified, with 122 of these selected for the sample.

A different approach was used for sampling secondary kitchens, where primary and secondary frame units were combined and an expanded sample selected. The change of approach was in part due to an unexpectedly large ineligibility rate for the initial sample of primary kitchens (about 50 percent). Even with the addition of a Bernoulli sample of secondaries, it was clear that the kitchen sample-size target could not be met without sampling additional primary kitchens. Once the necessary sample size had been determined, the primary and secondary frame units were combined and the new sample allocated to PSU i following the allocation procedures outlined above. To select the sample of kitchens (j=1,2) from PSU i, we ordered the primary and secondary kitchens by the permanent random number and designated the first $n_2(ij)$ kitchens for the sample, using the revised PSU sample sizes. Because each PSU's sample allocation was either substantially greater than or equal to its initial allocation, this approach included all members of the initial sample in the expanded sample. The expanded frame contained 4,871 primary kitchens and 544 secondary kitchens, from which 2,764 primary kitchens and 393 secondary kitchens were selected.

For kitchens, the within-PSU sample can be regarded as a simple random sample of primary and secondary kitchens. Let $N(i1)$ be the total number of primary kitchens identified for PSU *i,* and let $N(i2)$ be the total number of secondary kitchens discovered during interviewing for PSU *i*. Let $n_2(i+)$ be the number of primary and secondary kitchens sampled from PSU *i*. Then the conditional weight CONDWT(*ijk*) for sampled kitchen *k* of facility type *j* from PSU *i* is

$$(4) \qquad CONDWT(ijk) = \frac{N(i3)+N(i4)}{n_2(i3)+n_2(i4)(/h(i)}$$

where j=1 for primary kitchens and j=2 for secondary kitchens and $h(i)$ is the number of times the PSU is selected. Note that the *CONDWT*(*ijk*) is identical within PSUs for primary and secondary kitchens.

For pantries, the primary and secondary sample designs differed, although the selection rates within PSUs were identical for primary pantries (j=3) and secondary pantries (j=4). Primary pantries had a simple random sample of size $n_2(i3)$, selected without replacement from PSU *i*. In contrast, each secondary pantry was given the same probability of selection as primary pantries, but an independent decision was made for each secondary pantry resulting in a variable-sized secondary sample. Let $N(i3)$ be the total number of primary pantries identified for PSU *i* and $N(i4)$ be the total number of secondary pantries discovered during interviewing for PSU *i*. Let $n_2(i3)$ be the total number of primary pantries sampled from PSU *i* and $n_2(i4)$ be the total number of secondary pantries sampled from PSU *i*. Then the conditional weight *CONDWT*(*ijk*) for sampled pantry *k* of facility type *j* from PSU *i* is

$$(5) \qquad CONDWT(ijk) = \frac{N(i3)}{n_2(i3)(/h(i)}$$

where $h(i)$ is the number of times the *i*th PSU was selected.

The overall sampling weight *SAMPWT*(*ijk*) for the *k*th facility of type *j* from PSU *i* is the product of the PSU sampling weight and the conditional facility sampling weight or

$$(6) \qquad SAMPWT(ijk) = PSUWT(i)\, CONDWT(ijk).$$

ADDITIONAL WEIGHTING ISSUES

Following are a number of additional procedures used in constructing the final analysis weights.

Poststratification Adjustment

The sampling weights can differ from the frame counts in minor ways, particularly for the pantry sample, where the sample size was variable for the secondary pantry portion. To ensure accurate estimation of population totals, a poststratification adjustment was made to the sampling weights.

Poststratification adjustment was chosen to counter the weight variability associated with variable sample sizes for the secondary pantries and to provide more accurate estimates of population totals. Both pantries and kitchen weight totals were forced to the estimated total number of pantries and kitchens within poststrata.[10]

We began by defining poststrata for the pantry and kitchen samples as PSUs or groups of PSUs that contain 20 or more responding eligible entities. The same process was used in defining the poststrata, but the poststrata differed for the two types of facilities. To define the poststrata, we first ordered the PSUs by selection order.[11] To define the poststrata for each type of facility, we grouped PSUs sequentially until we had 20 or more responding eligible pantries and kitchens. That was the first poststratum. We continued the process to define the remaining poststrata, relaxing the rule of 20 responding eligible pantries and 20 eligible responding kitchens when enforcement would require combining PSUs from different Census regions or different metropolitan statuses.

The poststratification adjustment $ADJ_{post}(p)$ for sampled facilities in poststratum p was defined as follows:

$$(7) \quad ADJ_{post}(p) = \frac{\sum_{ijk \in p} PSUWT(ijk) * h(i) * [N(i1) + N(i2)]}{\sum_{ijk \in p} SAMPWT(ijk)}$$

The numerator of this adjustment is an estimate of the total number of facilities in the contiguous United States, hereafter referred to as the "Frame Estimate."

The adjustment $ADJ_{post}(p)$ was then applied to the sampling weight $SAMPWT(ijk)$ for the ijkth case from the pth poststratum as follows:

$$(8) \qquad POSTWT(pijk) = ADJ_{post}(p) \quad SAMPWT(ijk)$$

Note that all sampled cases had a poststratified weight defined for them.

Definition of weighting classes

Nonresponse adjustments were made within weighting classes. The first step in defining the weighting classes was to classify each sampled facility in terms of response and eligibility. To do this, we defined a response and eligibility indicator ELIGRESP as follows:

---

[10]The kitchen sample had fixed sample sizes and hence the sum of the weights automatically yielded this estimated total.

[11]For sample selection, the frame was sorted by Census region, metropolitan status, percent minority, and total population (including nonpoor and poor). Metropolitan status was defined as the four-digit MSA/PMSA code plus a defined value for rural areas. Percent minority was defined as the total black population plus the total Hispanic population divided by the total PSU population. Before sorting by percent minority, we reclassified it as 0-9, 10-24, 25-49, and 50+. A serpentine sorting procedure was used prior to selecting the sample using Chromy's procedure.

ELIGRESP =	0	eligibility status unknown
	1	identified as ineligible
	2	identified as eligible and questionnaire completed
	3	identified as eligible but questionnaire not completed
	4	identified as a duplicate of a previous frame listing.

Note that codes of 1, 2, 3, and 4 imply that eligibility status is known and hence the case is a respondent for the purposes of eligibility-status determination. Codes 1, 2, and 4 are considered to be questionnaire respondents. Codes 1 and 4 are a special form of questionnaire respondent because *by definition* no questionnaire is needed from any ineligible and hence complete information has been obtained. Code 4 cases were identified via an examination of the frame, and hence interview data were not required.

The frame for this study was a compilation of hardcopy lists for each PSU that had been deduplicated manually. After the sample was selected, the deduplication process was repeated for the sampled cases to ensure that no sampled listings were duplicated in the sample frame. When duplicate entries were detected, the sampled listing was considered eligible for interview *only* when it was the first such listing associated with the facility.[12] Sampled listings were labeled as ineligible when they were associated with the second or higher listing of a particular facility. Weighting classes c were defined based on the poststratum p, with frame duplicates removed to a separate weighting class. The weighting class variable was used for all nonresponse adjustments.

Screening nonresponse adjustment

The first nonresponse adjustment accounted for nonresponse to screening. The underlying assumption behind the adjustment was that the percentage of eligibles within each weighting class among cases with unknown eligibility was the same as among cases with known eligibility within that class.

The screening nonresponse adjustment $ADJ_{sc}(c)$ for screening respondents in weighting class c was defined as follows:

$$(9) \qquad ADJ_{sc}(c) = \frac{\sum_{ijk \in c} POSTWT(ijk)}{\sum_{ijk \in c} \delta_{sc}(ijk) \; POSTWT(ijk)}$$

where $\delta_{sc}(ijk)$ is equal to 1 for cases where eligibility was determined and 0 otherwise.
The adjustment $ADJ_{sc}(c)$ was then applied to the poststratified weights to obtain the screening nonresponse adjusted weight $SCADJWT(cijk)$ for the ijkth case from the cth weighting class or:

$$(10) \quad SCADJWT(cijk) = \delta_{sc}(ijk) \;\; ADJ_{sc}(c) \;\; POSTWT(ijk)$$

Note that screening nonrespondents have $\delta_{sc}(ijk)$ equal to 0 and hence a screening nonresponse adjusted weight of 0.

[12]This approach eliminated multiple selection opportunities associated with duplicate listings.

Questionnaire nonresponse adjustment

The next step in weighting was to adjust for the loss of completed questionnaires from beneficiaries known to be eligible. To create new weighting classes c' for this adjustment, ineligible cases were removed from the previous weighting classes and placed in a separate weighting class. (Screening nonrespondents were also placed in a separate class and given an adjustment factor of zero.) Then a questionnaire nonresponse adjustment factor $ADJ_{ques}(c')$ was calculated for weighting class c' as:

$$(11) \quad ADJ_{ques}(c') = \frac{\sum_{i \in c'} \delta_{sc}(ijk) \quad SCADJWT(cijk)}{\sum_{i \in c'} \delta_{ques}(ijk) \quad SCADJWT(cijk)}$$

where $\delta_{ques}(ijk)$ is equal to 1 for questionnaire respondents and 0 otherwise.
Ineligible respondents by definition provided all required data so $\delta_{ques}(ijk) = \delta_{sc}(ijk) = 1$ for ineligible cases. Hence, the questionnaire nonresponse adjustment factor for the class composed of ineligibles has $ADJ_{ques}(c') = 1$.
The questionnaire nonresponse adjusted weight $QADJWT(cc'ijk)$ for the *ijk*th facility in the *c* and c'th weighting classes is calculated as the product of its screening nonresponse adjusted weight and its questionnaire nonresponse adjustment factor or:

$$(12) \quad QADJWT(c'cijk) = \delta_{ques}(ijk) \ ADJ_{ques}(c') \ SCADJWT(cijk)$$

Note that eligible nonrespondents and screening nonrespondents have $\delta_{ques}(ijk) = 0$ and hence a questionnaire nonresponse adjusted weight of 0.

Weight truncation and smoothing

The final step in weighting was to examine the distribution of the questionnaire nonresponse adjusted weights to determine whether weight truncation was needed. In developing weights, some attention must be given to the increase in variance caused by unequal weights. Frequently, very large (or very small) weights are truncated and then the weights adjusted so that they sum to the total of the untruncated weights. This process is referred to as *weight truncation* or *smoothing*.

The sample of kitchens and pantries was designed so that equal weights would have resulted if the size measure had been well correlated to the number of facilities. Although steps were taken to reduce the unequal weighting resulting from our less than perfectly correlated size measure, we could not remove all such unequal weighting across PSUs. As a consequence, a few records had their weights truncated. For pantries, we truncated weights for one PSU that had a very large questionnaire-nonresponse adjusted weight of 114 (the next largest weight was 52). A maximum weight of 60 was set for pantry weights. For kitchens, we encountered a few weights that were very small (less than one). This event occurred when the PSU was so large that it was selected multiple times (*PSUWT* < 1) but had so few listed kitchens that we were forced to sample all its kitchens with certainty (CONDWT=1). A minimum weight of 1 was set for kitchen weights.

Having determined the minimum weight allowed and the maximum weight allowed, we proceeded to calculate the truncated weight. Let MINWT be the minimum weight and MAXWT be the maximum weight allowed. The truncated weight $TRUNCWT(ijk)$ for facility k from frame type j and PSU i was defined as follows:

$$
\begin{aligned}
&TRUNCWT(cc'ijk) = \text{MINWT} && \text{if } QADJWT(cc'ijk) < \text{MINWT} \\
(13)\quad &TRUNCWT(cc'ijk) = QADJWT(cc\square ijk) && \text{if MINWT}\square QADJWT(cc'ijk)\square\text{MAXWT} \\
&TRUNCWT(cc'ijk) = \text{MAXWT} && \text{if } QADJWT(cc'ijk) > \text{MAXWT}.
\end{aligned}
$$

These truncated weights no longer added to the population total.

The next step was to calculate a smoothing adjustment factor to correct the weight total by adjusting the weights of the untruncated cases only. The smoothing adjustment factor $ADJ_{sm}(cc')$ for class c' was calculated using the same weighting classes as were used for the questionnaire nonresponse adjustment:

$$
(14)\quad ADJ_{sm}(c') = \begin{cases} \dfrac{\sum_{ijk\in c'} QADJWT(c'cijk) - \sum_{ijk\in c'} \delta_{trunc}(ijk) TRUNCWT(c'cijk)}{\sum_{ijk\in c'} TRUNCWT(c'cijk) - \sum_{ijk\in c'} \delta_{trunc}(ijk) TRUNCWT(c'cijk)} & if \ \ \delta_{trunc}(ijk) = 0 \\[2ex] 1 & \text{if } \delta_{trunc}(ijk) = 1 \end{cases}
$$

where $\delta_{trunc}(ijk)$ is 1 if the ijkth facility had its weight truncated and 0 otherwise. This adjustment factor was applied to all untruncated weights [$\delta_{trunc}(ijk) = 0$] to recover the weight lost (or gained) due to truncation. The final analysis weight is the product of the smoothing adjustment and the truncated weight or

$$
(15)\quad ANALWT\left(c'cijk\right) = \delta_{ques}\left(ijk\right)\ ADJ_{sm}\left(c'cijk\right)\ TRUNCWT\left(c'cijk\right).
$$