



HUNGER IN AMERICA 2014

Technical Volume
September 2014



Hunger in America 2014 (HIA 2014)

Technical Volume Prepared for Feeding America

Authors

Jill Montaquila, PhD, Sampling Team Leader
Nancy S. Weinfield, PhD, Project Director

September 15, 2014

Prepared for:
Theresa Del Vecchio-Dys
Manager of Social Policy
Research & Analysis
Feeding America

Prepared by:
Westat
An Employee-Owned Research Corporation®
1600 Research Boulevard
Rockville, Maryland 20850-3129
(301) 251-1500

Table of Contents

<u>Chapter</u>		<u>Page</u>
1	Introduction	1
	1.1 Overview of the Hunger in America 2014 Study	1
	1.2 Target Population and Survey Population	2
	1.3 Design Objectives	7
	1.4 Sampling Unit and Unit of Analysis	8
2	Client Survey Sample Design and Selection	9
	2.1 Overview of the Sample Design	9
	2.1.1 Multistage Design	9
	2.1.2 Dual Frame Sampling	10
	2.1.3 PPS Sampling	10
	2.1.4 Stratification	11
	2.1.5 Target Sample Sizes and Related Assumptions	11
	2.1.6 General Precision Considerations	15
	2.2 Sampling Agencies and Programs	17
	2.2.1 Sampling Frames	17
	<i>Agency Frames</i>	18
	<i>Lists of Programs in Sampled Agencies</i>	19
	<i>Missing Data in the Sampling Frames</i>	19
	2.2.2 Measures of Size and Probabilities of Selection	21
	2.2.3 Identification of Certainty Agencies and Undersized Agencies and Programs	22
	2.2.4 Stratification and Allocation	23
	2.2.5. Selection of Noncertainties	24
	2.2.6 Post Hoc Revisions to the Sample of Programs	25
	<i>Selection of Reserve Sample</i>	25
	<i>Release of Non-Sampled Large Programs</i>	26
	<i>Creation of Additional Program Visits for Multiple Sites in Large Programs</i>	26
	<i>Subsampling To Reduce the Number of Program Visits</i>	27
	2.3 Sampling Clients	31

<u>Chapter</u>		<u>Page</u>
	2.3.1 The Role of Data Collectors in the Client Survey	32
	2.3.2 Assigning the Survey Day.....	32
	<i>Step 1: Assigning Each Sampled Program to a Release</i>	
	<i>Group by Block of Weeks</i>	33
	<i>Step 2: Assigning a Survey Day</i>	33
	2.3.3 Sampling Clients during the Survey Day/Hours	34
	2.3.4 Data Required for Determining the Client Sampling Rate.....	35
	2.3.5 Special Situations Involving Client Sampling	36
	<i>Programs with Multiple Sites in Operation during the</i> <i>Survey Window</i>	36
	<i>Mobile Programs</i>	36
	<i>Limits to the Lengths of Data Collectors' Shifts</i>	37
	<i>Multiple Visits to Programs by the Same Client</i>	38
	2.3.6 Handling Missed Survey Days	38
	<i>Staffing/ Resource Issues</i>	39
	<i>Weather-related Impediments</i>	39
	<i>Other Events Affecting Client Demand Levels</i>	39
3	Weighting and Estimation.....	40
	3.1 Introduction.....	40
	3.2 General Approach for Computing Unduplicated Client Count Estimates.....	40
	3.2.1 Dual-Frame Sampling and Estimation	40
	3.2.2 General Form of the Estimator.....	42
	3.2.3 Estimation of Duplicated Client Counts (Client Visits).....	43
	3.2.4 Estimation of Unduplicated Client Counts	43
	3.2.5 Estimation at Different Levels	44
	3.2.6 Estimation over Different Time Periods	46
	3.2.7 Client Survey Items Used in Computing Unduplicated Client Count Estimates	47
	3.3 General Methodology for Computing Survey Weights.....	48

<u>Chapter</u>	<u>Page</u>
3.3.1 Base Weights	49
3.3.2 Adjusting for Nonresponse.....	49
3.4 Details of the Computation of Survey Weights for the Agency Survey and the Client Survey	52
3.4.1 Agency Survey Weights	53
3.4.2 Client Survey Weights	53
3.5 Standard Error Computation	58
3.6 Survey Weights Provided in Analysis Files	60
3.6.1 Weights for Analysis of Agency Survey Data.....	60
3.6.2 Weights for Analysis of Client Survey Data	60

<u>Tables</u>	<u>Page</u>
1-1. Food banks in the Feeding America network, with food bank level poundage and HIA 2014 participation status.....	3
2-1. Nominal and targeted released Client Survey sample sizes and associated assumptions	12
2-2. Program eligibility by detailed program type.....	14
2-3. Effective and nominal sample sizes to attain CVs of 10, 15, and 20 percent for an estimate of a proportion of 0.5	16
2-4. Expected CV of estimates of a proportion of 0.5 at the national and food bank levels, under various assumptions about the DEFF	17
2-5. Effective and nominal sample sizes for hypothesis testing of two proportions	17
2-6. Food bank-level counts of numbers of programs originally sampled, added to sample, and subsampled out.....	27

Contents (continued)

Tables (continued) **Page**

3-1. Overall weighted response rates at each stage of data collection 58

Exhibit **Page**

3-1. List of weight variables included in the Client Survey analysis files..... 61

This report provides technical details on the sample design and selection, weighting, and estimation processes for the Hunger in America 2014 (HIA 2014) study. Chapter 1 provides an introduction to the HIA 2014 study, including its target population and design objectives. Chapter 2 describes the sample design and selection for the Client Survey. Details of the approach for weighting and estimation are given in Chapter 3.

1.1 Overview of the Hunger in America 2014 Study

Feeding America, the nation's premier hunger relief charity, works with its network of more than 200 food banks to provide food and grocery products to agencies and their programs that serve people in need of food assistance across the United States. Hunger in America 2014 (HIA 2014) is being used to characterize the current efforts of the Feeding America network and provide a profile of the individuals and households served. Specifically, HIA 2014 describes the partner agencies and their programs that provide assistance; yields estimates of the numbers of duplicated and unduplicated clients served by these food assistance programs; and delves into the characteristics, circumstances, and coping strategies of clients served by these agencies.

The Feeding America network is structured hierarchically. Food banks receive supplies both from Feeding America's national organization and from local sources, and they store these food and grocery items for distribution to their partner agencies. A single food bank can have hundreds of partner agencies. Agencies partnered with the food banks are nonprofit organizations, typically 501(c)(3)s or nonprofit religious organizations. Agencies obtain food from the food banks, and provide meals or groceries directly to those in need through running or sponsoring programs that provide emergency or non-emergency food assistance. Agencies can vary substantially in size, with some operating a single program and others operating multiple programs. HIA 2014 engaged all levels of the hierarchy in an effort to obtain information about hunger relief agencies and the people they serve.

Data collection for HIA 2014 began with the engagement of food banks. Feeding America asked each participating food bank¹ to provide a list of agencies served by the food bank. These lists formed the basis of the Agency List. Each agency in the Agency List was sent an invitation to participate in the Agency Survey (i.e., the Agency Survey was designed to be a census of agencies in participating food banks). The Agency Survey provided key information about the agencies and their programs; this information was used for analytic purposes (to characterize the agencies and

¹ As discussed in Section 1.2, a small number of food banks in the Feeding America network declined to participate in HIA 2014.

programs in the Feeding America network) as well as for sample design and selection for the Client Survey. Finally, a sample of programs administered or supported by agencies in the Feeding America network was selected, and clients receiving food services from those programs were sampled to be interviewed for the Client Survey. The Client Survey captured information about the characteristics of the clients served by these programs (demographics, health status and coverage, living situation, income and financial hardships, SNAP participation, food insecurity), their frequency of use of these programs, and their experiences related to obtaining food from these programs.

Because the Agency Survey was designed to be a census of agencies in food banks participating in HIA 2014², most of this chapter focuses on sample design and selection for the Client Survey.

1.2 Target Population and Survey Population

For the Agency Survey, the target population was the set of agencies partnering with food banks in the Feeding America network.

The target population for the Client Survey was adults 18 years or older who received food from meal programs and households that received food from grocery programs, where in both cases these programs are operated by agencies that are partnered with the Feeding America food banks.

For the Client Survey, the survey population excludes clients served by agencies in food banks that chose not to participate in HIA 2014 (either the full study or the Client Survey activities); however, the weights used to produce national estimates included an adjustment for food bank-level nonresponse, so that these nonparticipating food banks are represented in the national estimates. Table 1-1 lists the food banks and Partner Distribution Organizations (PDOs)³ that acted as food banks for the purpose of the HIA 2014 study and gives, for each, the 2012 food bank poundage figure used for weighting (derived using a combination of the 2012 Network Activity Report (NAR) and the agency-level poundage provided in the agency list), as well as the food bank's participation statuses in the HIA 2014 Agency Survey and Client Survey.

² As stated, the Agency Survey was intended to be a census. However, even after cleaning of the Agency List, Westat expects that there may still have been programs erroneously included on the list, as well as agencies missing from the list. This issue is discussed later where relevant within the context of this report.

³ While PDOs are not direct members of the Feeding America network, they are independent nonprofit organizations contracted to fulfill certain food banking responsibilities, such as product distribution management and food solicitation within a portion of a member's service area.

In addition to the exclusion from the Client Survey effort of programs that serve only children, programs that serve only persons with severe cognitive or mental health disabilities, as well as domestic violence shelters, were excluded. Within eligible programs, clients with severe cognitive or mental health disabilities were also excluded. Although there may have been clients who were physically unable to complete the Client Survey (e.g., some clients who are blind and amputees who are unable to use a computer touchscreen), such clients remain part of the target population and were treated as nonrespondents in estimation.

Table 1-1. Food banks in the Feeding America network, with food bank level poundage and HIA 2014 participation status

Food bank ID	Food bank poundage (2012) used for weighting	Food bank participated in HIA 2014 Agency Survey?	Food bank participated in HIA 2014 Client Survey?
001	27,211,513	Yes	Yes
002	7,238,956	Yes	Yes
003	9,127,597	Yes	Yes
004	33,478,591	Yes	Yes
005	13,499,593	Yes	Yes
006	26,040,271	Yes	Yes
007	22,392,393	Yes	Yes
008	7,780,572	No	No
009	38,718,153	Yes	Yes
010	65,439,562	Yes	Yes
011	10,860,609	Yes	Yes
012	10,757,885	Yes	Yes
014	36,506,930	Yes	Yes
015	11,114,262	Yes	Yes
016	62,703,841	Yes	Yes
017	19,010,006	Yes	Yes
018	33,647,321	Yes	Yes
019	13,407,054	Yes	Yes
020	27,638,595	Yes	Yes
021	43,805,317	Yes	Yes
022	38,126,005	Yes	Yes
023	5,935,911	Yes	Yes
024	46,668,119	Yes	Yes
025	39,068,186	Yes	Yes
026	7,673,765	Yes	Yes
027	43,951,522	Yes	Yes
028	7,496,885	Yes	Yes
029	9,710,467	Yes	Yes
030	18,224,077	Yes	Yes
031	25,240,948	Yes	Yes
033	10,802,615	No	No
034	22,731,911	Yes	Yes
035	6,694,321	Yes	Yes
036	14,529,781	Yes	Yes
037	69,161,179	Yes	Yes
038	20,230,226	Yes	Yes

Table 1-1. Food banks in the Feeding America network, with food bank level poundage and HIA 2014 participation status (continued)

Food bank ID	Food bank poundage (2012) used for weighting	Food bank participated in HIA 2014 Agency Survey?	Food bank participated in HIA 2014 Client Survey?
039	19,099,765	Yes	Yes
040	39,198,694	Yes	Yes
041	15,539,130	Yes	Yes
042	26,942,049	Yes	Yes
043	14,294,169	Yes	Yes
045	59,532,816	Yes	Yes
046	15,492,224	Yes	Yes
047	7,407,418	Yes	Yes
048	30,294,279	Yes	Yes
049	10,895,077	Yes	Yes
050	30,439,664	Yes	Yes
051	13,953,915	Yes	Yes
052	15,510,825	Yes	Yes
053	6,578,665	Yes	Yes
054	18,467,462	No	No
055	19,478,331	Yes	Yes
056	17,083,103	Yes	Yes
057	22,284,398	Yes	Yes
058	45,921,292	Yes	Yes
059	5,385,604	Yes	Yes
060	15,389,763	Yes	Yes
062	38,021,076	Yes	Yes
063	9,541,931	Yes	Yes
064	17,129,104	Yes	Yes
065	39,804,248	Yes	Yes
066	15,504,804	No	No
068	59,900,488	No	No
070	43,365,387	Yes	Yes
072	17,065,482	Yes	Yes
073	18,342,473	No	No
074	16,333,724	Yes	Yes
075	8,366,063	Yes	Yes
076	41,332,965	Yes	Yes
077	20,945,952	Yes	Yes
078	45,226,233	Yes	No
079	44,766,875	Yes	Yes
080	9,445,634	Yes	Yes
081	35,330,807	Yes	Yes
082	9,893,412	Yes	Yes
083	20,281,063	Yes	No
084	6,974,538	Yes	Yes
085	23,102,109	Yes	Yes
086	65,186,808	Yes	Yes
087	17,130,658	Yes	Yes
089	11,724,135	Yes	No
090	28,526,519	Yes	Yes
091	24,066,768	Yes	Yes
092	15,466,233	Yes	Yes

Table 1-1. Food banks in the Feeding America network, with food bank level poundage and HIA 2014 participation status (continued)

Food bank ID	Food bank poundage (2012) used for weighting	Food bank participated in HIA 2014 Agency Survey?	Food bank participated in HIA 2014 Client Survey?
093	20,915,518	Yes	Yes
094	9,685,280	Yes	Yes
095	31,997,119	Yes	Yes
096	9,342,549	Yes	Yes
097	22,295,632	Yes	Yes
098	3,110,221	Yes	Yes
100	19,504,261	Yes	Yes
102	9,919,468	Yes	Yes
103	10,063,056	Yes	Yes
105	10,095,559	Yes	Yes
106	6,100,838	Yes	Yes
109	11,412,563	Yes	No
110	7,537,786	Yes	Yes
112	34,399,725	Yes	Yes
113	6,549,267	No	No
114	2,417,249	Yes	Yes
116	25,216,794	Yes	Yes
117	5,285,648	Yes	Yes
119	4,379,807	Yes	Yes
121	5,650,202	Yes	Yes
123	4,774,005	Yes	Yes
124	11,401,122	Yes	Yes
126	7,261,964	Yes	Yes
129	6,613,885	Yes	Yes
133	1,189,373	Yes	Yes
134	7,842,160	Yes	Yes
135	10,519,218	No	No
136	3,324,909	Yes	Yes*
137	7,825,047	Yes	Yes
138	5,243,445	Yes	Yes
139	4,876,975	No	No
140	4,442,689	Yes	Yes
141	6,222,623	Yes	Yes
144	11,749,479	Yes	No
145	4,750,068	Yes	Yes
146	5,937,580	Yes	Yes
149	15,750,148	Yes	Yes
151	4,448,888	Yes	Yes
152	17,974,008	Yes	Yes
154	5,614,459	No	No
155	24,970,732	Yes	Yes
156	22,496,870	Yes	Yes
157	8,197,945	Yes	Yes
159	7,167,116	Yes	Yes
160	9,803,009	Yes	Yes
161	7,727,630	Yes	Yes
164	19,512,568	Yes	Yes
167	6,694,405	Yes	Yes
168	9,228,613	Yes	Yes
169	23,208,211	Yes	Yes

Table 1-1. Food banks in the Feeding America network, with food bank level poundage and HIA 2014 participation status (continued)

Food bank ID	Food bank poundage (2012) used for weighting	Food bank participated in HIA 2014 Agency Survey?	Food bank participated in HIA 2014 Client Survey?
170	10,090,347	Yes	Yes
174	2,685,140	Yes	Yes
175	5,842,253	Yes	Yes
177	7,453,462	Yes	Yes
178	5,940,588	Yes	Yes
184	13,439,669	Yes	Yes
185	17,297,328	Yes	Yes
186	5,276,101	Yes	Yes
187	5,246,997	Yes	Yes
188	12,191,621	Yes	Yes
189	8,569,228	Yes	Yes
191	3,840,413	Yes	Yes
192	11,704,621	Yes	Yes
193	4,579,498	Yes	Yes
195	5,407,102	Yes	Yes
196	2,492,797	Yes	Yes
197	6,115,736	Yes	Yes
199	10,646,733	No	No
201	6,879,610	Yes	Yes
202	7,335,723	Yes	Yes
206	4,488,392	Yes	Yes
207	5,288,349	Yes	Yes
208	11,225,739	Yes	Yes
209	3,048,477	Yes	Yes
214	9,593,557	Yes	Yes
216	16,366,641	Yes	Yes
218	4,336,751	Yes	Yes
219	7,609,471	Yes	Yes
220	7,459,414	Yes	Yes
221	12,821,297	Yes	Yes
222	6,208,585	Yes	Yes
225	6,751,790	Yes	No
229	4,292,607	No	No
258	6,175,032	No	No
259	12,514,402	Yes	Yes
276	5,427,456	Yes	Yes
278	5,633,239	Yes	Yes
279	38,310,584	Yes	Yes
280	15,281,127	Yes	Yes
281	18,297,036	Yes	Yes
284	5,669,134	Yes	Yes
285	6,693,044	Yes	Yes
287	4,320,550	Yes	Yes
290	3,283,252	Yes	Yes
291	13,768,958	Yes	Yes
294	7,876,862	Yes	Yes
296	23,378,000	Yes	Yes
297	39,573,385	No	No
298	9,335,241	Yes	Yes
303	3,279,802	No	No

Table 1-1. Food banks in the Feeding America network, with food bank level poundage and HIA 2014 participation status (continued)

Food bank ID	Food bank poundage (2012) used for weighting	Food bank participated in HIA 2014 Agency Survey?	Food bank participated in HIA 2014 Client Survey?
305	5,219,297	Yes	Yes
306	37,779,552	Yes	Yes
309	6,494,866	Yes	No
310	7,045,355	Yes	Yes
311	7,321,899	Yes	Yes
313	966,953	Yes	Yes
314	1,370,608	Yes	Yes
316	2,266,831	Yes	Yes
319	62,917,522	Yes	Yes
320	7,351,760	No	No
323	4,044,428	Yes	Yes
324	8,952,914	Yes	Yes
326	6,151,308	Yes	Yes
339	3,299,930	Yes	Yes
369	5,814,108	Yes	Yes
516	1,298,338	Yes	No
540	1,488,468	Yes	Yes
559	1,920,093	Yes	Yes
563	7,945,487	Yes	Yes
603	5,408,200	Yes	Yes
611	5,521,380	Yes	Yes
626	1,455,648	Yes	Yes
627	9,433,385	Yes	Yes
684	10,115,891	Yes	Yes
685	1,700,489	Yes	Yes
779	4,390,679	Yes	Yes
780	4,053,251	Yes	Yes
781	1,246,239	Yes	Yes

*Participated in Client Survey but was treated as a nonparticipant for analysis purposes.

1.3 Design Objectives

The primary design goal for the HIA 2014 Client Survey was to select a probability sample of clients and collect data to support the following estimates:

- National estimates of the total number of clients served.** While duplicated client count estimates (reflecting the number of times clients were reached through food distribution) were of interest, an important aspect of the design was producing estimates of *unduplicated* total client counts over time (by week, by month, and annualized), both overall and by program type (namely, meal programs and grocery programs), reflecting the unique number of clients served.

- **Food bank level estimates of the total number of clients served.** As with the national estimates, both duplicated and unduplicated client count estimates were of interest (by week, by month, and annualized), overall and by program type. For the food bank level estimates of the total number of unduplicated clients, the desired precision was a coefficient of variation below 20 percent. The ability to meet this precision requirement was driven largely by the program sample size within each food bank.

While estimates pertaining to more detailed program types (e.g., shelter, pantry, (soup) kitchen; mobile programs; senior programs) may be produced using HIA 2014 data, the sample was not designed specifically to support these estimates; that is, while some such estimates may be produced with adequate precision, others may not be. This is because the sample was not stratified at the more detailed program type level and thus was not designed to yield any specific sample sizes at this level.

The sample design supports national estimates by program type (meal/grocery), but was not designed to support food bank-level estimates for these two broad program types. Again, while some such estimates may be produced with adequate precision, others may not be. This is because the sample was stratified by food bank, but sample sizes at the program type level within food bank were based on operational constraints, not statistical considerations.

1.4 Sampling Unit and Unit of Analysis

In considering the sampling unit and unit of analysis for the HIA 2014 Client Survey, it is necessary to consider the concept of the “client.” For meal programs, the client is the individual adult, and it is only necessary to account for services received by that adult. For grocery programs, however, the client is everyone in the household who may benefit from the groceries received, so it is necessary to account for services received by the respondent and/or anyone else in the respondent’s household. Thus, for meal programs, the sampling unit and the unit of analysis is the individual adult. For grocery programs, the sampling unit is the household; the unit of analysis is generally the household, although estimates of client counts may be obtained at either the household or individual level. Thus, when counting and sampling clients in grocery programs, households were treated as a single unit; i.e., if multiple members of a household reported to the grocery program at the same time, they (as a group) were only to have been counted once, and if the household was sampled, one household member served as the household respondent. The household member chosen as the respondent was an adult household member who could reliably respond for the household. It was not necessary that this household member be randomly selected, as the data collected in the Client Survey pertain to either the household (as a unit) or each household member.

Client Survey Sample Design and Selection

2

2.1 Overview of the Sample Design

To meet the design objectives, the sample design for the Client Survey was a multistage dual frame design. This section outlines these design features and the general approach for sampling agencies, programs, and clients. Subsequent sections provide more details on each stage of selection

2.1.1 Multistage Design

The finite population of clients in the nation is unknown and there is no national frame of clients. Therefore, the sample of clients was selected using a multistage design. In order to develop a representative sample and to select each client with a known probability of selection, the approach was a nested selection in four stages:

- Stage 1 involved selecting agencies from the Agency List provided by food banks.
- Stage 2 involved selecting a sample of programs within sampled agencies. For the second stage of selection, a list of programs was compiled within each sampled agency.
- Stage 3 involved assigning a sampled program to a “survey day/hours” (a span of hours within a day during the survey data collection period); this was done in a manner that aimed to spread data collection over the entire survey period and capture the ebbs and flows in the way that clients are served with respect to hours of the day, days of the week, and weeks of the month.
- Stage 4 involved sending trained volunteer data collectors to the sampled program site on the survey day. The data collectors maintained a complete tally of all clients served during the survey hours and they were provided with the protocol for selecting a random sample of clients to complete the Client Survey (a systematic sample that was based on a random start and a sampling interval provided to the data collectors).

The sampling method at each sampling stage was designed to accommodate specific design objectives.

2.1.2 Dual Frame Sampling

For the stage 1 sample of agencies, a dual frame approach was used to facilitate estimation. (See Chapter 3 for details.) The dual frame approach used two agency frames – a list of agencies that offer meal programs (the meal frame) and a list of agencies that offer grocery programs (the grocery frame). Within a given food bank, the sampling of agencies in each frame was independent. Agencies in the meal frame were sampled for their meal programs and clients who receive meals at these programs. Agencies in the grocery frame were sampled for their grocery programs and clients who receive groceries at these programs. Agencies that offered both types of programs were included on both lists. Also, programs that offered both types of services were included on both lists. Depending on whether the agency was sampled from the meal or the grocery frame, the subsequent sampling of programs and clients was for the same service type. The exception is that some programs were misclassified because the agency reported an incorrect program type on the Agency Survey; in these cases, while the program was sampled based on the program type reported in the Agency Survey, the Client Survey administration was designed to take into account the correct program type. The dual frame estimator was modified to take this misclassification into account. With this approach, it was possible that an agency that offers programs with both types of service could have been sampled from both frames. In fact, it was possible that the same program could have been sampled from both frames, if the program provided both types of service.

2.1.3 PPS Sampling

Both agencies and programs were sampled with probabilities proportionate to size (PPS). PPS selection was used at these early stages because, theoretically, it can be used to obtain an approximately self-weighting sample of clients (i.e., a client sample with approximately equal overall probabilities of selection) while also yielding, in expectation, an approximately fixed-size sample of clients in each program. The ideal measure of size (MOS) for this selection would be the number of clients served (either duplicated or unduplicated) during the HIA 2014 Client Survey data collection period. Clearly, that exact information was not available for use in selecting the agencies and programs for the Client Survey. In its absence, the goal was to use an MOS that is correlated with the number of clients served during the data collection period.

To illustrate using PPS sampling for selecting agencies and programs within agencies, the overall sampling fraction f_h for sampling programs in food bank h is the following:

$$f_h = \frac{n_h MOS_i}{\sum_{i \in h} MOS_i} * \frac{n_i MOS_j}{\sum_{j \in i} MOS_j}$$

where n_h is the number of agencies to sample in food bank h , n_i is the number of programs to sample in agency i , and MOS_i and MOS_j denote the measures of size for agencies and programs, respectively. The first term in this product is the selection probability for agency i in food bank h and the second term is the conditional selection probability for program j within sampled agency i (and for ease of presentation, the “ i ” subscript is not included in the term involving j ; likewise, the subscript “ h ” is not included in these terms, but all sampling was within food bank). Technically, since a dual frame approach was used, and sampling from the two frames was independent, the terms in the expression above should also be indexed by the frame. However, for ease of illustration, we have eliminated the subscript that indexes the frame. (The subscript h could be interpreted to denote the frame by food bank combination.)

2.1.4 Stratification

Stratification is often used to improve the precision of survey estimates, to ensure sufficient sample sizes for domain estimation, and/or to ensure representation of sampling units with specific characteristics. Precision gains may be realized by forming strata that comprise homogeneous subsets of sampling units. In the HIA 2014 sample design, food banks comprised the strata in each of the two agency sampling frames.

2.1.5 Target Sample Sizes and Related Assumptions

The original sample size targets for the Client Survey were 15,000 participating eligible agencies and 70,000 completed client surveys, and the sample design originally aimed to achieve these sample sizes. Calculations in this section are based on these values. However, the final allocation of the agency sample had a total number of sampled agencies that was not designed to yield exactly 15,000 participating eligible agencies and 70,000 completed client surveys. (The final allocation and its implications are discussed further in Section 2.2.4.)

In an effort to attain the required number of completed client surveys, the actual sample sizes to release for data collection needed to be inflated to take into consideration potential sample losses at each sampling stage. The right side of Table 2-1 shows eligibility and response rate assumptions. The

left side of Table 2-1 gives the nominal expected sample sizes required to attain 70,000 completed client surveys based on these assumptions.

Table 2-1. Nominal and targeted released Client Survey sample sizes and associated assumptions

Sample size	Number	Assumed adjustment factors	Percent
Participating food banks	193		
Nominal participating eligible agencies	15,000	Agency/program eligibility rate	90.0
Nominal complete client surveys	70,000	Agency/program participation rate	90.0
Release sample size for agencies*	18,519	Client eligibility rate	95.0
Release sample size for programs	18,519	Client completion rate	75.0
Release sample size for clients**	98,246		

*The same agency may be counted twice here, if it was sampled from both the meal frame and the grocery frame.

**Target number of clients sampled in eligible, participating programs.

For the purpose of sampling clients for the Client Survey, data from the Agency Survey was used to identify those agencies with food programs eligible to be included in the Client Survey for HIA 2014; agencies and programs identified (based on these Agency Survey items) as ineligible were excluded from the sampling frames. Table 2-2 shows, for each detailed type of program, whether the program type was eligible. Some detailed program types (such as rehabilitation programs, residential programs, shelter, and transitional housing) may or may not have been eligible depending on the types of clients they served. Other programs such as community kitchens were eligible if their primary function was to serve meals or distribute groceries. They were not eligible if their primary function was something other than serving food (e.g., serving as workforce training and development program). Programs for children (e.g., Day Care, Afterschool Snack, Backpack, kids café, school pantry), food related benefits programs (e.g., SNAP, WIC Outreach, Nutrition Education), home delivery programs (e.g., Meals on Wheels), and community gardens were not eligible for the Client Survey. Mobile programs (e.g., mobile pantries and markets, and senior mobile programs) were newly eligible program types in HIA 2014.

For programs having detailed program types listed as eligible or “potentially eligible” in Table 2-2, those with CHILDREN_ONLY = 1 (from item Q32.d in the Agency Survey⁴) were treated as ineligible; consideration was given to treating those with ELIGIBLE_OTHER = 1 (from item Q30A.e in the Agency Survey) as ineligible, but due to concerns about misinterpretation of this item, the final decision was to not use ELIGIBLE_OTHER in determining eligibility. Ineligible programs were eliminated during the construction of the sampling frame for agency and program selection. If

⁴ Specifically, a “yes” response to item Q32.d as well as “no” responses to items Q32.a, Q32.b, and Q32.c is required; the latter condition was added due to concerns that some respondents may have misinterpreted Q32.d as asking whether the program serves children under 18, rather than asking whether the program serves only children under 18.

any of the required data for determining this status was missing, the program (and associated agency) was classified as having unknown eligibility and was retained in the sampling frame.

The eligibility status of an agency was linked to the program(s) it operated. An agency was:

- Eligible if it operated at least one program that was eligible for the Client Survey;
- Ineligible if none of the programs it operated were eligible for the Client Survey.

Even with the exclusion of agencies/programs that could clearly be identified as ineligible based on data from the Agency Survey or Agency List, the sampling frames still included some agencies/programs whose eligibility status was unknown or whose status changed after completion of the Agency Survey. To account for these, a combined eligibility rate of 90 percent for agencies and programs was assumed. Additionally, an agency/program participation rate (among agencies that completed the Agency Survey) of 90 percent was assumed. By selecting a sample of 18,519 agencies and sampling one eligible program in each sampled agency (with the exception of agencies sampled from both the meal frame and the grocery frame, for which both a meal program and a grocery program were sampled), a nominal sample of 15,000 eligible participating agencies/programs would be attained under these assumptions. (As discussed above, the final agency sample size was not 18,519; it was later determined, during the allocation of the agency sample to the food banks, that this original target could not be attained due to resource limitations.)

Table 2-2. Program eligibility by detailed program type

Program type	Detailed program type	Eligible for client data collection
Grocery	(Food) Pantries	Yes
Meal	(Soup) Kitchen	Yes
Grocery	Commodity Supplemental Food Program (CSFP)	Yes
Meal	FB operated meal program	Yes
Grocery	FB operated pantry program	Yes
Grocery	Mobile Pantries / Mobile Markets	Yes
Grocery	Senior Brown Bag/Food Box Distribution	Yes
Meal	Senior Congregate Meal	Yes
Grocery	Senior Grocery Program	Yes
Grocery	Senior Mobile Pantry/Just-in-Time Delivery	Yes
Grocery	Other Pantry Program	Potentially eligible
Meal	Community Kitchens	Potentially eligible
Meal	Group Home	Potentially eligible
Meal	Rehabilitation Program	Potentially eligible
Meal	Residential Program	Potentially eligible
Meal	Shelter	Potentially eligible
Meal	Transitional Housing	Potentially eligible
Meal	Afterschool Snack	No
Grocery	BackPack Programs	No
Meal	Child Congregate Feeding Programs (non-Kids Cafe)	No
Grocery	Community Gardens	No
Meal	Day Care	No
Grocery	Home Delivered Grocery Program	No
Meal	Home Delivered Meal (or Meals on Wheels)	No
Meal	Kids Cafe Programs	No
Food related benefits	Nutrition Education	No
Grocery	School Pantry Programs	No
Food related benefits	SNAP - Assistance/Outreach	No
Meal	Summer Food Service Program (SFSP)	No
Food related benefits	WIC Outreach	No
Food related benefits	National School Breakfast Program	No
Food related benefits	National School Lunch Program	No

To compute the number of clients to be sampled, the target number of completed client surveys was adjusted by both an expected client eligibility rate and an expected client response rate. The client eligibility rate was needed to account for the losses when some sampled clients who could not be readily identified by the data collector as ineligible prior to sampling were found to be ineligible to complete the survey (e.g., due to a particular cognitive or mental health disability or being under age 18). The client response rate was needed to compensate for losses from client refusals, language problems, or other conditions that lead to nonresponse. With an assumed client eligibility rate of 95 percent and an assumed client response rate of 75 percent, a sample of 98,246 clients was needed to yield 70,000 completed client surveys. Based on these assumptions, at each participating program, an average of 6.5 clients would need to be sampled, and an average of 4.7 client surveys would be completed. Since the number of programs sampled changed as a result of Feeding America's concerns about the food banks' abilities to support data collection in only a limited number of

programs, these numbers were revised. During agency and program sampling, a total of 16,869 programs were sampled. With this program sample size, at each participating program, an average of 7.2 clients needed to be sampled, and an average of 5.1 completed client surveys were expected.

A reserve sample of programs was set aside for possible release if the sample yield fell considerably short of targets. Following the original selection, some programs for the reserve sample were released into the sample.

2.1.6 General Precision Considerations

When considering the precision requirements of a complex sample survey such as HIA 2014, one must take into account the design effect (DEFF), or increase in variance relative to a simple random sample (SRS) of the same size. The DEFF is a function of various factors including the effect of clustering the sample, the effect of stratification, and the extent to which the survey weights vary (unequal weighting effect). Because of the clustering that was used to select clients within agencies and programs, the dual frame approach that was used, weights that need to be used to account for differential client selection probabilities due to inaccurate measures of size and due to limits imposed on data collection at a program, and adjustments that will be made during estimation to adjust for nonresponse, the effective sample sizes for analyses have to take into consideration the DEFF.

The optimum cluster size of clients per agency/program depends on the degree of homogeneity within clusters (the coefficient of intraclass correlation). A smaller cluster size was recommended for HIA 2014 following evaluations of the HIA 2010 design. To illustrate the gain, assuming a modest intraclass correlation of 0.02, a cluster size of 5 will result in relatively small DEFF of 1.08, or an 8 percent increase in variance relative to a SRS; in contrast, when the cluster size is 10, the DEFF is 1.18, or an 18 percent increase in variance relative to the SRS.

To illustrate, Table 2-3 shows the effective and nominal sample sizes required to achieve coefficients of variations (CV) of 10, 15 or 20 percent for an estimated proportion at 0.50. Under the assumption of a SRS design, a sample size of 100 completed interviews is needed to achieve a CV at 10 percent, a sample size of 45 is needed to achieve a CV at 15 percent, and a sample size of 25 is needed to achieve a CV at 20 percent. For a sampling and estimation scheme of the type that was used for the Client Survey (i.e., taking into account not only the effect of clustering, but also the effects of differential selection probabilities and adjustments for nonresponse, as well as the dual-frame estimation approach), the DEFF for various estimates may be in the range of 2 to 4 (or

higher). Then the nominal sample size should be two to four times those for the SRS design to achieve the same precision. For example, with an assumed DEFF=4, a nominal sample size of 180 completed surveys is needed to achieve a CV of 15 percent.

Table 2-3. Effective and nominal sample sizes to attain CVs of 10, 15, and 20 percent for an estimate of a proportion of 0.5

Sample sizes	An estimated proportion of 0.5 coefficient of variation (CV%)		
	10%	15%	20%
Effective completed surveys	100	45	25
Nominal completed surveys (assumed DEFF=2)	200	90	50
Nominal completed surveys (assumed DEFF=4)	400	180	100

Table 2-4 gives the expected precision, under different assumptions about the level of the DEFF, of an estimated proportion of 0.5; CVs are tabulated at the national overall level (i.e., for an estimate computed using the full client sample, $n = 70,000$), for national subgroups of varying sizes, and for food banks under various assumptions about the sample sizes within the food bank ($n = 700, 500, 300, \text{ and } 100$). In general, larger design effects would be expected at the national level than at the food bank level, due to the expected larger variation in client selection probabilities nationally than within food banks.

Table 2-5 shows the effective and the nominal responding sample sizes required for hypothesis testing to detect differences between two subgroup proportions. For comparisons of two hypothetical proportions (say, e.g., 0.50 vs 0.54; 0.50 vs 0.56; and 0.50 vs 0.60) at 0.80 power level and at a significance level of 0.05, the minimum effective sample sizes required in each cell involved in the comparison are displayed in the first row. For example, to detect the difference between a proportion of 0.50 in one subgroup and a proportion of 0.60 in another group would require an effective sample of at least 388 completed surveys in each of the two subgroups. At an assumed DEFF=2, twice as many (nominal) completed surveys would be required in each subgroup.

Table 2-4. Expected CV of estimates of a proportion of 0.5 at the national and food bank levels, under various assumptions about the DEFF

Level and sample size	Expected coefficient of variation (CV%) for an estimated proportion of 0.5		
	DEFF = 1	DEFF = 2	DEFF = 4
National (n = 70,000)	0.4	0.5	0.8
National subgroup (n = 20,000)	0.7	1.0	1.4
National subgroup (n = 10,000)	1.0	1.4	2.0
National subgroup (n = 1,000)	3.2	4.5	6.3
Food bank (n = 700)	3.8	5.3	7.6
Food bank (n = 500)	4.5	6.3	8.9
Food bank (n = 300)	5.8	8.2	11.5
Food bank (n = 100)	10.0	14.1	20.0

Table 2-5. Effective and nominal sample sizes for hypothesis testing of two proportions

Sample sizes	Sample sizes required in each subgroup (assuming 0.8 power and significance level of 0.05) to detect a difference between a proportion of 0.5 and a second proportion of:		
	54%	56%	60%
Effective completed surveys	2,448	1,086	388
Nominal completed surveys (assumed DEFF=2)	4,896	2,172	776
Nominal completed surveys (assumed DEFF=4)	9,792	4,334	1,552

2.2 Sampling Agencies and Programs

A general overview of the sample design was given in Section 2.1. In this section, details of specific aspects of the sampling of agencies and programs are discussed.

2.2.1 Sampling Frames

In this section, the creation of sampling frames is discussed. First, the construction of the agency frames is described. Next, the construction of lists of programs within sampled agencies is considered. Finally, the implications of missing data in key frame variables are discussed.

Agency Frames

The sources of data for building the agency sampling frames included (1) items from the Agency Lists provided by the food banks and (2) Agency Survey questionnaire items for agencies participating in the Agency Survey. Two agency frames were created:

- The meal frame included all agencies that operated one or more meal programs.
- The grocery frame included all agencies that operated one or more grocery programs.

Agencies that include both types of programs appeared in both frames. The key agency-level data items in the agency frames included identifiers; total poundage for the agency as recorded in the Agency List (i.e., total poundage provided by the primary food bank); and the agency's proportion of poundage that is provided by the primary food bank as reported in the Agency Survey (item Q2.a).

Some agencies receive food from more than one food bank. (These are referred to in this document as “shared agencies”.) In those cases, in an effort to avoid duplication, a food bank was instructed to list the agency if the given food bank is the primary food bank in terms of support to the agency, and not to list the agency otherwise.

During data collection for the Agency Survey, it was determined that in some cases, food banks had listed programs rather than agencies in providing their agency lists. Efforts were undertaken to repair the Agency List and administer new Agency Surveys to agencies that were originally missing from the list. However, it is likely that even with these efforts, the final Agency List included some programs listed as agencies. The two-stage process of sampling agencies and then sampling one program within each sampled agency thus became, effectively, a one-stage process in the case of programs listed as agencies. The sampling implication was that it was possible for more than one program from the same agency to have been sampled from the same frame. This could increase the design effect associated with clustering; however, such occurrences were rare, so any such increase in design effect would be expected to be small. The implication for analysis of the Agency Survey is that some items intended to be agency-level may have been misreported at the program level, and/or multiple programs within the same agency may have reported the same agency-level data.

Lists of Programs in Sampled Agencies

To sample programs in the sampled agencies, the data were obtained from the Agency Survey section in which agencies listed the programs they operated. Following the path of the agency sample, for agencies sampled in the grocery frame, the program lists included programs that distributed groceries. Likewise, for those agencies sampled in the meal frame, the program list included programs that served meals. The designation of program type (meal or grocery) was based on the detailed program type reported in the Agency Survey, using the classifications given in Table 2-2. In some cases, the incorrect program type was reported in the Agency Survey and this sometimes led to programs having been assigned to the wrong frame. The program-level data items in these lists included the items from the Agency Survey such as program name and detailed program type, days and times the program distributes groceries/serves meals, percentage of food from the primary food bank, and monthly counts (duplicated and unduplicated) of clients reported by the survey respondent as served by the program.

Missing Data in the Sampling Frames

Despite efforts aimed at capturing items that were critical for sampling purposes, some level of item nonresponse to critical items resulted. Consideration was given to excluding (as nonrespondents) from the sampling frames agencies/programs with missing data for the critical sampling items, if the proportion of such agencies/programs was very low. However, if that proportion was not very low, then such exclusions could restrict the number of agencies/programs available to participate in the Client Survey and could also result in substantial bias in the estimates (if the characteristics of agencies/programs that are missing these items or the clients they serve differ from the characteristics of agencies/programs that are not missing these items or the clients they serve). Thus, the final decision was to address missingness in the critical sampling items through imputation.

With the decision to tolerate missingness in the critical sampling items, it was necessary to specify criteria for determining when the Agency Survey for a particular agency was to be considered “complete” for sampling purposes. To be considered “complete” for sampling purposes, the Agency Survey data had to meet the following criteria:

- There must be at least one program-level record (i.e., program-level data for at least one program associated with the agency);

- the program name and detailed program type must have been reported (i.e., not missing); and
- there must be valid reported data for at least one critical sampling item other than detailed program type.

The following procedures were used to impute for the missing critical sampling items:

- Missing poundage from the agency list (*Valid poundage information was missing from the agency list for 2.8 percent of all agencies, and 3.7 percent of agencies in the final agency sampling frame. Although the overall missingness rate for agency-level poundage was only 3.7 percent, this varied from 0 missing in several food banks to 38.2 percent missing.*):
 - If client counts were available, the missing poundage for agency i was imputed as $\bar{r}x_i$, where \bar{r} is the overall ratio of poundage to total client count for agencies for which complete data were available and x_i is the total client count for agency i .
 - If client counts are not available, the missing poundage was imputed as the median poundage among agencies in the same food bank.
- Missing proportion of total agency poundage provided by the primary food bank:
 - This was imputed using the median among agencies in the same food bank.
- Missing monthly program-level client counts (*Among programs included in the program sampling frame, 58 percent were missing at least some information about monthly duplicated and/or unduplicated client counts, and 24 percent had no information at all about either monthly duplicated client counts or monthly unduplicated client counts.*):
 - First, these were imputed using the mean of the program's counts for the two adjacent months (if available).
 - If one set of client counts was missing (either the unduplicated counts from Q31C or the duplicated counts from Q31D) for a given program, it was imputed using the overall ratio of unduplicated to duplicated client counts for the program type (meal/grocery). That is, among programs for which both sets of client counts were available, the monthly counts of each type were summed and the ratio was computed. Then, for programs that were missing the unduplicated counts, the duplicated counts were multiplied by this ratio. For programs that were missing the duplicated counts, the unduplicated counts were multiplied by the reciprocal of this ratio.
 - If both the unduplicated client counts (from Q31C) and the duplicated client counts (from Q31D) were missing for a given program, hot deck imputation was used to impute counts from a donor with the same detailed program type.

- While considerable variation in actual numbers of sampled clients was anticipated, the lack of any information about the size of the program was particularly problematic in this regard.
- Missing days/times the program distributes groceries/serves meals (*Days of operation data in the Agency Survey were consistent enough--in terms of providing responses to all of the relevant questions in the applicable path of questions--to permit calculation of monthly days of operation for only 52 percent of programs listed in the Agency Survey.*):
 - For programs with missing information about days of operation, the implicit assumption (used to assign the survey day) was that the program serves clients every day during the data collection period. Rules were established for determining a replacement day, and these rules were used in the event that the program was not open on the assigned survey day.
 - For programs with missing information about the times the program serves clients on the designated survey day, fall-back rules were applied to determine the survey window.
 - While these operational rules were sufficient for data collection purposes, missing or invalid days of operation data were problematic for estimation purposes. Thus, the food banks were asked to review the monthly days of operation counts and either confirm or correct these counts. Inaccurate days of operation counts could introduce substantial bias into client count estimates because error in these counts would be expected to translate to proportionate error in client count estimates. For example, a monthly days open count that accounts for only half the days the program was actually open that month would translate to a monthly client count estimate (for that program) that is half of what it would be if an accurate monthly days open count had been reported.
 - For weighting and estimation purposes, the monthly days of operation counts were imputed when necessary, as follows. Using the set of programs that either provided monthly days of operation counts in the Agency Survey or provided these when given the opportunity (during the Client Survey data collection period) to confirm/correct the days of operation data, we computed the mean days open for each month by detailed program type (e.g., CSFP, community kitchens, FB operated meal program, etc.). For all programs whose agency did not provide usable days of operation data for the program in the Agency Survey and for which the days open counts were not confirmed/corrected, we used these monthly means at the detailed program type level as the final days open counts.

2.2.2 Measures of Size and Probabilities of Selection

As discussed in Section 2.1.3, PPS sampling was used for both the selection of agencies and the selection of one program within each sampled agency. For sampling agencies and programs, the

MOS should be strongly correlated with the total number of clients served by meal and grocery programs. For sampling agencies, HIA 2010 had used a function of the annual poundage of food from the food bank. While this measure was a reasonable choice, Feeding America expressed a need to take a new approach because (1) some agencies can receive food from more than one food bank, and (2) in many food banks, the distribution of food across agencies is highly skewed.

In HIA 2014, additional information was collected in the Agency Survey to better reflect the sizes of agencies and programs. For agency sampling, the MOS was based on adjusted poundage, i.e., total poundage provided by the primary food bank (from the Agency List) divided by the proportion of the agency's total pounds that were provided by the primary food bank (from item Q2.a in the Agency Survey). Adjusted poundage is an estimate of the total poundage of food the agency gave out or served to clients over the past 12 months. For agencies with only meal or grocery programs, the MOS was the adjusted poundage. For agencies with both meal and grocery programs, the total adjusted poundage was allocated to the two frames based on the relative sizes of the total duplicated client counts in order to compute the MOS of the agency for the particular frame.

For the sampling of one program within each sampled agency, the MOS was the sum of the monthly duplicated client counts (from item Q31D in the Agency Survey).

2.2.3 Identification of Certainty Agencies and Undersized Agencies and Programs

Under this PPS sampling scheme, agencies with very large selection probabilities—those with MOS above a specified threshold—were designated for certainty selection. For each of the two frames, the threshold used for this was 75 percent of the initial sampling interval. (The initial sampling interval was computed as the total MOS for the given program type in the food bank divided by the number of agencies to be sampled within the food bank.)

In addition to these very large agencies, it would have been possible for some agencies to be designated for certainty selection irrespective of size. This option could have been used, for example, to ensure the inclusion in the sample of agencies that serve specific geographic areas or minority populations, or that have unique characteristics. This approach was considered for HIA 2014 but ultimately the decision was made to identify certainties based on MOS alone. (However, subsequent to the random selection of agencies and programs, some additional programs were added to the sample with certainty.)

The number of agencies to be sampled (in a noncertainty selection) was obtained by subtracting the certainty selections from the overall target number of sampled agencies. Small agencies/programs were deemed “undersized” if the conditional client selection probability needed in order to achieve an equal probability sample of clients would exceed 1.0. To avoid these situations (which would require weighting factors to be applied), the original plan was for these small agencies/programs to be paired/grouped other larger agencies/programs to form sampling units that satisfy a minimum size requirement. With these groupings, if one of these sampling units was selected, all agencies/programs that were grouped to form the unit would have been included in the Client Survey data collection effort. Since it is possible for an agency to be sufficiently large (i.e., not deemed to be undersized) but contain at least one program that is undersized, it was necessary to identify and address undersized programs during the construction of the agency sampling frame, i.e., prior to agency sampling.

In theory, this approach of pairing agencies/programs is effective in eliminating the problem of undersized programs. However, for HIA 2014, a very large proportion of programs were identified as undersized (partly as a result of missing client count data), and the resulting extent of pairing would have resulted in operational complications, since each set of paired programs is effectively one unit for sampling purposes but requires two separate program visits. Thus, in selecting the agency/program samples, the final decision was to not pair agencies/programs, but rather to allow undersized programs to remain as such on the sampling frame. In order to reduce the likelihood of such programs being sampled, the smallest 2 percent of agencies in each food bank were excluded from the sampling frame.

2.2.4 Stratification and Allocation

This section discusses considerations for stratifying and allocating the sample of agencies. Once the agencies had been sampled, one program was sampled from each sampled agency. Thus, at the program level, there was no need or opportunity for further stratification/allocation considerations, although allocation of the agency sample to the two frames is considered here, in light of objectives for estimation by program type.

During the design of the HIA 2014 Client Survey sample, considerable attention was given to allocation options. The data collection constraints for the HIA 2014 Client Survey were local in nature; since the food banks provided the volunteer or food bank staff data collectors, the maximum numbers of programs that could be included in the Client Survey data collection effort were

dependent on the food banks' resources. Thus, in contrast to most national studies, HIA 2014 did not have an essentially fixed pool of data collection resources to be allocated in the most efficient manner; instead, the resources were fixed at the food bank level (and this varied from food bank to food bank), and this limited the ability to attain allocations that are optimal for either national or food bank-level estimates (or any reasonable compromise allocation). These limitations were the key driving factors in determining the final allocation of the HIA 2014 agency sample; discussion of important allocation considerations was provided in memoranda and sample design reports.

In spite of the allocation limitations, the goals were still to attain an overall agency/program sample size that was considerably larger than that in HIA 2010 and to ensure sufficient sample sizes within food banks to support food bank-level estimates. However, Feeding America expressed serious reservations regarding the food banks' abilities to staff data collection at the numbers of agencies indicated in early versions of the allocation. While there was a genuine interest in attaining a large agency/program sample size, the need to put forth an allocation that would be feasible was recognized as imperative. As a result, in the end, Feeding America specified the number of agencies to be sampled in each food bank, based on resource considerations.

2.2.5. Selection of Noncertainties

As discussed above, agencies were selected using PPS sampling. Within each sampled agency, one program was selected with PPS. A key consideration in the sampling of noncertainty agencies was the sub-stratification to be used within food banks.

Two characteristics of agencies were considered for sub-stratification: geography and size. A simulation study was undertaken to evaluate three possible options for sub-stratification within food bank⁵:

- Stratification based on a geographic sort (using the county to which the agency's address geocodes);
- Stratification based on the agency's MOS; and
- Stratification based on the agency's MOS (three roughly equal-sized groups), with a geographic sort of agencies within each size grouping.

⁵ Note that in each case, implicit stratification (through sorting) was used, rather than explicit stratification.

These two characteristics—geography and size—were used in this exercise because Feeding America indicated that these are the two primary characteristics on which it is important to try to ensure appropriate representation of the sample within food banks. Further details of this simulation study were provided in memoranda and sample design reports.

After considering the results of this simulation, Feeding America decided that the third approach—stratification based on the agency’s MOS (three roughly equal-sized groups), with a geographic sort of agencies within each size grouping—should be used. This approach allowed for the sample size to be controlled by size of the agency (to ensure that both large and small agencies were sampled) and also helped balance the geographic distribution of the sample within the food banks.

2.2.6 Post Hoc Revisions to the Sample of Programs

Following the selection of the sample of programs, a number of revisions—both random and non-random—were made to affect the set of programs visited in HIA 2014. These post hoc revisions are described here. Table 2-6 gives counts, for each food bank, of the numbers of programs originally sampled, the numbers of program visits added after the original selection (for each reason below), and the number of programs subsampled out.

Selection of Reserve Sample

The original program sample size was 16,869, and the expected 13,664 completed program visits to these sampled programs were expected to yield totals of 98,244 sampled clients and 70,000 completed client surveys. However, these sample sizes were based on assumptions about eligibility and response rates that were somewhat intractable during the design stage. HIA 2014 introduced new modes of data collection and new survey methodology. Because of these changes, past experience was viewed as not directly relevant for determining these assumptions. For comparison, the sample size for HIA 2010 was 12,502 agencies and 82,301 clients. In that study, the overall eligibility rate of agencies in the Master List of Agencies was 64 percent; and the agency response rate was near 70 percent. The final data showed that clients were sampled in 6,551 agencies/programs and 62,143 clients completed the survey, with a client response rate near 80 percent. The HIA 2014 study is different from the HIA 2010 in many ways. HIA 2010 did not use the Agency Survey as the basis for constructing the sampling frame for the Client Survey, did not

include certain types of programs (e.g., mobile programs), and did not use an automated mode of data collection for the Client Survey.

In addition to the 16,869 sampled programs, a reserve sample of programs was prepared that included all eligible non-sampled programs. Reserve sample could be selected and released if actual program visit rates or eligibility rates were considerably lower than expected, or if the food bank determined it had sufficient resources to visit additional programs (without impacting the ability to complete program visits with originally sampled programs). An important limitation was that in several food banks, the allocation involved sampling all available agencies (i.e., all eligible agencies that completed the Agency Survey), so the only reserve sample available were additional programs in sampled agencies.

In cases where food banks found that they were able to handle more program visits than the number of sampled programs, Feeding America (with input from the food banks) requested random supplementation of the program sample. A variation on this is that in some cases, Feeding America requested supplementation only within a specific county, to support the ability to produce county-level estimates for that county. In response to such requests, programs were randomly sampled from among all available programs in the food bank (or all available programs in the food bank that were associated with the particular county).

Release of Non-Sampled Large Programs

Subsequent to the sampling of programs, Feeding America (with input from the food banks) determined that some very large programs had not been included in the sample of programs. In some cases, this was because either the agency was not included in the agency list provided by the food bank, or the program was not listed in the Agency Survey for the agency. In other cases, these programs were not sampled because the agency MOS and/or the program's MOS were not sufficiently large, due to issues with the data used to construct these measures of size. At Feeding America's request, such programs were released into the sample.

Creation of Additional Program Visits for Multiple Sites in Large Programs

For some large programs, Feeding America (with input from the food banks) requested additional program visits. These additional visits were designed to allow additional sites (in programs with

multiple sites) to be visited. In such cases, Feeding America provided lists that uniquely identified the program's sites, and additional site(s) were sampled from that list.

Subsampling To Reduce the Number of Program Visits

In some cases, the number of program visits required for the sampled programs was determined to exceed the resources available for data collection in the food bank. In such cases, Feeding America (with input from the food banks) requested subsampling to reduce the number of program visits. We randomly subsampled out program visits from among all remaining (not completed) program visits in the food bank.

Table 2-6. Food bank-level counts of numbers of programs originally sampled, added to sample, and subsampled out

Food bank	Total number of program visits ever released	Number of originally sampled programs	Number of randomly added programs	Number of additional program visits assigned for multiple sites	Number of very large, nonsampled programs added to sample	Number of programs subsampled out
1	128	110	18	0	0	0
2	63	62	1	0	0	0
3	114	110	4	0	0	0
4	119	110	9	0	0	0
5	110	110	0	0	0	0
6	110	110	0	0	0	0
7	110	110	0	0	0	0
9	110	110	0	0	0	0
10	110	110	0	0	0	0
11	110	110	0	0	0	0
12	108	108	0	0	0	0
14	110	110	0	0	0	0
15	100	100	0	0	0	0
16	110	110	0	0	0	0
17	100	80	0	0	0	20
18	113	110	0	0	3	0
19	110	110	0	0	0	0
20	114	107	0	0	4	3
21	110	110	0	0	0	0
22	99	81	0	3	6	9
23	80	80	0	0	0	0
24	110	110	0	0	0	0
25	110	110	0	0	0	0
26	110	110	0	0	0	0
27	110	110	0	0	0	0
28	108	108	0	0	0	0
29	108	108	0	0	0	0

Table 2-6. Food bank-level counts of numbers of programs originally sampled, added to sample, and subsampled out (continued)

Food bank	Total number of program visits ever released	Number of originally sampled programs	Number of randomly added programs	Number of additional program visits assigned for multiple sites	Number of very large, nonsampled programs added to sample	Number of programs subsampled out
30	110	110	0	0	0	0
31	110	110	0	0	0	0
34	110	110	0	0	0	0
35	81	80	0	1	0	0
36	110	110	0	0	0	0
37	111	110	1	0	0	0
38	110	110	0	0	0	0
39	110	90	0	0	0	20
40	110	110	0	0	0	0
41	110	110	0	0	0	0
42	110	110	0	0	0	0
43	120	110	10	0	0	0
45	110	110	0	0	0	0
46	108	80	0	0	0	28
47	80	80	0	0	0	0
48	114	110	0	0	4	0
49	110	110	0	0	0	0
50	110	110	0	0	0	0
51	121	110	11	0	0	0
52	110	110	0	0	0	0
53	110	110	0	0	0	0
55	110	110	0	0	0	0
56	110	110	0	0	0	0
57	112	110	2	0	0	0
58	110	110	0	0	0	0
59	80	80	0	0	0	0
60	110	110	0	0	0	0
62	110	110	0	0	0	0
63	80	80	0	0	0	0
64	110	110	0	0	0	0
65	110	110	0	0	0	0
70	110	110	0	0	0	0
72	108	108	0	0	0	0
74	110	110	0	0	0	0
75	121	110	11	0	0	0
76	110	110	0	0	0	0
77	110	110	0	0	0	0
79	114	100	0	6	8	0
80	100	80	0	0	0	20
81	112	110	2	0	0	0
82	111	108	0	1	2	0
84	109	91	0	0	1	17
85	110	110	0	0	0	0
86	110	110	0	0	0	0
87	110	110	0	0	0	0
90	110	110	0	0	0	0
91	127	108	2	13	4	0

Table 2-6. Food bank-level counts of numbers of programs originally sampled, added to sample, and subsampled out (continued)

Food bank	Total number of program visits ever released	Number of originally sampled programs	Number of randomly added programs	Number of additional program visits assigned for multiple sites	Number of very large, nonsampled programs added to sample	Number of programs subsampled out
92	108	108	0	0	0	0
93	100	100	0	0	0	0
94	115	108	7	0	0	0
95	111	110	0	0	1	0
96	63	48	0	6	9	0
97	117	110	6	0	1	0
98	70	70	0	0	0	0
100	117	110	7	0	0	0
102	80	80	0	0	0	0
103	111	110	1	0	0	0
105	114	104	2	2	2	4
106	90	90	0	0	0	0
110	37	37	0	0	0	0
112	118	110	8	0	0	0
114	65	65	0	0	0	0
116	73	70	2	0	1	0
117	55	55	0	0	0	0
119	80	80	0	0	0	0
121	55	55	0	0	0	0
123	51	51	0	0	0	0
124	106	106	0	0	0	0
126	108	91	0	0	0	17
129	57	55	2	0	0	0
133	23	20	3	0	0	0
134	110	110	0	0	0	0
136	55	55	0	0	0	0
137	28	28	0	0	0	0
138	110	110	0	0	0	0
140	64	64	0	0	0	0
141	49	44	0	5	0	0
145	100	100	0	0	0	0
146	93	90	0	0	0	3
149	85	85	0	0	0	0
151	59	59	0	0	0	0
152	110	110	0	0	0	0
155	110	110	0	0	0	0
156	112	76	0	0	4	32
157	50	42	8	0	0	0
159	55	55	0	0	0	0
160	80	80	0	0	0	0
161	110	110	0	0	0	0
164	90	90	0	0	0	0
167	59	55	0	4	0	0
168	115	110	5	0	0	0
169	110	110	0	0	0	0
170	42	40	2	0	0	0
174	77	77	0	0	0	0

Table 2-6. Food bank-level counts of numbers of programs originally sampled, added to sample, and subsampled out (continued)

Food bank	Total number of program visits ever released	Number of originally sampled programs	Number of randomly added programs	Number of additional program visits assigned for multiple sites	Number of very large, nonsampled programs added to sample	Number of programs subsampled out
175	108	108	0	0	0	0
177	55	55	0	0	0	0
178	31	31	0	0	0	0
184	78	78	0	0	0	0
185	83	69	0	1	6	7
186	65	65	0	0	0	0
187	70	70	0	0	0	0
188	110	110	0	0	0	0
189	112	110	0	2	0	0
191	55	55	0	0	0	0
192	128	108	20	0	0	0
193	74	61	13	0	0	0
195	54	54	0	0	0	0
196	43	43	0	0	0	0
197	108	108	0	0	0	0
201	80	80	0	0	0	0
202	60	55	5	0	0	0
206	72	62	0	0	10	0
207	110	110	0	0	0	0
208	108	108	0	0	0	0
209	60	60	0	0	0	0
214	108	108	0	0	0	0
216	108	85	0	0	0	23
218	51	51	0	0	0	0
219	100	100	0	0	0	0
220	108	108	0	0	0	0
221	108	108	0	0	0	0
222	67	67	0	0	0	0
259	110	110	0	0	0	0
276	28	28	0	0	0	0
278	108	108	0	0	0	0
279	50	50	0	0	0	0
280	108	108	0	0	0	0
281	110	110	0	0	0	0
284	67	55	10	0	2	0
285	55	55	0	0	0	0
287	77	71	2	0	4	0
290	53	53	0	0	0	0
291	110	100	0	0	0	10
294	47	47	0	0	0	0
296	108	108	0	0	0	0
298	79	79	0	0	0	0
305	60	55	5	0	0	0
306	90	90	0	0	0	0
310	108	108	0	0	0	0
313	25	25	0	0	0	0
314	10	6	0	4	0	0

Table 2-6. Food bank-level counts of numbers of programs originally sampled, added to sample, and subsampled out (continued)

Food bank	Total number of program visits ever released	Number of originally sampled programs	Number of randomly added programs	Number of additional program visits assigned for multiple sites	Number of very large, nonsampled programs added to sample	Number of programs subsampled out
316	60	60	0	0	0	0
319	110	110	0	0	0	0
323	108	108	0	0	0	0
324	110	110	0	0	0	0
326	64	64	0	0	0	0
339	48	41	6	0	1	0
369	92	92	0	0	0	0
540	36	36	0	0	0	0
559	40	40	0	0	0	0
563	108	108	0	0	0	0
603	25	20	0	0	5	0
611	84	82	2	0	0	0
626	30	30	0	0	0	0
627	52	52	0	0	0	0
684	70	70	0	0	0	0
685	50	50	0	0	0	0
779	92	90	2	0	0	0
780	57	57	0	0	0	0
781	78	68	10	0	0	0
Total	16,921	16,383	199	48	78	213

2.3 Sampling Clients

Following the selection of agencies and programs, clients served by the sampled programs were sampled. The sampling of clients within sampled programs was conducted in two stages. The first stage involved selecting a survey day (a day on which the program is open and serving clients) for conducting the sampling and data collection activities. The second stage was to select clients who visited the program site(s) during the assigned survey day/hours. The Client Survey was administered during the survey day/hours in the same service visit during which the client was sampled.

For a given sampled program, the average daily number of duplicated clients served was estimated, and the sampling interval for client selection was computed by dividing the estimated average daily number of duplicated clients by 7.2, the target number of clients to be sampled in each program

(based on the assumptions given in Section 2.1.5).⁶ The estimated average daily number of duplicated clients was initially estimated (and an initial client sampling interval computed based on this estimate) based on the client count data (reported or imputed) in the Agency Survey; however, prior to the program visit, the food banks' Hunger Study Coordinators had the opportunity to revise this estimate (and obtain a revised client sampling interval) based on information obtained during the pre-visit call with the program contact person.

2.3.1 The Role of Data Collectors in the Client Survey

The HIA 2014 Client Survey was conducted using trained food bank staff or volunteer data collectors who were at the program site during the survey day/hours to:

- Tally and record the total number of client visits to the program during the hours the program is open within the survey day/hours;
- Follow a sampling protocol to sample clients;
- Enter data about the sampled clients (e.g., time of visit, consent and response status, and observations of characteristics that will be considered for nonresponse adjustments);
- Invite sampled clients to participate and administer survey consent; and
- Assist the sampled clients in completing a self-administered ACASI survey.

For meal programs, the clients eligible for the survey were adults receiving a meal at the program site. For grocery programs, the clients were households. The data collection procedures called for a family or household unit sharing the same grocery order to be counted once and for one adult member of a sampled household to be asked to complete the Client Survey.

2.3.2 Assigning the Survey Day

Data collection for the Client Survey was conducted between April 17 and August 30, 2013⁷. To manage data collection over this period and ensure temporal balance, each sampled program was

⁶ The maximum sampling interval was capped at 100; computed intervals above that cap were reset to 100.

⁷ For a select group of agencies served by the Food Bank 003, a truncated data collection period of April 15 through May 19 was used due to logistical and operational concerns.

randomly assigned a survey day for conducting the survey. The specific survey day for each program depended on the day(s)/week(s)/month(s) that the program was open to serve food to clients. The assignment process included the two steps described in *Step 1* and *Step 2*. Below.

Step 1: Assigning Each Sampled Program to a Release Group by Block of Weeks

The entire 4 1/2-month survey period was divided into six blocks (four-week periods, with the exception of the first and last blocks) in each food bank: April 17 – April 28; April 29 – May 26; May 27 – June 23; June 24 – July 21; July 22 – August 18; and August 18 – August 25. The remaining time (August 26-30) was reserved for follow-up (due to missed survey days near the end of the data collection period) if required. Sampled programs that were open to serve clients on alternative months, quarterly, semi-annually, or annually, were assigned (with probabilities proportionate to the number of weeks in the block, except that the first block was assigned with a reduced probability designed to affect a weekly workload that was about 25 percent of the workload in the other blocks) to one block among the blocks during which they were open during the survey period. For the programs that were open to serve clients at least once in every block, the assignment was to one of the four possible months in the survey period. The random assignments were done such that the blocks were balanced in size, to the extent possible given the programs' service schedules, with the exception of the first block (as noted above). The overall objectives were to achieve (a) balance of programs among the blocks for estimation purposes, and (b) even workloads that could be accomplished by trained data collectors in each food bank.

Step 2: Assigning a Survey Day

Within its assigned block, each program was assigned a survey day in two steps: (1) by week, and (2) by days within the selected week. Across programs in the same block, the objectives were to constrain the selection such that, within food banks, there would be (a) an approximately equal number of programs assigned each week, and (b) no more than 2 programs assigned each day.

2.3.3 Sampling Clients during the Survey Day/Hours

This section describes the method that was used to determine the selection probability for sampling clients during the sampled program's survey day/hours. The probability of selection of a particular client visit v can be expressed as

$$\pi_v^f = P_f(v) = P_f(i)P_f(j|i)P_f(k|i,j)P_f(v|i,j,k)$$

where $P()$ denotes a probability of selection, i indicates the agency, j indicates the program, and k indicates the survey day/hours, and f denotes the frame (meal or grocery) from which the agency (leading to the sampled client visit) was sampled.

When a target overall selection probability is fixed, we can set the conditional client selection probability to

$$P_f(v|i,j,k) = \frac{\pi_v^f}{P_f(i)P_f(j|i)P_f(k|i,j)}$$

in order to attain the target overall probability.

By using PPS sampling at the agency and program stages, and random selections of opening days in a month as the survey day, we were able to set the conditional client selection probability as a function of the selection probability at each preceding sampling stage and achieve a client sample that was designed to be (a) self-weighting within each food bank and (b) a fixed client sample size within each program (in expected value). Note that these conditions depend on the allocation of the agency sample (to the two program types and to the food banks) being proportional. To the extent that the allocation deviated from proportional, departures from either or both of these two outcomes were possible. That is, to maintain the self-weighting aspect, the target client sample sizes within programs would need to vary. To maintain the fixed expected client sample sizes, the client selection probabilities would need to vary. Because of the operational constraints and limitations, a design that allows client sample sizes within programs to vary was not feasible. (As discussed below, variations were expected due to limitations in the data used to determine selection probabilities and client sampling intervals and due to fluctuations in client flows, so there was no room for additional variation by design.) Thus, the client selection probabilities varied to maintain the fixed expected client sample sizes.

The conditional client selection probability shown above is a fixed value for a given agency/program/survey day. This probability was achieved by sampling clients systematically. As clients arrived for service, the data collectors determined an ordering (e.g., by using their place in line, assigning a number, or using some other mechanism to create an ordering). The sampling instructions provided to the data collectors included a random start and a sampling interval (e.g., select the 4th client and every 12th client thereafter; so the 4th, 16th, 28th, 40th, etc. clients are sampled). Although the client sampling parameters were established in a way that was designed to yield a specific target number of sampled clients at a program, the actual number of sampled clients varied from that target due to issues with the data used to compute these parameters and/or variations in client flows.

2.3.4 Data Required for Determining the Client Sampling Rate

The data required to determine the selection probability of clients in each program in a food bank were:

- The overall target number of clients sampled at a program (7.2, based on the assumptions given in Section 2.1.5); and
- An estimate of the number of clients expected to report to the program for service during the survey day/hours. This was initially derived based on duplicated client counts reported (or imputed) in the Agency Survey, as well as information provided in the Agency Survey on the program's days and hours of operation.

Based on this information, an initial client sampling interval was computed for each program visit. In some cases, the client flow rate was expected to be such that it would have been difficult for the data collectors (or the equipment available for administration of the ACASI interview) to keep up with data collection for a sample selected using the specified interval. Alternatively, the intervals might have been so large as to be expected to preclude selection of any client (or to result in selection of only, say, 1 or 2 clients). As a result, prior to the visit, the Hunger Study Coordinator was able (through the Program Visit Report) to update the estimated number of clients used in this calculation, and when this was done, a new sampling interval was provided. Data collectors also had the opportunity to call the Westat Help Desk at the beginning of the program visit (prior to the start of data collection at the program) to obtain a revised sampling interval.

2.3.5 Special Situations Involving Client Sampling

This section contains discussion of a few special situations that, due to operational considerations or constraints, had an effect on client sampling.

Programs with Multiple Sites in Operation during the Survey Window

The design assumed that for each sampled program, the program site for conducting the survey was linked to the survey day. To avoid introducing a weighting factor that would result in additional variation in the survey weights, if program j was open in s site(s) during the survey day k , then the Client Survey should be conducted in the s site(s). However, the decision was that for programs that were open in multiple sites during the survey day, only one site would be visited; in such cases, subsampling of program sites was necessary. (Mobile programs, discussed below, are a special case of this.) Ideally, any such subsampling of program sites should have been done prior to determining the client sampling interval (so that the necessary adjustments could have been made to the client sampling interval). Although an advance review of sampled programs would have been ideal, Feeding America determined that such a review was not feasible, so it was necessary to have procedures for subsampling sites that could be implemented in the field.

Mobile Programs

For most programs, the selection of the program was the final stage of sampling prior to sampling clients. As discussed earlier, ideally, client sampling and data collection would have occurred in all program locations that were in operation for the given program during the specified survey day. However, for mobile programs, one or two additional stages of sampling were required for operational purposes. If the mobile program had more than one vehicle in operation at the same time during the specified survey day/hours, then exactly one vehicle was selected for data collection. This selection was done in the field, by applying a predetermined rule. In this case, information about the number of eligible vehicles was recorded and was used during weighting to adjust the program's probability of selection.

If the sampled mobile program traveled to more than one location on the survey day, then it was necessary, for operational reasons, to limit client sampling and data collection to just one location. (Again, ideally, client sampling and data collection would have occurred at all locations served by the

mobile program during the survey day. However, this was not operationally feasible.) In this case, the data collectors were instructed to go to only the first location served by the mobile program on the survey day. It was expected that this restriction of the first location (rather than a location randomly selected from among all locations served by the program during the survey day) may introduce a small amount of bias into the estimates, but it was deemed a necessary restriction from an operational standpoint. The number of hours the mobile program was at the first location and the total number of hours the mobile program was in service during the survey day were recorded and this information was used in weighting and estimation.

Limits to the Lengths of Data Collectors' Shifts

Typically, client counting, sampling, and data collection occurred during all hours the program was serving clients during the survey day. (Although multi-day data collection was possible from a sampling standpoint, it was ruled out for operational reasons.) However, programs whose service hours spanned a long period during the survey day would have been operationally difficult to staff with food bank staff or volunteer data collectors. As a result, for HIA 2014, Feeding America decided to limit the survey hours to no more than 6 hours in a day.

Programs that are open 24 hours a day are often shelters (or other residential programs) that have a kitchen from which people can get food. These may be open 24 hours a day, but not for “service” in quite the same way as other programs. For these programs there is often an intake process, and virtually all daily clients could be expected to be served the dinnertime meal. Thus, for these 24-hour programs, the data collection hours were set (by the Hunger Study Coordinators) to be a time span that included the full dinnertime service period. The key assumption in this case was that virtually all clients served during that day would have been present during that time span, so that any bias that might result from restricting the data collection hours would have been expected to be negligible.

For other programs with service periods spanning more than 6 hours a day (including programs open 24 hours a day other than shelters or other residential programs), one 6-hour period was randomly selected from all possible 6-hour periods during the survey day. The sampling of hours within the survey day was used in adjusting the program visit's probability of selection in the computation of the survey weights.

In some cases, information about the program's hours of operation were missing or inconsistently reported in the Agency Survey (the items in Q22). In those cases, a process was used to randomly

designate the survey data collection hours as either “from the opening of the program until 6 hours later” or “from 6 hours prior to the closing of the program until closing”. Again, in this case, the program’s total number of hours of operation during the survey day were used in adjusting the program visit’s probability of selection in the computation of the survey weights.

Multiple Visits to Programs by the Same Client

With the sampling approach described above, it is possible that clients who visit programs multiple times on the same day may have been sampled more than once. If a particular program is of the type that many of the same clients are served at each meal on a given day, one option would have been for this information to be used to restrict the survey data collection period to just the hours spanning one meal and adjust the client sampling parameters (random start and sampling interval) accordingly (e.g., if three meals are offered, randomly select one of the three, and multiply the original client sampling interval by a factor of one-third). This option was ruled out for HIA 2014 due to its requirement for the food banks or Hunger Study Coordinators to review the selections, identify such situations, and transmit information back regarding hours associated with each meal. However, the restriction of the survey data collection period to a 6-hour window is expected to have helped limit overlap in clients served. If a given client was sampled multiple times (either at the same program or at different programs), the client was to have been asked to complete the Client Survey each time they were sampled, due to the inability to directly link these cases.

2.3.6 Handling Missed Survey Days

Westat emphasized the importance of ensuring that Client Survey data collection occurred at the sampled program on the specified survey day during the specified hours. However, even with diligent efforts to ensure that data collection occurs as prescribed, there were situations in which it was not possible to administer Client Survey data collection as originally assigned. In such cases, a replacement day was issued (using a pre-specified procedure) and a disposition code captured the fact that the designated day was missed. Up to two replacement days were issued before the program was coded out as final program-level nonresponse.

Staffing/Resource Issues

If a survey day was missed due to staffing or resource issues (e.g., inability to transfer computers from one data collection team to another), the replacement day assignment procedure described above was used.

Weather-related Impediments

If a survey day was missed because the data collectors could not get to the program site due to weather-related issues, a replacement day was issued as described above. If weather-related issues resulted in very heavy demands on the program (e.g., disaster relief efforts), technically, data collection should have occurred as scheduled. (The temporary heavy demands on the program are outcomes that, technically, should have been measured and incorporated into the estimates.) However, in some cases, it may not have been logistically possible to conduct data collection in such circumstances.

Other Events Affecting Client Demand Levels

Feeding America indicated that there are situations in which client demand levels may be affected by other temporary situations (e.g., a whale is caught in an Alaska fishing village, so there is a temporary hiatus from a need for the program's meal and/or grocery services). In such situations where the demand for the program's services was affected by external events, Client Survey data collection should have proceeded as prescribed. Even if no clients reported for services on the survey day, that outcome that should have been measured and reflected in the estimates (through the reporting of no sampled clients).

3.1 Introduction

Feeding America has the following three analytic objectives for its ongoing research into hunger in America: (1) to create a national profile of charitable food recipients in the United States, (2) to construct precise weekly, monthly and annualized national and local estimates of the numbers of charitable food recipients, and (3) to summarize the breadth and depth of services available in the Feeding America network and the role of those services in the food system, broadening the understanding from that achieved in the Hunger in America 2010 (HIA 2010) Study.

The first objective requires the development of survey weights that are appropriate for the analysis of data from the Client Survey in the HIA 2014 Study. The second objective requires the collection of data and the development of procedures to produce unduplicated estimates of client counts; “unduplicated” involves de-duplicating visits such that each unique client served during the target time period is counted only once in the estimate. The third objective requires the development of survey weights that are appropriate for the analysis of data from the HIA 2014 Agency Survey.

3.2 General Approach for Computing Unduplicated Client Count Estimates

3.2.1 Dual-Frame Sampling and Estimation

As discussed in Chapter 2, sample selection for the HIA2014 Client Survey was based on a dual-frame approach, with separate frames (and independent sampling from those frames) for clients served by meal programs and clients served by grocery programs. Each agency was assigned to one or both of these frames depending on the type(s) of programs offered by the agency. If an agency was sampled from the meal frame, one meal program offered by that agency was sampled, and the Client Survey was administered to clients from the meal program, and likewise for an agency sampled from the grocery frame. (Note that there were a few exceptions due to the addition of programs to the sample after the original selection.)

Clients were sampled based on their reporting for service at the particular program during a pre-specified (randomly determined) survey day. A given client might report to an eligible program for service multiple times within the analytic window of time (i.e., week, month, or year). Accounting for these multiplicities in estimation is a process referred to in this survey program as “unduplication.”

Ignoring duplication for the moment, the term “client visit” will be used to refer to the reporting for service of a particular client during a particular survey day/hours. (It is these client visits that are unduplicated to compute unduplicated client count estimates.) The probability of selection of a particular client visit v is

$$\pi_v^f = P_f(v) = P_f(i)P_f(j|i)P_f(k|i,j)P_f(v|i,j,k) \quad (3.1)$$

where $P()$ denotes a probability of selection, i indicates the agency, j indicates the program, and k indicates the survey day/hours, and f denotes the frame (meal or grocery) from which the agency (leading to the sampled client visit) was sampled.

An advantage of the dual-frame estimation approach is that it is not necessary to determine each unit’s probability of selection from each frame; it is sufficient to determine the unit’s probability of selection in the frame from which the unit was sampled, along with the frame(s) from which the unit could have been sampled. That is, for a client sampled from the meal frame, we simply need to know that client’s probability of selection from the meal frame and whether or not the client could have been sampled from the grocery frame (i.e., whether or not the client obtained services from a grocery program). If we were to use a single-frame estimator, we would need to know the client’s probability of selection from the meal frame as well as that client’s probability of selection from the grocery frame.

It should be noted that in accounting for multiplicities, it is necessary to consider the concept of the “client.” For meal programs, since the client is the individual adult, it is only necessary to account for services received by that adult. However, for grocery programs, since the client is the household, it is necessary to account for services received by the respondent and anyone else in the respondent’s household. Additionally, when counting and sampling clients in grocery programs, households were treated as a single unit; i.e., if multiple members of a household reported to the grocery program at the same time, the data collection procedures called for them (as a group) to only be counted once.

3.2.2 General Form of the Estimator

For estimating client counts (duplicated or unduplicated) by broad program type, an expansion estimator (i.e., weighted sum) based on clients from only the given program type is used. For example, to estimate client counts for meal program clients (alone), a weighted sum of clients sampled in meal programs is used. (The computation of the weights is discussed in Section 3.3.)

For obtaining overall (across both program types) estimates of client counts (\hat{Y}_T), a dual-frame estimator is being used; i.e., the estimated number of clients served by meal programs is

$$\hat{Y}_T = \hat{Y}_M^m + \hat{Y}_G^g + \alpha \hat{Y}_{MG}^m + (1 - \alpha) \hat{Y}_{MG}^g, \quad (3.2)$$

where \hat{Y}_M^m is an estimate of clients who were served only by meal programs (M), based on clients sampled from the meal frame (m); \hat{Y}_{MG}^m is an estimate of clients who were served by both meal and grocery programs (MG), based on clients sampled from the meal frame (m); and \hat{Y}_{MG}^g is an estimate of clients who were served by both meal and grocery programs (MG), based on clients sampled from the grocery frame (g); \hat{Y}_G^g is an estimate of clients who were served only by grocery programs (G), based on clients sampled from the grocery frame (g); and $0 < \alpha < 1$ is the compositing factor.

There are several possible approaches for determining the compositing factor α . (See, for example, Lohr and Rao 2006⁸). One approach is to simply use the constant value $\alpha = 0.5$, i.e., take the simple average of the two estimates for the overlap group (MG). For HIA 2014, in order to account for the disparity in sample sizes between the two frames, the values of α are set based on the relative sample sizes; e.g., for estimating the total for meal programs α is set equal to the proportion of clients sampled from the meal frame, and for estimating totals for grocery programs α is set equal to the proportion of clients sampled from the grocery frame. While other approaches are available for choosing α (e.g., approaches based on minimizing variance or mean squared error), using a relatively simple approach such as the recommended approach based on sample sizes may be preferred because (1) the overlap set (clients served by both meal and grocery programs) is expected to be small relative to the sizes of the sets of clients served by only meal programs or by only grocery programs, so the terms involving α are not likely to contribute much (relatively speaking) to the overall totals or the variances of those totals (thus, rendering the choice of α rather insignificant); (2) this approach is relatively simple to implement, understand, and explain; and (3) some of the other

⁸ Lohr, S., and Rao, J.N.K. (2006). Estimation in multiple-frame surveys. *Journal of the American Statistical Association*, 101, 1019-1030.

approaches for choosing α (e.g., approaches based on minimizing variance or mean squared error) result in the need for different sets of weights for each specific variable.

3.2.3 Estimation of Duplicated Client Counts (Client Visits)

Disregarding nonresponse for now, for estimating duplicated client counts (i.e., client visits or food distributions), the components of the estimators shown above use weights that are based only on base weights (i.e., sampling weights that are reciprocals of the probability of selection of the client visit). For example,

$$\hat{Y}_M^m = T^* \sum_{v \in m} I_{v,M} \frac{1}{\pi_v^m} \quad (3.3)$$

where $I_{v,M}$ is an indicator that the client associated with client visit v was served only by meal programs during the given timeframe (week, month, or year), with a value of 1 indicating that the client was served only by meal programs and a value of 0 indicating that the client was not served only by meal programs. \hat{Y}_{MG}^m , \hat{Y}_{MG}^g , and \hat{Y}_G^g are analogous in form. T^* is a factor that adjusts for the difference between the HIA 2014 data collection period and the target timeframe (week, month, or year). Note that a single value of T^* is used for virtually all programs (for a given target timeframe), with one exception; for a small set of agencies served by the Food Bank of Alaska, a truncated data collection period was used. Thus, in these agencies, a different value of T^* (that accounts for the truncated data collection period) is used.

In practice, these estimates are computed by replacing the base weights in expression (3.3) with the final client-level weights that include adjustments for nonresponse, as described in Section 3.3.

3.2.4 Estimation of Unduplicated Client Counts

Again disregarding nonresponse for now, for estimating unduplicated weekly and monthly client counts, the estimators are similar in form to those used for duplicated client counts, but include a factor that accounts for multiple visits by the same client during the given timeframe. (The estimator for unduplicated annual client counts takes a different form.) For example,

$$\hat{Y}_M^m = T^* \sum_{v \in m} d_{v,m} I_{v,M} \frac{1}{\pi_v^m} \quad (3.4)$$

where $d_{v,m}$ is, for the client associated with client visit v (a client served by only meal programs), the reciprocal of the number of times that client was served by meal programs during the given timeframe (week or month). \hat{Y}_{MG}^m , \hat{Y}_{MG}^g , and \hat{Y}_G^g are analogous in form.

In practice, these estimates are computed by replacing the base weights in expression (3.4) with the final client-level weights that include adjustments for nonresponse.

3.2.5 Estimation at Different Levels

HIA 2014 was designed to support estimation of duplicated and unduplicated client counts at both the food bank and national levels. For estimation at the national level, the estimators shown above involve summing over all client visits sampled from the given frame (nationally), and the counts of the numbers of times the client was served by the given broad type of program that are used to compute $d_{v,m}$ and $d_{v,g}$ pertain to any programs associated with food banks in the Feeding America network.

For estimation at the food bank level, the estimators involve summing over all client visits sampled in programs associated with the specific food bank; the identification of whether the client was served only by meal programs, only by grocery programs, or by both meal and grocery programs technically should be done at the food bank level (e.g., if the client was served only by a meal program supported by Food Bank 1 and a grocery program administered by Food Bank 2, that client should be treated as a meal-only client for food bank-level estimates for Food Bank 1). Similarly, for unduplicated food bank-level estimates, the counts of the numbers of times the client was served by the given broad type of program that are used to compute $d_{v,m}$ and $d_{v,g}$ should also pertain only to programs associated with the given food bank. For HIA 2014, it was determined that it would not have been feasible for clients to identify the specific programs from which they received services (and, therefore, it would not be possible to associate these programs with particular food banks). The implications are that misclassification of clients (clients classified as both meal and grocery clients, when they should have been classified as only meal or only grocery clients) may reduce the precision of food bank-level estimates, and deflated values of $d_{v,m}$ and $d_{v,g}$ (as a result of the counts in their denominators including service received by the client from other food banks) may result in underestimation of unduplicated client counts at the food bank level.

In some cases, clients are served by programs that are supported by more than one food bank. (The agencies providing such programs are referred to as “shared agencies”.) As discussed in Section 2.2.1, shared agencies were listed only by their primary food bank, to avoid duplication within the sampling frame. However, for estimation purposes, it was decided that clients of shared agencies should be counted in the client counts of both supporting food banks. For food bank level estimates, the approach described above would result in these clients being counted in the food bank-level estimates for the primary food bank, but not in those for the secondary food bank. In order to count these clients in the food bank-level estimates for both supporting food banks, the shared agency records (and records of clients associated with shared agencies) were duplicated during the computation of the set of Client Survey weights used for food bank-level analyses. As a result, the food bank-level estimates, by design, do not sum to the national totals.

State-level estimates or estimates at other geographic levels (e.g., the combined service area of two food banks) may also be of interest. In theory, the weights and estimators described here still pertain, and the issues discussed above that apply to food bank-level estimates also apply to state-level estimates. However, there are additional issues with state-level estimates.

One issue concerned the weights to be used for state-level estimates. Consideration was given to using the weights developed for food bank-level analyses and aggregating food banks within the state. However, this would result in double-counting of clients of shared agencies (if the agencies were shared by two food banks in the same state), and would not account for any nonparticipating food banks within the state. An alternative approach is to use the weights developed for national analyses, and just subset to the food banks within the state. However, the national analysis weights include an adjustment for food bank nonparticipation (nationally), which is particularly problematic if all food banks in the given state participated in the study. It is important to note that official State level reports were issued only for States in which all food banks participated in HIA 2014. State reports prepared outside this subset (in particular, those for Indiana and New York) pertain only to the subset of food banks that participated, without adjustment for nonparticipating food banks. The decision was, in general, to use as weights for state-level analyses the weights intended for food bank-level analyses. The exception, used in three states (Arizona, Michigan, and Pennsylvania) in which a high level of shared agencies between food banks within the state made this approach inadvisable, was to construct weights for state-level analyses by removing the food bank nonparticipation factor from the national weights.

A second issue involved determining which records should be included in a given state’s analysis. Due to situations in which a food bank’s service area spans more than one state, the decision of which records should be included in a particular state’s estimates was, in general, based on the geocoding of agency addresses (followed by select changes to the state/county assignment that were provided by Feeding America); agencies whose addresses geocoded to a location within the state were included in that state’s estimates, and all other agencies were excluded. For two State reports, Rhode Island and Utah, the States are primarily served by a single food bank, so the small numbers of agencies in those States partnered with Food Banks from outside the State were not included in State estimates.

3.2.6 Estimation over Different Time Periods

Estimates of duplicated and unduplicated client counts are needed at the weekly, monthly, and annual levels. As discussed in Section 2.3.2, the survey days were assigned in a manner that balances the temporal assignment (across weeks in a month and across months in the data collection period) among programs administered by a given Food Bank, by broad program type. Additionally, each of the estimators includes a factor that accounts for temporal variations, based on the monthly days of operation data provided by the food banks. (The original plan had been to use the prior year’s client counts from the Agency Survey—either the duplicated client counts in item Q31D or the unduplicated client counts in item Q31C—but data quality concerns led to the decision to use the days of operation counts instead.) This “seasonal” factor takes the form \bar{C}_j / C_{jt} , where \bar{C}_j is the average monthly days of operation for program j , and C_{jt} is the monthly days of operation count for program j for the same month as the month including the survey day (but one year earlier).

An additional factor, a “time reference factor”, is used to account for the amount of time spanned by the particular estimate (i.e., to account for whether the estimate is a weekly, monthly, or annual estimate). This factor takes the form a/b , where a is the number of days in the reference period pertaining to the estimate (7 for weekly estimates, 365/12 for monthly estimates) and b is the number of days in the HIA 2014 data collection period (33 for programs of “remote” agencies in Food Bank 003, 137 for all other programs). For estimates other than client count estimates, weights that use the monthly time reference factor are used.

When computing these estimates, besides the need to account for temporal variations and the length of the time reference period, there is a need to use indicators ($I_{v,M}$, $I_{v,G}$, and $I_{v,MG}$) that reflect the

type(s) of service received during the allotted time period (past week or past month) and—for the unduplicated client count estimates--unduplication factors ($d_{v,m}$ and $d_{v,g}$) that account for numbers of times service was received during the allotted time period.

For unduplicated annual client count estimates only, the estimator takes a different form than that used for all other client count estimates. The unduplicated annual estimate uses the unduplicated monthly estimate as its basis, and accounts for the estimated proportion of clients that are “newcomers” to the type of program in a month. That is, the unduplicated annual client count estimate for program type p is computed as:

$$\hat{Y}_{p,ann} = (1 + 11\tau_p)\hat{Y}_{p,mnth}, \quad (3.5)$$

where $\hat{Y}_{p,mnth}$ is the unduplicated monthly client count estimate for program type p , and τ_p is the estimated monthly newcomer rate for program type p . “Newcomers” are defined as grocery clients with $Q49 = 1$ and $Q50 = 1$ and $Q51 = 1$, as well as meal clients with $Q55 = 1$ and $Q56 = 1$ and $Q57 = 1$. (Both sets of conditions flag clients reporting that they have received services once in the past year—the visit during which they were sampled.)

Section 3.2.5 contained a discussion of considerations for estimation at different levels. A final consideration for producing estimates at different geographic levels is the geographic scope used for the calculation of the various factors. The compositing factor is computed at the geographic level of the analysis (e.g., using national ratios for national estimates, food bank ratios for food bank estimates). Consideration was also given to computing the newcomer rate at the geographic level of the analysis, but due to concerns about imprecise estimates (due to small sample sizes for some sub-national estimates), the decision was to use the national newcomer rates. All other factors in the computation of the client count estimates were invariant to the geographic level of the estimate.

3.2.7 Client Survey Items Used in Computing Unduplicated Client Count Estimates

The following are considerations pertaining to Client Survey items that are used for computing the estimates of unduplicated client counts:

- In order to determine the indicators $I_{v,M}$, $I_{v,G}$, and $I_{v,MG}$, it is necessary to know whether the client was served only by meal programs, only by grocery programs, or by

both meal and grocery programs. Since estimates are needed for different reference periods, this information must be captured for the “past week” and for the “past month”. This information is captured in items Q46 and Q47 (for grocery programs) and Q52 and Q53 (for meal programs) in the Client Survey.

- For computing the unduplication factors that account for the number of times the client was served by the particular broad type of programs during the given timeframe ($d_{v,m}$ and $d_{v,g}$), it is necessary to capture the following:
 - For clients sampled through a meal program, the number of times the client was served by meal programs in the past week and in the past month (items Q49 and Q50 in the Client Survey).
 - For clients sampled through a grocery program, the number of times the client was served by grocery programs in the past week and in the past month (items Q55 and Q56 in the Client Survey).

Strictly speaking, the information discussed above should be based only on services provided by Food Banks in the Feeding America network. However, this would have required capturing (in the Client Survey) the names of the particular programs providing services to the client (so that these could be traced back to the associated Food Bank); as discussed in Section 3.2.5, it was determined that obtaining reliable client reports of the names of programs providing services to them would not be feasible. As a result, there is the potential for bias in the client count estimates (at both the national and Food Bank levels) due to accounting for services provided by programs that are not part of the Feeding America network.

The client count estimates provided in the analysis reports are all at the individual level. While meal clients are individuals, grocery clients are households. Thus, to compute individual-level client counts, a household size factor needs to be applied to convert grocery clients (households) to individuals. The Client Survey item used as this factor is Q1, which asked the respondent for the number of household members.

3.3 General Methodology for Computing Survey Weights

This section contains a description of the general approach for computing survey weights for HIA 2014. In Section 3.4, specific details of the weighting process are provided.

3.3.1 Base Weights

Base weights account for differential probabilities of selection, and are computed as the reciprocal of the unit's selection probability. These weights reflect stages of sampling that occur prior to data collection, as well as those that occur during the data collection effort. The computation of base weights is the first step in the process of developing weights that are appropriate for the analysis of Agency Survey or Client Survey data.

For the Agency Survey, all agencies in food banks participating in HIA 2014 were included; i.e., the Agency Survey was a census of these agencies. Thus, each agency invited to participate in the Agency Survey had a probability of selection of 1.

For the Client Survey, the calculation of base weights should reflect all stages of sampling. The probability of selection of a given client visit was shown in expression (3.1). During the computation of Client Survey weights, the terms in this expression were adjusted as necessary to reflect additional sampling done in the field (e.g., sampling one vehicle from more than one in operation for a mobile program on the survey day; sampling a 6-hour period within the survey day; etc.).

3.3.2 Adjusting for Nonresponse

In HIA 2014, there are several stages at which nonresponse occurred:

- Nonparticipation of food banks: A small number of food banks declined to participate in the HIA 2014 study, and a small number of food banks that participated in the Agency Survey did not participate in the Client Survey. (See Section 1.2 for further discussion of this and for a list of these food banks.)
- Agency nonresponse to the Agency Survey: Some agencies listed in the Agency List (compiled from lists provided by the food banks) failed to complete an Agency Survey. Two different sets of criteria were used to classify agencies as respondents or nonrespondents. For the purpose of Client Survey sampling, an agency was considered a respondent if it listed and provided the program type for a least one program. For the purpose of Agency Survey analysis, more stringent criteria were used. .
- Agency/program nonparticipation in the Client Survey: The agencies that met the criteria for being considered a respondent for Client Survey sampling purposes were used to construct the sampling frame for agency/program sampling for the Client Survey. Among sampled programs, some were not visited for Client Survey data collection. For example, for various reasons (weather, illness, etc.) the volunteer data

collectors might not have been able to report to the sampled program on the designated survey day. In such cases, a replacement survey day was assigned (and up to 2 such replacement days were permitted for a given program). However, even with replacements, Client Survey data collection did not occur at some programs.

- Client nonresponse to the Client Survey: Some clients sampled for the Client Survey failed to complete the survey. For example, some clients refused to complete the survey, some did not have time, and others were unable to complete it because they do not speak any of the languages in which the survey is administered. Even among those who did submit a client survey, in some cases, the survey was deemed to not be complete enough to be considered a response.⁹

For estimates of totals such as the total unduplicated client counts that are key estimates for HIA 2014, nonresponse results in bias (underestimation) because clients served by nonresponding food banks, agencies, and programs are not included in the estimated totals. For other types of estimates (means, proportions, ratios, regression coefficients, etc.), bias in the survey estimates may result if the characteristics of nonrespondents differ from the characteristics of respondents.

In an effort to reduce bias in the HIA 2014 survey estimates due to nonresponse, the base weights were adjusted for nonresponse at the agency, program, and client levels. The weights used for national estimates were adjusted for food bank nonparticipation; the weights used for food bank-level and other local-level analyses were not adjusted for food bank nonparticipation because this adjustment involves inflating the weights of respondents in participating food banks to account for the nonparticipating food banks, and doing so would introduce a positive bias in food bank level client count estimates. Below, we describe the general approach used for making these adjustments.

Separate adjustments were made to the weights to adjust for each stage of nonresponse. For nonresponse at the agency, program, and client levels, these adjustments applied a sample-based weighting class adjustment (Kalton and Flores Cervantes, 2003¹⁰). Each unit was classified as either a respondent (*R*), a nonrespondent (*NR*), an ineligible case (*I*), or a case with unknown eligibility (*U*). The weights were adjusted within weighting classes, also referred to as *nonresponse adjustment cells*. The variables used to form the nonresponse adjustment cells must be available for both respondents and nonrespondents. For nonresponse at the food bank level, the adjustment involves the ratio of total poundage among all food banks to total poundage of participating food banks.

⁹ To be considered a response, at least 50 percent of core survey questions—those not involving skip logic—had to have been answered.

¹⁰ Kalton, G., and Flores Cervantes, I. (2003). Weighting methods. *Journal of Official Statistics*, 19(2), 81-97.

For the adjustment to be effective in reducing nonresponse bias in key estimates from the survey, the variables selected to form the cells should be associated with both the response propensities (the probabilities of response) and those key survey items. For HIA 2014, at each stage, there is generally little information available about the nonrespondents. The following are the variables that were used in forming nonresponse adjustment cells for each stage:

- **Agency nonresponse to the Agency Survey:** For agency nonrespondents, the only information available is that from variables that were included in the Agency List. For this adjustment, agency poundage was used in the computation of the adjustment factor. Different adjustment factors pertained to the Agency Survey and the Client Survey, since the sets of agencies considered “respondents” differed for Agency Survey analysis versus Client Survey sampling. The cells used for this adjustment were the food banks.
- **Non-visited programs in the Client Survey:** The program visit rates varied considerably among the food banks. Thus, the cells used for this adjustment were the combinations of food bank and program type (meal vs. grocery).
- **Client nonresponse to the Client Survey:** For each client sampled for the Client Survey, the volunteer data collectors were instructed to record the client’s gender and broad age classification (young/old) based on observation. These two variables were considered for use in forming nonresponse adjustment cells. However, due to missing information on these variables, the inability to link these data to the completed client surveys, and extreme adjustment factors when these were used, the decision was to not use these demographics and to use only the combinations of food bank and program type to form cells for client nonresponse adjustment.

At each stage of adjustment, once the cells were constructed, the input weights of the respondents within a nonresponse adjustment cell were inflated to compensate for the input weights of nonrespondents in that cell, as well as a proportion of the cases with unknown eligibility estimated to be eligible (where applicable). The proportion of the unknowns who would be eligible was assumed to be the same as the proportion among cases with known eligibility. That is, the nonresponse adjustment factor, f_{1b} , applied to each respondent in adjustment cell b was:

$$f_{1b} = \frac{\sum_{j \in R_b \cup NR_b} w_{Bj} + \sum_{j \in U_b} \left(\frac{\sum_{k \in R_b \cup NR_b} w_{Bk}}{\sum_{k \in R_b \cup NR_b \cup I_b} w_{Bk}} \right) w_{Bj}}{\sum_{j \in R_b} w_{Bj}} \quad (3.6)$$

where R_b , NR_b , U_b , and I_b are the sets of respondents, nonrespondents, cases with unknown eligibility, and ineligible, respectively, in adjustment cell b , and w_{Bi} is the input weight for unit i . For unit i in adjustment cell b , the nonresponse adjusted weight, w_{Ni} , is

$$w_{Ni} = \begin{cases} w_{Bi}f_{1b}, & i \in R_b \\ w_{Bi}, & i \in I_b \\ 0, & i \in NR_b \cup U_b \end{cases} \quad (3.7)$$

Note that in this adjustment, the weights of ineligibles are not affected. It is not appropriate to adjust the weights of these cases at this stage.

At each stage of adjustment, the distribution of the weights was examined to determine whether any weight trimming was warranted. Outlier weights were candidates for trimming, to reduce their potential influence on estimates. Additionally, the unequal weighting effect (estimated as $1 + cv_w^2$, where cv_w^2 is the square of the coefficient of variation of the weights; see Kish, 1965¹¹) was computed at each stage to assess the extent to which the particular adjustment was expected to affect precision.

The final agency weights that are used in the analysis of data from the Agency Survey include only the adjustment for agency nonresponse to the Agency Survey (and, for national estimates, the adjustment for food bank-level nonparticipation). The final client weights that are used in the analysis of data from the Client Survey (including the production of unduplicated client count estimates) include all of the adjustments discussed above.

3.4 Details of the Computation of Survey Weights for the Agency Survey and the Client Survey

The general approach to computing survey weights for analyses of the Agency Survey and the Client Survey was discussed in Section 3.3. In this section, we provide specific details on the computation of these weights.

¹¹ Kish, L. (1965). *Survey sampling*. New York: John Wiley and Sons.

3.4.1 Agency Survey Weights

Since the Agency Survey was intended to be a census, each agency had a base weight of 1. The computation of the Agency Survey weights included just two adjustments, the results of which follow.

Adjusting for food bank non-participation. Among the 212 food banks (including PDOs that participated separately, essentially acting as food banks for the purpose of this study), 196 participated in the Agency Survey. The overall adjustment for food bank nonparticipation in the Agency Survey, which was based on food bank poundage, was 1.075 (meaning, the weighted food bank participation rate for the Agency Survey was 93 percent). This adjustment was applied only to the weights for national-level analyses. There were a few special situations involving food banks' participation in the Agency Survey:

- Clark County, Washington, is technically served by Food Bank 068, which did not participate in the HIA 2014 study. Since this was the only county in Washington that was not being included in the study, the Agency Survey for the agencies that serve Clark County was administered through Food Bank 081 in Washington.
- At Feeding America's request, food banks 207 and 311 were essentially treated as a single food bank for sampling and estimation purposes.

Adjusting for agency nonresponse. Among the 44,659 agencies eligible for the Agency Survey, completed surveys (i.e., deemed complete enough for analysis purposes) were received from 32,677 agencies. The adjustment for agency nonresponse (within participating food banks) was computed separately for each food bank, and was based on agency poundage; the adjustment factor was the reciprocal of the poundage-weighted Agency Survey response rate for the food bank. The overall Agency Survey response rate (weighted by poundage) was 71.5 percent. The lowest weighted Agency Survey response rates were in food banks 603 (15%), 080 (16%), 090 (29%), 047 (32%), 064 (33%), 276 (37%), and 010 (37%).

3.4.2 Client Survey Weights

The following provides information on each of the computations and adjustments applied in calculating the Client Survey weights.

Program base weights. As discussed in Section 2.2.6, changes were made after the original selection to the set of programs assigned to be visited in HIA 2014. Because of the departures from probability sampling used to identify programs to be visited, it was not possible to assign program base weights based directly on program probabilities of selection (i.e., assign program base weights as the reciprocals of the program selection probabilities). Thus, an alternative approach was developed for this purpose. Under this alternative approach:

- The very large programs that were added to the sample at the request of Feeding America were classified as certainties; Feeding America also identified all sampled programs that, had they not been sampled, they would have requested to add to the sample (based on the same criteria used to request the addition of the former group). All of these programs were treated as certainties and assigned a program base weight of 1.
- For the remaining programs, if the food bank's sample was supplemented (other than the addition of the very large programs), the sample is viewed as having been selected in two phases—the original sample (phase 1) and the supplemental portion of the sample (phase 2).
 - If the program was selected in the phase 1 sample, it retained its original probability of selection, $P(1)$.
 - If the program was not selected in the phase 1 sample but was added through subsequent random supplementation, its phase 2 probability of selection, $P(2)$, was approximated as $P(2) = [1 - P(1)] \left[\frac{m_2}{M_2} \right]$, where m_2 is the total number of noncertainty programs added to the sample for the given food bank through random supplementation, and M_2 is the number of noncertainty programs in the given food bank that were available for supplemental selection. If there was county-specific supplementation, m_2 and M_2 were computed separately for (a) the supplemented county and (b) the remainder of the food bank.

The program's base weight was then computed as $1/[P(1) + P(2)]$.

Adjusting for food bank nonparticipation. Among the 212 food banks (including PDOs that participated separately, essentially acting as food banks for the purpose of this study), 188 participated in the Client Survey. The overall adjustment for food bank nonparticipation in the Agency Survey, which was based on food bank poundage, was 1.117 (meaning, the weighted food bank participation rate for the Client Survey was 89.5 percent). There were a few special situations involving food banks' participation in the Client Survey:

- For Food Bank 136, there was one particular sampled program with a very large base weight (relative to others in that food bank) for which a program visit was not completed. Due to the adverse effects of that program on the estimates for this food bank, Feeding America decided to treat that food bank as a nonparticipating food bank for the purpose of computing national Client Survey estimates.
- At Feeding America's request, food banks 207 and 311 were essentially treated as a single food bank for sampling and estimation purposes.

Adjusting for agency nonresponse. Among the 44,659 agencies eligible for the Agency Survey, 36,211 returned surveys that were deemed complete enough for sampling purposes. (Note that not all of these were deemed complete enough for analysis purposes.) The adjustment for agency nonresponse (within food banks participating in the Client Survey) was computed separately for each food bank, and was based on agency poundage. The overall Agency Survey response rate (weighted by poundage, where “response” is defined as complete enough for sampling purposes, as opposed to complete enough for analysis purposes) was 84.9 percent. The lowest weighted Agency Survey response rates were in food banks 603 (15.3%), 184 (45.9%), and 007 (51.4%).

Adjusting for exclusion of very small agencies. During program sampling for the Client Survey, the smallest two percent of agencies in each food bank were dropped from the sampling frame, in an attempt to reduce the likelihood of situations in which very small programs (likely to have very low client flows) were sampled. A weight adjustment was included to account for these very small agencies. This adjustment accounts for the food bank's proportion of total poundage associated with the very small agencies. The mean adjustment factor was 1.0001, and the maximum adjustment was 1.0032.

Adjusting for the random assignment of survey day and hours and the sampling of sites/vehicles within programs. An adjustment was applied to account for the sampling of a single day on which to visit the program, and for the sampling of hours within that day. For programs with multiple sites or multiple vehicles in operation on the survey day, a single site/vehicle was sampled and an adjustment was applied to the weights to account for this selection. The adjustment for the sampling of a single day was equal to the estimated number of days the program was open during the data collection period. The mean adjustment was 47.5, and the maximum was 137 (the total number of days in the data collection period). The adjustment for the sampling of hours was equal to 1 if the duration of the program visit was 6 hours or less, and equal to the number of hours the program was open on the survey day divided by the duration of the visit, for visits that were 6 hours or longer. The mean adjustment was 1.01, and the maximum was 4. The adjustment for the sampling of sites/vehicles was equal to the number of sites/vehicles in operation

on the survey day. For sites, the mean adjustment was 1.02, and the maximum was 55; for vehicles, the mean adjustment was 1.002, and the maximum was 17.

Adjusting for program visit nonresponse. Among the 16,709 assigned program visits, 12,511 were completed; counts of clients served during the assigned survey day/time were obtained for another 63 programs at which visits could not occur (e.g., drive-through pantry programs). The adjustment for program visit nonresponse was done separately within cells defined by program type within food bank; the adjustment factor was the reciprocal of the weighted response rate for the cell. Collapsing of cells was done in a few cases with no completed visits in the cell or very large adjustment factors; in those cases, for the national weights the collapsing was with the same program type from other food banks in the same state, and for the food bank weights the collapsing was with the other program type within the same food bank. The overall weighted program visit response rate was 78.8 percent. The lowest weighted program visit response rates were in food banks 015 (29.8%), 062 (30.3%), 222 (35.1%), 314 (35.3%), 188 (38.7%), and 117 (39.4%).

Adjusting for all within-program sampling. An adjustment was applied to account for the sampling of clients. In some cases, clients formed multiple lines at the program, and a single line was sampled. In such cases, the weights were adjusted to account for the number of lines. The mean adjustment for multiple lines was 1.008, and the maximum was 6. Next, an adjustment was applied to account for the systematic sampling of eligible clients reporting to receive services on the survey day. This adjustment was equal to the sampling interval; the mean adjustment (among programs) was 8.59 and the maximum was 100.

Adjusting for client nonresponse. The next adjustment compensated for eligible clients who failed to complete a client survey. Failure to complete a client survey could have been due to client refusal or nonresponse for other reasons (e.g., the client did not have enough time for the survey), or the result of a client who submitted a survey that was later deemed to not be complete enough for analysis purposes. As clients were sampled for the Client Survey, the data collectors recorded information about each sampled client's demographics (gender and broad age range). The original intention was to use these demographics (together with food bank and program type) to form cells for client nonresponse adjustment. For example, within a given food bank, all senior male grocery clients would form one cell; all non-senior male grocery clients would form another cell; all senior female grocery clients would form another cell; etc. However, there was a need to modify that plan due to missing site surveys (the mechanism used to transmit the information about sampled clients), extremely small cell sizes, large factors, and impossible situations (e.g., the site surveys indicated a

total of 8 senior male meal clients were sampled, but 10 completed surveys were received from senior male meal clients). Instead, adjustment cells were formed by program type within food bank (without regard to the demographics); the response rate was the reciprocal of the weighted response rate within the cell. The mean client response rate was 61.9 percent, and the lowest client response rates were for meal programs in food bank 279 (7.9%), meal programs in food bank 010 (13.5%), meal programs in food bank 222 (14.1%), meal programs in food bank 294 (15.9%), and meal programs in food bank 316 (19.1%). Although these low response rates result in some large factors, we decided not to collapse cells further or trim the factors themselves, since the next step in the weighting process was to trim the weights.

Trimming outlier client weights. Following the sequence of adjustments described above, excessive client weights were identified and trimmed. For each program type within each food bank, a trimming threshold was computed as $\sqrt{10 \frac{\sum w_i^2}{n}}$, where n is the number of completed client surveys for the given program type in the food bank, and w_i is the client weight (after all of the adjustments described above had been applied) for client i . (This is the same computation that was used to trim the weights in the HIA 2010 study.) For a program type within a given food bank, if any client weights exceeded this value (the trimming threshold), those weights were trimmed back to that value, and the excess (the total amount trimmed off the weights) was redistributed proportionately to all respondents in the program type in the food bank. Overall, 2 percent of the client weights were trimmed.

Trimming weights of clients in programs with excessive contributions. Following the trimming of the client weights, we compared the distribution of client weights by food bank to the distribution of total poundage by food bank. This comparison revealed some food banks whose share of the sum of weights far exceeded their share of poundage. Delving further into this, we observed that these food banks often included a few very large programs for which the estimated number of clients served by the program alone (without any of the aforementioned adjustments) seemed very excessive. (This estimated number of clients served was obtained by backing out of the client weight the program base weight and all adjustments for nonparticipating food banks, agencies, and programs.) As a result, we implemented a program-level trimming adjustment.

The program-level trimming was done in two steps:

- First, in situations in which the estimated number of clients served by the program annually exceeds 14,600 (a total corresponding to 40 clients per day for each of the 365 days in the year), if the estimated number of clients served by the program annually also exceeds the total of the 12 monthly client counts from the Agency Survey, we flagged the program as potentially eligible for trimming. This resulted in 899 programs in total (307 meal and 592 grocery) being flagged.
- Next, among the programs flagged in the above step, if a program's annual duplicated count was at or above the 90th percentile for annual duplicated counts among all programs of the given program type, then the client weights for all clients of that program were trimmed to yield a trimmed annual duplicated count equal to the higher of the following two values: (1) the 90th percentile cap, or (2) the duplicated count from the Agency Survey.

Table 3-1. Overall weighted response rates at each stage of data collection

Stage	Weighted response rate (%)
Food bank participation	90
Agency Survey	85
Program visit	79
Client	62

3.5 Standard Error Computation

As described in Section 1.1, the Agency Survey was designed to be a census of agencies in the Feeding America network. As a result, there is no sampling error, *per se*, in Agency Survey estimates. However, there was nonresponse to the Agency Survey, and some nonresponse models view nonresponse as a phase of selection¹². Thus, standard errors were computed to account for the variability in the Agency Survey estimates due to nonresponse error. For the Client Survey, the precision estimates for the survey estimates reflect the stages of sampling as well as the adjustments for nonresponse. Two broad classes of methods have been developed for computation of standard errors of estimates from complex sample surveys: (1) Taylor series linearization and (2) replication methods. The HIA 2014 data files contain the information necessary for analysts to use either of these approaches to compute standard errors of estimates.

¹² See, for example, Särndal, C.E., and Swensson, B. (1987). A general view of estimation for two phases of selection with applications to two-phase sampling and nonresponse. *International Statistical Review*, 279-294.

The Taylor series linearization approach uses a mathematical technique to approximate a nonlinear statistic with a linear form. The variance of the nonlinear statistic is approximated by the variance of the linear function. Software packages that use Taylor series linearization to estimate variances of statistics from complex sample surveys require the user to specify design information including “stratum” and “cluster” variables. Unlike those based on replication methods (discussed below), variance estimates based on Taylor series linearization do not fully capture the effects of all of the weight adjustments; however, in most cases, the differences between Taylor series standard errors and replication-based standard errors are not large. The “stratum” and “cluster” variables that allow users to compute variance estimates using Taylor series linearization are provided on the HIA 2014 datasets. The stratum is the food bank. In most cases, the cluster is the agency. In cases in which all agencies in a food bank were included in the sample or other cases in which an agency is sampled with certainty, if an agency has more than one eligible program and not all of the agency’s eligible programs were included in the sample, then the cluster is the program. In cases in which all agencies in a food bank were included in the sample or other cases in which an agency was sampled with certainty, if the agency has only one eligible program or all eligible programs of the agency were included in the sample, the cluster is the client.

Replication methods provide a relatively simple way of estimating variances and have some advantages over other variance estimation methods (e.g., linearization approach). The basic idea behind the replication approach is to identify subsets of the full sample (“replicates”), to calculate the estimate of interest for each replicate, and then to use the variability among these replicate estimates to estimate the variance of the full sample statistics. Different approaches can be used to create these replicates.

For HIA 2014, the replication approach that was used is Fay’s method¹³, a variation on balanced repeated replication where, instead of creating half-samples, the full sample base weights are down-weighted and up-weighted by factors of k and $(2-k)$; for HIA 2014, $k=0.3$.

With the replication approach, each of the weighting adjustments described in Section 3.3 were repeated for each of the replicates (separately), in addition to the full-sample weight. In this way, the replication variance estimates reflect the precision effects of all of these adjustments.

¹³ Judkins, D.R. (1990). Fay’s method for variance estimation. *Journal of Official Statistics*, 6(3), 223-239.

3.6 Survey Weights Provided in Analysis Files

The variables provided on the analysis datasets are intended for use with software that uses the replication approach to compute variance estimates for complex survey data. In these software packages, it is necessary to specify the replication method (in this case, Fay's method with Fay's $k = 0.3$), the full-sample weight variable, and the replicate weight variables.

3.6.1 Weights for Analysis of Agency Survey Data

The following is a list of weight variables included in the Agency Survey analysis files (both the Agency files and the Program files). In each case, the full-sample weight variable name is given first, followed by the corresponding replicate weight variable names in parentheses.

- Weight for national level analyses: WAN_W_NA (WAN_W_NA1 - WAN_W_NA100)
- Weight for food bank-level analyses: WAN_W_FA (WAN_W_FA1 – WAN_W_FA100)
- Weight for State-level analyses: WAN_W_SA (WAN_W_SA1 – WAN_W_SA100)

As noted above, the Agency Survey was designed as a census, so the variance estimates reflect variance due to nonresponse only. As a result, a finite population correction should be applied to the variance. (In this case, the finite population correction is equal to 1 minus the Agency Survey response rate.) The STRATRATES file contains these Agency Survey response rates, in the variable `_RATE_`. Thus, the software's instructions for specifying a finite population correction should be followed, using the `_RATE_` variable as the "sampling fraction."

3.6.2 Weights for Analysis of Client Survey Data

Exhibit 3-1 contains a list of weight variables included in the Client Survey analysis files. In each case, the full-sample weight variable name is given first, followed by the corresponding replicate weight variable names in parentheses.

Exhibit 3-1. List of weight variables included in the Client Survey analysis files

Weights for national level analyses		
	F_WCN_W_NCG_Mnth_HH (F_WCN_W_NCG_Mnth_HH1- F_WCN_W_NCG_Mnth_HH100)	Weight used to produce monthly household-level overall client count estimate and all other overall household-level estimates in report
	F_WCN_W_NCG_Mnth_MG_HH (F_WCN_W_NCG_Mnth_MG_HH1- F_WCN_W_NCG_Mnth_MG_HH100)	Weight used to produce monthly household-level client count estimates by type of program and all other household-level estimates by type of program in report
	F_WCN_W_NCG_Mnth_IN (F_WCN_W_NCG_Mnth_IN1- F_WCN_W_NCG_Mnth_IN100)	Weight used to produce monthly individual-level overall client count estimate and all other overall individual-level estimates in report
	F_WCN_W_NCG_Mnth_MG_IN F_WCN_W_NCG_Mnth_MG_IN1- F_WCN_W_NCG_Mnth_MG_IN100	Weight used to produce monthly individual-level client count estimates by type of program and all other individual-level estimates by type of program in report
	F_WCN_W_NCG_Ann_HH F_WCN_W_NCG_Ann_HH1- F_WCN_W_NCG_Ann_HH100	Weight used to produce annual household-level overall client count estimate
	F_WCN_W_NCG_Ann_MG_HH F_WCN_W_NCG_Ann_MG_HH1- F_WCN_W_NCG_Ann_MG_HH100	Weight used to produce annual household-level client count estimates by type of program
	F_WCN_W_NCG_Ann_IN F_WCN_W_NCG_Ann_IN1- F_WCN_W_NCG_Ann_IN100	Weight used to produce annual individual-level overall client count estimate
	F_WCN_W_NCG_Ann_MG_IN F_WCN_W_NCG_Ann_MG_IN1- F_WCN_W_NCG_Ann_MG_IN100	Weight used to produce annual individual-level client count estimates by type of program
Weights for food bank-level analyses		
	F_WCN_W_FCG_Mnth_HH F_WCN_W_FCG_Mnth_HH1- F_WCN_W_FCG_Mnth_HH100	Weight used to produce monthly household-level overall client count estimate and all other overall household-level estimates in report
	F_WCN_W_FCG_Mnth_MG_HH F_WCN_W_FCG_Mnth_MG_HH1- F_WCN_W_FCG_Mnth_MG_HH100	Weight used to produce monthly household-level client count estimates by type of program and all other household-level estimates by type of program in report
	F_WCN_W_FCG_Mnth_IN F_WCN_W_FCG_Mnth_IN1- F_WCN_W_FCG_Mnth_IN100	Weight used to produce monthly individual-level overall client count estimate and all other overall individual-level estimates in report
	F_WCN_W_FCG_Mnth_MG_IN F_WCN_W_FCG_Mnth_MG_IN1- F_WCN_W_FCG_Mnth_MG_IN100	Weight used to produce monthly individual-level client count estimates by type of program and all other individual-level estimates by type of program in report

Exhibit 3-1. List of weight variables included in the Client Survey analysis files (continued)

F_WCN_W_FCG_Ann_HH F_WCN_W_FCG_Ann_HH1- F_WCN_W_FCG_Ann_HH100	Weight used to produce annual household-level overall client count estimate
F_WCN_W_FCG_Ann_MG_HH F_WCN_W_FCG_Ann_MG_HH1- F_WCN_W_FCG_Ann_MG_HH100	Weight used to produce annual household-level client count estimates by type of program
F_WCN_W_FCG_Ann_IN F_WCN_W_FCG_Ann_IN1- F_WCN_W_FCG_Ann_IN100	Weight used to produce annual individual-level overall client count estimate
F_WCN_W_FCG_Ann_MG_IN F_WCN_W_FCG_Ann_MG_IN1- F_WCN_W_FCG_Ann_MG_IN100	Weight used to produce annual individual-level client count estimates by type of program
Weights for state-level analyses	
F_WCN_W_SCG_Mnth_HH F_WCN_W_SCG_Mnth_HH1- F_WCN_W_SCG_Mnth_HH100	Weight used to produce monthly household-level overall client count estimate and all other overall household-level estimates in report
F_WCN_W_SCG_Mnth_MG_HH F_WCN_W_SCG_Mnth_MG_HH1- F_WCN_W_SCG_Mnth_MG_HH100	Weight used to produce monthly household-level client count estimates by type of program and all other household-level estimates by type of program in report
F_WCN_W_SCG_Mnth_IN F_WCN_W_SCG_Mnth_IN1- F_WCN_W_SCG_Mnth_IN100	Weight used to produce monthly individual-level overall client count estimate and all other overall individual-level estimates in report
F_WCN_W_SCG_Mnth_MG_IN F_WCN_W_SCG_Mnth_MG_IN1- F_WCN_W_SCG_Mnth_MG_IN100	Weight used to produce monthly individual-level client count estimates by type of program and all other individual-level estimates by type of program in report
F_WCN_W_SCG_Ann_HH F_WCN_W_SCG_Ann_HH1- F_WCN_W_SCG_Ann_HH100	Weight used to produce annual household-level overall client count estimate
F_WCN_W_SCG_Ann_MG_HH F_WCN_W_SCG_Ann_MG_HH1- F_WCN_W_SCG_Ann_MG_HH100	Weight used to produce annual household-level client count estimates by type of program
F_WCN_W_SCG_Ann_IN F_WCN_W_SCG_Ann_IN1- F_WCN_W_SCG_Ann_IN100	Weight used to produce annual individual-level overall client count estimate
F_WCN_W_SCG_Ann_MG_IN F_WCN_W_SCG_Ann_MG_IN1- F_WCN_W_SCG_Ann_MG_IN100	Weight used to produce annual individual-level client count estimates by type of program