

SUPPORTING STATEMENT PART B FOR

Second National Household Food Acquisition and Purchase Survey (FoodAPS-2) Field Test

**Food Assistance Branch
Food Economics Division
Economic Research Service
US Department of Agriculture
1400 Independence Avenue, SW, Mail Stop 1800
Washington, DC 20250-1800
Phone: 202-694-5398
Fax: 202-694-5661
E-mail: jeffrey.gonzalez@usda.gov**

Table of Contents

<u>Section</u>	<u>Page</u>
Part B. Collections of Information Employing Statistical Methods.....	1
B.1 Respondent Universe and Sampling Methods.....	1
Respondent Universe.....	1
Sample Sizes and Expected Response Rates.....	6
B.2 Procedures for the Collection of Information.....	10
Statistical Methodology.....	10
Estimation.....	15
Degree of Accuracy.....	17
Specialized Sampling Procedures.....	18
Any Use of Periodic (Less Frequent Than Annual) Data Collection Cycles to Reduce Burden.....	22
B.3 Methods to Maximize Response Rates and to Deal With Issues of Nonresponse.....	22
Maximizing Response Rates.....	22
Nonresponse Bias Analysis.....	24
B.4 Test of Procedures or Methods to be Undertaken.....	25
B.5 Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data.....	28
References.....	30

Table

Table B.1. Field Test Universe Percentages and Expected Sample Sizes.....	6
---	---

Table B.2.Assumed Response Rates ¹ for the Field Test	7
Table B.3. Comparison of assumed field test response rates vs. comparable NFS Pilot response rates, unweighted.....	8
Table B.4. Screener Sample Size per Case for a Simple Screening Exercise	11
Table B.5a. Standard errors and coefficients of variation for average household food acquisitions during survey week, NFS Pilot.....	17
Table B.5b. Standard errors and coefficients of variation for estimated percentage of overall perception of Food Log for primary respondents, NFS Pilot.....	18
Table B.6a. Incentive Plan for the Field Test.....	20
Table B.6b.Incentive Experiment for the Field Test	21
Table B.7. Minimum Detectable Differences for the Incentive Field Test, n=430 Households.....	21

Figures

- B.1. Response rate declines for major in-person surveys
- B.2. Flow of Sample Design

Attachments

- U. Final Analysis Report and Recommendations for Revisions to the FoodLogger from the First Round of Usability Testing
- V1. Welcome to NFS Video Script
- V2. Getting Started Video Script
- V3. Entering Grocery Store Food and Drink Video Script
- V4. Entering Restaurant Food and Drink Video Script
- V5. Completing Profile and Income Questionnaire Video Script
- V6. Completing the Day Overview Video Script
- V7. Grocery Store Purchase Video Script
- V8. How to Report Meals from a Fast-Food Restaurant Video Script
- X. Final Analysis Report and Recommendations for Revisions to the FoodLogger from the Second Round of Usability Testing

Part B. Collections of Information Employing Statistical Methods

B.1 Respondent Universe and Sampling Methods

Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection method to be used. Data on the number of entities (e.g., establishments, state and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.

Respondent Universe

The respondent universe is described in this section with a general description of the sampling methods. More details on the sampling methods are provided in Section B.2.

The primary goal of the Field Test is to evaluate and finalize main survey design procedures and data collection protocols for Full Survey data collection. That is, the Field Test is a “dress rehearsal” for the Full Survey data collection. It will help detect any problems that may arise in the planned survey design and/or processing system. This dress rehearsal may also provide an opportunity to obtain data on survey costs and estimate population variances. Estimates of more current population variances for the sample design under consideration would help to better predict the sampling errors of the proposed sample plans in the Full Survey, since the FoodAPS-1 would have been conducted almost ten years ago by the time this field test begins. Current administrative costs, including average interview times as well as cash outlays, are obviously useful in deciding staffing and budgetary arrangements. In addition, the Field Test will include the implementation of

the new native smartphone application, FoodLogger. FoodLogger will serve as the primary mode of data collection and was designed to reduce respondent burden and improve data quality.

The Field Test will not be a complete “dress rehearsal” for the Full Survey, because in the Field Test a debriefing interview will be used in the place of a final interview in the planned Full Survey. The debriefing interview will collect information from respondents about their experiences and participation in study, including, but not limited to, whether the smartphone application was easy or difficult to use, what they found difficult about the survey protocol, and whether they’d participate in a similar survey in the future. This information will help hone the survey protocols for the Full Survey and enable the Field Test to provide meaningful information for the Full Survey about important or new design features (such as the app, the use of the mail screener, adaptive survey design and so on), under current design constraints and for similar target populations. Such information is critical to the successful implementation of a complex study such as the FoodAPS-2 Full Survey.

The implementation of the Field Test will also aim to create a database that allows the following research question to be answered: What is an optimal incentive strategy among strategies considered that will stem the drop-off in reporting across the week? It is critical to the success of FoodAPS that respondents complete the Food Log each day of the data collection week. Both FoodAPS-1 and the National Food Study (NFS) Pilot¹ suffered from reporting attrition throughout the data collection week; a goal for the Full Survey is to reduce this attrition rate and to minimize the potential for nonresponse bias. Therefore, the Field Test will include an incentive experiment that varies the amount of incentive a respondent can receive for Food Log completion during the week. To evaluate the effectiveness of this protocol, a split sample incentive design is proposed. The incentives experiment is described in the section titled “Specialized Sampling Procedures”.

¹ The National Food Study Pilot was also known internally at USDA as the Alternative Data Collection Method (ADCM) Pilot. For background information on the ADCM Pilot, see <https://www.reginfo.gov/public/do/DownloadDocument?objectID=68178702>.

The Field Test will also give us a preview of the proportion of Food Log entries completed on the app (FoodLogger), on the web, and by the phone. In particular, the uptake rate of the FoodLogger will inform the Full Survey's expectations. We expect about 22%, or about 95 households, to use the web mode to complete their Food Log entries during the survey week². To some extent we can examine the burden and completeness between those using FoodLogger, i.e., the smartphone application, and those completing the Food Log entries on the web³. In particular, we can assess whether data completeness and accuracy differ by response mode after controlling for demographic characteristics of respondents who responded to different modes. We can also broadly compare the characteristics of respondents across each mode of administration; this comparison may provide indications of potential impacts on bias and completion rates by mode and demographic characteristics in the Full Survey, allowing us to proactively address any potential issues.

Although the FoodAPS-2 Full Survey sample design ultimately needs to provide adequate precision for key survey estimates in seven domains, producing reliable estimates for those seven domains is not a primary goal of the Field Test. Some of the domain-specific sample sizes will likely be quite small, however, acquisition patterns can still be observed to a limited extent and described by analytic domain and this information may provide useful inputs for sample design planning for the Full Survey. The seven analytic domains for FoodAPS-2 (defined for each household at the time of the survey) are households: (1) SNAP, (2) WIC, (3) Income-to-poverty ratio at or below 130 percent of the poverty guideline that are not SNAP or WIC, (4) Income-to-poverty ratio above 130 percent of the poverty guideline that are not SNAP or WIC, (5) With children, (6) Without children, and (7) All households. The last three domains are overlapping with the first four domains.

² Expectations regarding the distribution of respondents using the smartphone, web, and phone to complete their Food Log entries during the survey week are derived from the ADCM Pilot.

³ We expect the percentage of cases completing their Food Log entries by phone to be small (3% or less) so this subgroup will be excluded from any analyses comparing data collection modes with respect to respondent burden and data completeness.

To reduce cost and save time, we will conduct the Field Test on 430 households in the same 12 Primary Sampling Units (PSUs), the same 120 Secondary Sampling Units (SSUs), and use the same address sampling frames within SSUs in which the NFS Pilot was conducted. Therefore, to describe the Field Test sample design, the following background on the FoodAPS-1 and NFS Pilot sample designs are provided.

Background on NFS Pilot Sampling Methodology

The sample for the NFS Pilot was selected late in 2016 and data were collected in the first half of 2017. One of the objectives of the NFS Pilot was to evaluate results against FoodAPS-1 results, therefore the NFS Pilot sample data was collected in a random subset of areas where FoodAPS-1 data were also collected. For the NFS Pilot, the selection occurred in three stages. The first stage was a subsample of 12 PSUs from the 50 PSUs selected for FoodAPS-1, where PSUs were defined as counties or groups of counties. The FoodAPS-1 PSUs were nationally representative and selected using a stratified probability proportionate-to-size (PPS) design. In FoodAPS-1, there was one PSU selected with certainty, and the other 49 PSUs were selected using Chromy's method of sequential random sampling (Chromy, 1979) in SAS Proc SurveySelect. Prior to selection, the PSUs were sorted by metro status and region, which was defined by seven USDA Food and Nutrition Service (FNS) administrative regions. The PSUs were selected using a composite measure of size that was a function of the estimated number of households in key sampling domains (based on SNAP participation and income level) and their associated overall sampling rates. The estimated number of households by sampling domain was derived from the American Community Survey (ACS) Public Use Microdata Sample (PUMS) files.

To select PSUs for the NFS Pilot, the one certainty PSU in FoodAPS-1 was assigned a measure of size of 1.5 to reflect the multiple selection hits from probability proportionate to size sampling, and the other PSUs each received a measure of size of 1. By assigning the measure of size in this manner for the NFS Pilot, the approach retained the key features of the composite measure of size assigned to PSUs in FoodAPS-1. To ensure a wide

representation of different characteristics of geographic areas, the 50 PSUs were sorted by FNS region, urbanicity, and the percentage of the population on SNAP. The subsample of 12 PSUs obtained a diverse mix of communities (e.g. having both urban and rural communities) to ensure the web-based data collection procedures were tested under a variety of real-world conditions. The subsample was selected using systematic sampling from the sorted list, and probabilities proportionate to the measure of size. In the end, there were seven FNS regions and nine states that had at least one selected PSU. There were two PSUs that contained a mix of counties inside and outside metropolitan/micropolitan statistical areas.

In the second stage of sampling for the NFS Pilot, an average of 10 SSUs (individual block groups or combination of adjacent block groups) were selected per PSU.⁴ While the NFS Pilot included fewer PSUs than FoodAPS-1 for cost efficiency, the number of sampled SSUs within a PSU was increased to 10 from 8 to spread the sample across the sampled PSU and cover diverse communities. The SSUs were selected using a similar composite measure of size as in FoodAPS-1, where the composite measure of size was calculated as a function of the SSU population count and the sampling fraction of the SNAP domain, and other sampling domains. The SSU population count was based on the 2010-2014 ACS summary file data. The composite measure-of-size gives SSUs with a larger number of SNAP participants a higher chance of selection.

In the third stage, residential addresses were selected within the selected SSUs from two strata (addresses on the state's SNAP address list and addresses not on the SNAP address list). To do so, a sampling frame of addresses was obtained from two sources of address listings. The first source of address listings was a list of all residential addresses from a vendor-maintained address-based sampling (ABS) frame based on the most recent United States Postal Service (USPS) computerized delivery sequence (CDS) file. The USPS address lists included street addresses along with the carrier route information from a qualified vendor who receives updated lists from the USPS on a bimonthly or weekly basis. For the second source for creating

⁴ On average there are about 69 block groups per county in the US with about 604 housing units in each.

the frame of addresses, Westat obtained lists of addresses with SNAP participants from the states. The request was for lists as of June 30, 2016. A matching operation was conducted to determine which ABS addresses were not on the SNAP administrative list of addresses. The matching operation effectively stratified the addresses into two strata, 1) SNAP list addresses and 2) addresses not on the SNAP list. The addresses were sorted geographically⁵, and a systematic random sample of addresses was selected within each stratum from these lists, with addresses on the SNAP administrative lists sampled at a higher rate. Once addresses were selected, a screening interview was conducted to identify the sampling domain of both the SNAP and non-SNAP households. A random proportion of those who were classified as non-SNAP in the highest income category were subsampled to reduce the number of high-income participants in the sample. The remaining households were eligible for the main survey.

As mentioned above, the same sampling frames constructed for the NFS Pilot will be used for the Field Test, with the same stratification of SNAP and non-SNAP addresses, while ensuring to avoid addresses selected in the NFS Pilot by removing the NFS Pilot addresses from the sampling frame prior to selecting the Field Test sample. The address lists will be at least five years old by the time the Field Test is conducted. As a quality check, the total number of addresses will be compared to counts from an address vendor by SSU. If the counts differ significantly, a new ABS address list will be requested. Due to the age of the SNAP administrative lists, there will be uncertainty about the number of SNAP respondents due to changes in household SNAP participation. Two approaches will be implemented to help reduce the uncertainty. The first approach will involve further stratifying the addresses, where the selected SSUs will be classified into high or low concentration, based on the percentage of SNAP participating households from the most recently available American Community Survey. A higher selection rate will be assigned to strata with higher concentration. The sampling rate will be initially assigned to three times the rate for lower concentration areas to constrain the impact of unequal selection rates on

⁵ The geographic sort used the following when available: ZIP, carrier route, walk sequence, and street name. Street name was used for the SNAP households that were not on the ABS list and have missing values of carrier route and walk sequence.

variance estimates. The design effect due to assigning unequal selection rates will be reviewed and sampling rates will be adjusted if the design effect is greater than 1.3. The second approach is to include a question on SNAP participation status on the mail screener.

The respondent universe percentages of households in each sampling domain from the 2019 Current Population Survey, out of the approximately 129 million households in the nation, are shown in Table B.1. Also shown in Table B.1 are the expected sample sizes for the Field Test, which are like those that resulted from the NFS Pilot. The oversample of SNAP households is accomplished as described above -- using SNAP lists from state agencies from the NFS Pilot, using a mail screener, and using an oversample of addresses in high concentration of SNAP households in selected SSUs. There is no need to request new administrative SNAP lists for the Field Test, since 1) we have procedures for cleaning, matching, and forming strata using the SNAP and address-based sample lists developed in the NFS Pilot, and 2) the above sampling procedures address potential concerns regarding the age of these lists. For the Field Test, there will be no direct oversample procedure for WIC participants; however, we expect 12 percent of the SNAP households to also participate in WIC in addition to other households in the ABS frame.

Table B.1. Field Test Universe Percentages and Expected Sample Sizes

Analysis domain	Estimated population # of households (in millions)^a	Estimated population percent^a	Estimated # of respondent households
SNAP households	12	9%	100
WIC households	3	2%	14
Non-SNAP and non-WIC households with income less than 130% poverty guideline	13	10%	83
Non-SNAP and non-WIC households with income above 130% of poverty guideline	102	79%	245
Total	129	-	430

^a Current Population Survey 2019

NOTE: The SNAP and WIC analysis domains are not mutually exclusive, so counts do not sum to totals.

Sample Sizes and Expected Response Rates

We plan to target the same nominal overall sample size (430 completed final interviews) as the NFS Pilot. Table B.1 provided the target respondent sample sizes by domain, where like the NFS Pilot, has an oversample of SNAP households. The definition of a completed household data collection for the Field Test will be based on the same definition of complete used for the NFS Pilot. The criteria for the Field Test are:

- The screening interview must be completed to the point where survey eligibility was determined.
- The Initial Interview must be completed.
- The Debriefing Interview must be completed.

To obtain 430 completed interviews, we plan to mail out 4,125 mail screeners. An additional 700 households sampled from administrative lists will go directly to the in-person screening stage for a total of 4,825 unique households sampled. We performed a preliminary evaluation based on FoodAPS-1 and NFS Pilot data to test whether, given reasonable assumptions related to response, contact and eligibility displayed in Table B.2, the planned number of mail screeners was likely to result in the target number of completes by domain. This evaluation was also used to set initial sampling rates. The evaluation is described in detail in this section.

Table B.2. Assumed Response Rates¹ for the Field Test

Stage	Expected sample size	Response Rate
Mailout sample	4,125	
Mail screener responses	825	20.0%
Screen-in rate for mail responses	582	70.5%
Mail screener nonresponse	3,300	80.0%
Subsample of nonrespondents	1,100	33.3%
Selected from admin lists (no mail screener)	700	-
Total for in-person screening	2,382	-
Occupied DU rate	2,025	85.0%
Screener completions	733	36.2%

Screened-in cases	660	90.0%
Initial Interview respondents	468	70.9%
Debriefing Interview respondents	430	92.0%

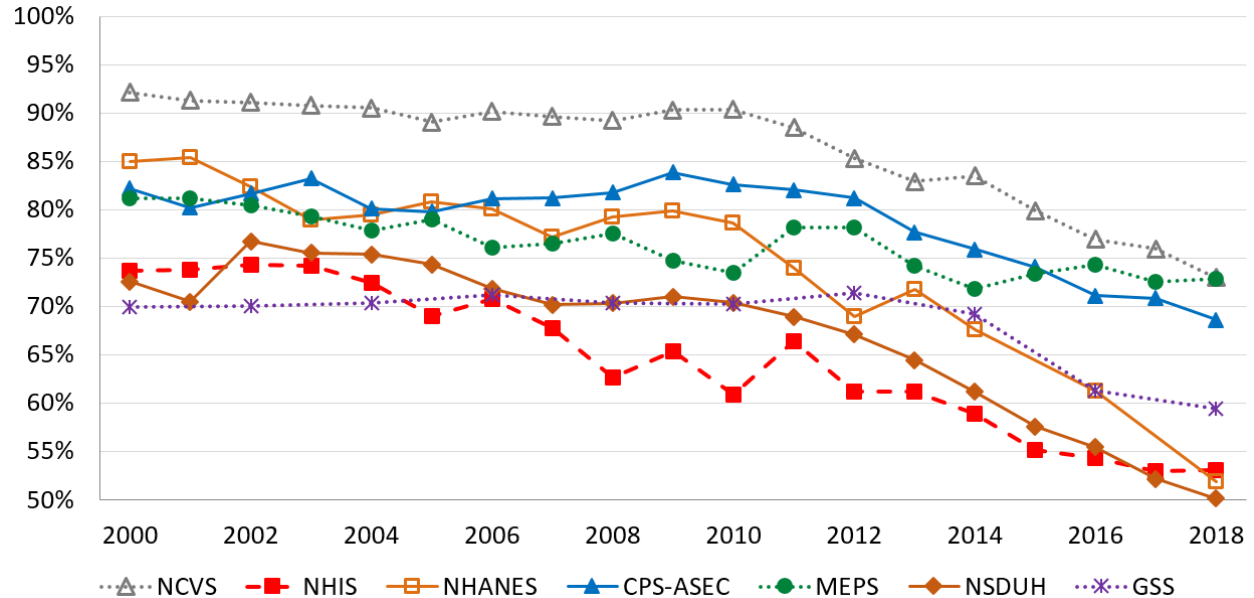
¹ Response rates will be computed based on the American Association for Public Opinion Research Response Rate 3 (RR3) formula, given in https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/Standard-Definitions2015_8thEd.pdf.

The assumed response rates in the evaluation for the Field Test in-person screener, Initial Interview, and Debriefing Interview are like the NFS Pilot response rates as shown in Table B.3. The approaches discussed in Section B.3 are designed to achieve a higher rate, however it is balanced with continued evidence of response rate decline since 2017 (Figure B.1). The COVID-19 pandemic may also impact response rates. For example, in a videoconference with the Agency for Health Research and Quality and Westat on August 11, Stephen Blumberg (National Center for Health Statistics) reported that for the National Health Interview Survey they had a 55 percent response rate for the adult survey in quarter 1 2020 for all in-person interviews. In quarter 2, the mode of collection switched to telephone and there was at least a 15-percentage point drop-in response rate. In quarter 3 the two modes are mixed, and the response rate is estimated to be about 46 percent, which is in between quarter 1 and quarter 2.

Table B.3. Comparison of assumed field test response rates vs. comparable NFS Pilot response rates, unweighted

	Assumed Field Test Response Rate (Lower Bound)	NFS Pilot Unweighted Response Rate	Comments
Screener	36.2	36.2	Same as NFS Pilot
Initial Interview	70.9	70.9	Same as NFS Pilot
Debriefing Interview	92.0	90.9	Slightly higher than in NFS Pilot due to enhanced procedures (follow-up calls)
Overall	23.6	23.3	

Figure B.1. Response rate declines for major in-person surveys



NCVS: National Crime Victimization Survey, NHIS: National Health Interview Survey, NHANES: National Health and Nutrition Examination Survey, CPS-ASEC: Current Population Survey – Annual Social and Economic Supplement, MEPS: Medical Expenditure Panel Survey, NSDUH: National Survey on Drug Abuse and Health, GSS: General Social Survey.

Source: Internal Westat presentation on October 17, 2019, entitled “Connect, Convince and Convert – Current Trends and Training on Gaining Cooperation” by Williams, Woods and Genoversa.

The evaluation to assess whether the planned number of mail screeners was likely to result in the target number of completes by domain used the NFS Pilot sampling frame and simulated the planned sampling process. The NFS Pilot frame has 91,953 dwelling unit records and includes SNAP status indicators, drawn from administrative lists obtained in July of 2016. Because the lists will be about six years old at the time of data collection, we assumed that only half of addresses sampled as SNAP would still be SNAP addresses⁶. We generated a “final” income domain for each record using the population prevalence rates from Table B.1. We also simulated a “screener” income domain for each record, assuming misclassification patterns similar to those in FoodAPS-1 and the NFS Pilot. Misclassification issues in FoodAPS-1

⁶ It is important to note that by “SNAP addresses” we do not mean that the same SNAP-eligible HH resides at the address as at the time of the ADCM. The household could have moved since the ADCM; as long as the current HH residing at that address receives SNAP, the address will count as a SNAP HH for the Field Test. WIC addresses were also collected at the time of the NFS Pilot, however, these were not used for sample selection purposes.

were studied in Krenzke and Kali (2016). Misclassification arises when a household classification into analytic domains from the short screener questionnaire is different from their classification based on the more thorough questions in the main survey. One example is that the screener response to household income tends to be lower than the sum of individual items on income that are asked during the main survey. This is a concern when attempting to achieve the sample size goal in the low population percentage in the low-income non-SNAP non-WIC domain.

As mentioned above, the evaluation was used to assign sampling rates. To do so, for the mail screener, the subsampling stratum for households not participating in SNAP or WIC was broken down into five categories, with cut points at income-to-poverty ratios of 100 percent, 130 percent, 185 percent, and 300 percent. More granular screener information on income categories will control misclassification. For example, based on previous data, we can assume that 50 percent of households that screen in at 100 percent income-to-poverty will have a final income-to-poverty ratio less than 130 percent, but the comparable rate is only 30 percent for households that screen in with income-to-poverty ratios between 100 percent and 130 percent. In the evaluation, we selected a simulated sample of 4,000 cases for mail screening (since the time of the evaluation, we increased the total sample for mail screening to 4,125). We subsampled cases at each stage according to the assumed response rates given in Table B.3.

In the evaluation, at each screening stage (mail screener and in-person screener), we subsampled cases with differential rates according to their subsampling stratum. We took all cases in the SNAP, WIC, and low-income (income-to-poverty ratio less than 130 percent) subsampling strata with certainty at both stages. We sampled higher-income cases to achieve the desired overall subsampling rate, but we set the sampling fractions within each income band to be proportional to the square root of the prevalence of the low-income domain in that band (Kalton and Anderson, 1986; Kalton, 2009). This is an efficient allocation that allows us to select more cases in lower income bands and reduce the number of high-income cases selected.

We simulated this sampling process 500 times to evaluate whether the proposed Field Test sampling procedure would consistently achieve the target sample sizes at each stage. We found that the target of 430 completed debriefing interviews was achieved in 70 percent of simulation runs, with a minimum final sample size of 397 and a maximum of 466. The average number of households in the final low-income domain was 83, with a minimum of 57 and a maximum of 117. This minimum is higher than the 45 low-income households in the NFS Pilot, which is promising. In the simulation, the number of in-person screeners necessary is consistent with our expectations, ranging between 2,423 and 2,506 with an average of 2,465. The simulations show that, if our response rate assumptions are reasonably accurate, the planned initial sampling rates are likely to produce the required sample.

The planned sample monitoring procedures will allow us to continually monitor whether the assumed response rates are reasonable. There is uncertainty about the impact that COVID-19 has on response rates, and therefore a reserve sample will be selected, sample yield will be closely monitored, and more sample will be released in portions relative to need to achieve the desired 430 completes. Screener subsampling rates will be reduced if a surplus is expected.

B.2 Procedures for the Collection of Information

Describe the procedures for the collection of information including:

- **Statistical methodology for stratification and sample selection,**
- **Estimation procedure,**
- **Degree of accuracy needed for the purpose described in the justification,**
- **Unusual problems requiring specialized sampling procedures, and**

- **Any use of periodic (less frequent than annual) data collection cycles to reduce burden.**

Statistical Methodology

FoodAPS-2, as did FoodAPS-1, seeks to represent several population groups that have a low population percentage (e.g., below 15 percent) and are not easily identified from commonly available information, specifically, the SNAP, WIC, and non-SNAP non-WIC low-income household domains. SNAP, WIC, and low-income non-SNAP non-WIC households would require large initial sample sizes under simple random sampling. Table B.4 shows the prevalence rates, as well as the resulting screening sample size that would be required to identify one household in the domain, for each of the first four key analytic domains specified for the FoodAPS-2 Field Test. As can be seen from this table, WIC households are particularly rare, with a prevalence rate of two percent. If one were to apply a simple screening exercise, 50 households would need to be screened to identify one WIC household. Much less screening would be needed to identify a SNAP household and a low-income household, but still the screening task is sizable.

Table B.4. Screener Sample Size per Case for a Simple Screening Exercise

Analytic Domain	Prevalence rate	Screening sample size per case
1. SNAP households	9%	11
2. WIC households	2%	50
3. Households with income-to-poverty ratios at or below 130 percent of the poverty guideline that do not participate in SNAP or WIC	10%	10
4. Households with income-to-poverty ratios above 130 percent that do not participate in SNAP or WIC	79%	1.3

Note: Population percentages were estimated from the 2019 Current Population Survey.

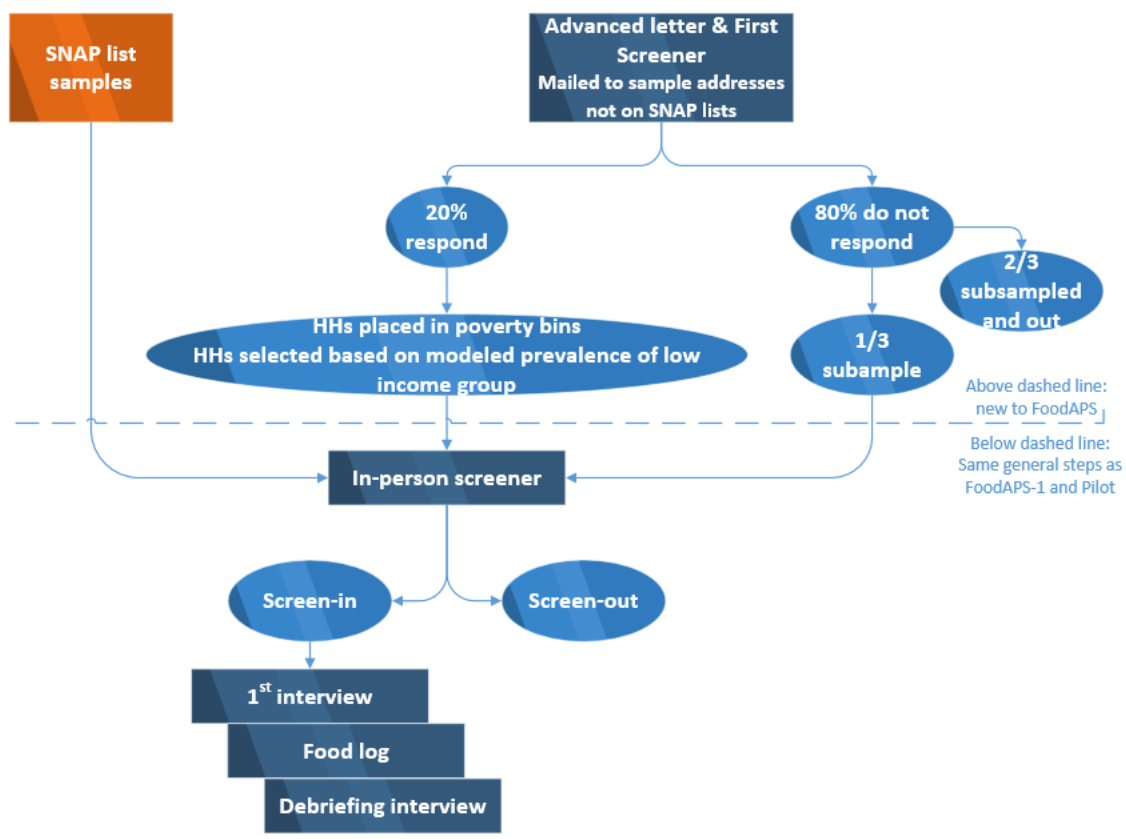
A second challenge associated with these domains is the potential for misclassification. The degree of misclassification is a parameter that affects the effective sample size if sampling domains are sampled at different rates. Misclassification occurs when a sampled household is classified at any stage of the sampling process into a sampling domain other than its final analytic

domain. This can occur because of out-of-date administrative information or inaccurate self-identification. The planned procedures for the Field Test are designed to address the challenges discussed above. The multi-stage sample consists of a sample of PSUs, a sample of SSUs within those PSUs, and then a sample of dwelling units within the SSUs. As mentioned in Section B.1, to reduce cost and save time, we will select the sample of dwelling units in the same PSUs and SSUs as used in the NFS Pilot, and use the same frame and stratification of dwelling units (DUs). Much of the background on the NFS Pilot sample was provided in Section B.1.

As in the NFS Pilot, we will apply systematic sampling, with households on the SNAP address list sampled at a higher rate. For the households not on the SNAP address list, we will draw an initial sample and send sampled households a mail screener asking their income level, current WIC participation status, and current SNAP status. The mail screener will help to efficiently classify cases prior to in-person visits. For example, any household selected from the non-SNAP household list that responds as a SNAP participant will be kept in the sample for in-person screening. A subsample of high-income households will occur prior to the in-person screener being conducted. This process also will avoid overlap between the NFS Pilot and Field Test samples by removing the NFS sample from the selection process. Sampling weights will be calibrated to the original number of addresses in 2016 in the NFS areas. The calibration adjustment is expected to be small due to the low sampling rates. That is, the impact on estimates due to removing NFS cases prior to sampling is expected to be minimal, because the NFS sampling frame contains 91,953 dwelling units, and the sampling rates across the strata (defined by SNAP/non-SNAP within each SSU) were very low, ranging from 0.008 to 0.213 with a median of 0.022.

The sample screening flow for the Field Test is shown in Figure B.2. The challenge in FoodAPS-1 and the NFS Pilot was classifying the households into the correct income domain based on screener information alone. The resulting misclassification of the low-income group heavily drives the sample size needed to achieve the target number of completes for this group and, thus, survey costs.

Figure B.2. Flow of Sample Design



For addresses that cannot be classified into a domain based on the administrative data, a mail screener will be used to screen the households. We will send a screener and an advance letter to the sampled addresses explaining the study. Using the mail screener information and frame data⁷ in an income prediction model, we will classify households into an income domain. The income prediction model will incorporate mail screener income domain along with other information about the household to predict the household's final income domain more accurately. The estimation of the model parameters from FoodAPS-1 will be validated using NFS Pilot data. That is, we plan to use the FoodAPS-1 data to predict data collected in the

⁷ Variables considered for the model will include, but are not limited to: tract-level demographic, income, and housing information from ACS, administrative data on food access/food desert status, paradata such as number of screener attempts and interviewer observations, and household screener responses for household size, household demographics (e.g., overall HH size, anyone less than 18 years old in HH, anyone age 65 and older in HH, and reported screener income group (including reported SNAP and/or WIC status)).

NFS Pilot data. Before data collection begins, we will explore data from FoodAPS-1 and the NFS Pilot, including area-level characteristics, and relevant household characteristics collected by the screener and initial interview to identify the strongest additional predictors of final income domain. The same model and estimated parameters would be used to predict responses for the Field Test. If the same model structure is used for the Full Survey, the parameters can be re-estimated using the Field Test data.

The income prediction model will aim to reduce the misclassification and improve the efficiency of the sampling process before we send interviewers out to screen households in person. The motivation for the income prediction model is as follows. Both FoodAPS-1 and the NFS Pilot subsampled screened cases for the main study based on assigned screener income domain alone. However, while the existing screener is effective at identifying SNAP and WIC households, it is much less effective at separating out households by income domain; only 36 to 46 percent of households that screened into a low-income domain in FoodAPS-1 or the NFS Pilot remained in that domain after the final interview. Improving this rate would mean that we are identifying low-income households more accurately and would reduce screening costs as well as reduce weight variation due to misclassification (and thus increase effective sample sizes). This model will allow us to generate timely and improved predictions of final income domain for each case while the survey is still in the field.

Due to their prevalence, it is likely more households with incomes above 130 percent of the poverty level will respond than are needed. We will subsample some of these households. To reduce bias due to nonresponse, we will continue to include one-third of the households that do not respond to the mail screener (income level remains unknown). The selected addresses will be administered an in-person screening interview to identify the sampling domain of both the SNAP and non-SNAP households. That is, those cases that screen-in via the mail screener, and the one-third of the households that do not respond to the mail screener, will be assigned to a field interviewer who will visit the address and conduct a second screener in-person that will ultimately determine the household's eligibility status to participate in the

study. Among the screener completes, all who report to be on SNAP will be retained for the initial interview. A random proportion of those who are classified as non-SNAP in the highest income category will be screened out to reduce the number of high-income participants in the sample. The remaining households will continue on with the survey.

Approaches to Reduce Coverage Error. We will conduct an operational test on two approaches to reduce coverage error. First, research presented in Link (2010) has shown that the Computerized Delivery Sequence (CDS)-based ABS lists have about a 98 percent coverage rate for residential addresses overall. However, some of the addresses are not usable for in-person surveys because the interviewers could not locate them if they were sampled. These unlocatable addresses include P.O. boxes, rural routes, highway contract boxes, and simplified addresses. These unlocatable addresses, if dropped from the sampling frame, create some noncoverage for which compensation is needed. Prior studies have estimated the resulting coverage rate to be approximately 92 percent (Kalton, Kali, and Sigman, 2014). Unlocatable addresses are particularly prevalent in certain highly rural areas. However, our plan is to keep the unlocatable addresses and send the mail screener which includes a request for the respondent to provide a physical street address. That way the coverage is effectively maintained at 98 percent.

Second, there will be more than one household living at some sampled addresses. In order to capture them in the sample, we will apply our Hidden DU procedure. We want to be sure that every household has been given a chance to participate in the survey. Therefore, in the screener questionnaire, we will apply the Hidden DU procedure by asking if there any other person or group of people who lives and receives mail {here/at ADDRESS}, and add them to the sample.

From the households responding to the mail screener, sampled households on SNAP lists, and a subset of non-responding households, we will select the sample for in-person data collection. We will assign these cases to a field interviewer who will visit the address and conduct a second screener in

person that will ultimately determine the household's eligibility status to participate in the study.

Estimation

The estimation process for the FoodAPS-2 Field Test will include weighting and variance estimation. Survey weights will be used when calculating statistics and estimating model parameters.

To help identify potential issues with data quality, and although not the primary analytic goal of this Field Test, it may be helpful to check results against the NFS Pilot and FoodAPS-1 samples. We fully acknowledge the inherent limitations in these comparisons given the potential differences in socio-demographic distributions from the two samples which were collected 5-6 and 11 years ago, respectively. If such comparisons are completed for the Field Test, the analyses will be appropriately caveated by noting these limitations. That said, for the Field Test we will still create sampling weights in the same manner as in the NFS Pilot⁸. Base weights will be constructed as the inverse of the selection probability to account for the differential probabilities of selection. To reduce nonresponse and noncoverage bias, the weights will be calibrated to control totals derived from the Current Population Survey Annual Social and Economic Supplement and American Community Survey. Minimal weight trimming will be performed within domains to decrease variation of the weights and prevent a small number of observations from dominating domain estimates. After trimming, the raking process will be repeated to produce the final household weights. The raking dimension will be the same as for the NFS Pilot: 1) Race/ethnicity of the primary respondent (Hispanic, non-Hispanic white race only, non-Hispanic black race only, all other non-Hispanic persons); 2) Household size (1, 2 to 4, 5 or more); 3) Number of children under 18 in the household (0, 1, 2, 3, 4 or more); 4) Whether one or more people 60 years of age and over reside in household (Yes, No); 5) SNAP participation (On SNAP, Not on SNAP); and 6) Annual household income (\$0 to \$14,999, \$15,000 to \$49,999, \$50,000 and over).

⁸ The NFS Pilot analysis report is not publicly available, but a redacted version is available upon request for further information.

Replicate weights will allow for variance estimation under a complex sample design. As in the NFS Pilot, we will create twenty-four replicates from the twelve PSUs using the jackknife (JK1) replication method. The replicate weights will be calibrated to control totals in the same manner as the full sample weights. Any analyses will use appropriate software (e.g., SAS[®] Proc SurveyReg) to make use of the replicate weights for variance estimation. We will report all statistics and models using survey weights.

For use in analysis, we will attach nutrient and Food Pattern Equivalent (FPE) values (Bowman, et al. (2018)) and location information of food acquisition. Respondents will record acquisitions of foods and beverages by scanning UPCs, entering Produce Look-Up (PLU) codes, or entering free text when they are unable to identify a match. In each case, we will ensure that the food item acquired matches food codes from the USDA databases (Food and Nutrient Database for Dietary Studies (FNDDS)⁹ and Food Data Central, formerly the National Nutrient Database for Standard Reference)¹⁰ so that we can append complete nutrient and FPE values to the resulting dataset. The methods to ensure these links vary depending on how the respondent reported the acquisitions. ERS' Linkages database provides established links between UPCs, USDA food codes, nutrients, and prices for many food items. Links among PLUs, UPCs, other IDs, nutrient database food codes, and USDA food categories fall into a grey area requiring matching and review. For location and distance information of food acquisition, we will determine travel distances to stores and restaurants by GIS and produce geocodes for a respondent's residence and frequented food places that will be obtained during data collection through Google Maps.

The evaluation of the FoodAPS-2 Field Test data will focus on three aspects of data quality: data completeness, accuracy, and response rates. Measures for assessing completeness include the average total number of items reported, number of events reported, number of free events reported, the

⁹ U.S. Department of Agriculture, Agricultural Research Service. 2020. *USDA Food and Nutrient Database for Dietary Studies 2017-2018*. Food Surveys Research Group Home Page, <http://www.ars.usda.gov/nea/bhnrc/fsrg>

¹⁰ Food Data Central: U.S. Department of Agriculture, Agricultural Research Service. FoodData Central, 2019. fdc.nal.usda.gov

expenditures reported, and the number of missing values for key variables. We will assess the accuracy of the data by examining the percentage of outliers and comparing to alternative sources of data such as administrative records and receipts. We will calculate weighted response rates from the point of screening to the completion of the debriefing interview. In addition, closely aligning with the NFS Pilot results, about 75 percent of the household completes are expected to complete their Food Log entries by smartphone, about 22 percent via web, and 3 percent by telephone. We will also examine data completeness and accuracy by response mode. Specifically, we will assess whether data completeness and accuracy differ by response mode (given that we expect less than 3% of respondents to report by phone, this analysis will focus on a comparison between the web and smartphone) after controlling for demographic characteristics of respondents who responded to different modes. To a certain extent, characteristics of respondents can be compared across each mode of administration to provide indications of potential impacts on bias and completion rates.

Degree of Accuracy

Because the Field Test sample design follows very closely to the NFS Pilot, the degree of accuracy will be about the same. Table B.5a provides the standard errors and coefficients of variation (CVs) for select food acquisition estimated averages from the NFS Pilot. The table shows that the CVs range from 7.4 percent to 10.1 percent across the eight estimates. Table B.5b provides the standard errors and CVs for the estimated percentage of overall perception of Food Log for primary respondents from NFS Pilot data. The table shows that the CVs range from 8.2 percent to 25.0 percent across the five estimates. Note that by calculation the CV is a relative term in that when a CV's denominator for percentages is small, the CV will be quite large. As shown in the table, the standard errors are reasonable.

Table B.5a. Standard errors and coefficients of variation for average household food acquisitions during survey week, NFS Pilot

Food acquisition estimates	Unweigh ted Count	Weight ed Mean	Standa rd Error	Coefficient of Variation
Average number of food events	430	10.64	0.903	8.5%
Average number of food items	430	33.68	2.493	7.4%
Average number of FAFH events	430	6.03	0.612	10.1%
Average number of FAFH items	430	12.00	1.150	9.6%
Average number of FAH events	430	4.61	0.396	8.6%
Average number of FAH items	430	21.68	1.651	7.6%
Average number of free events	430	2.72	0.225	8.3%
Average number of free items	430	5.72	0.489	8.6%

Table B.5b. Standard errors and coefficients of variation for estimated percentage of overall perception of Food Log for primary respondents, NFS Pilot

Overall perception of Food Log	Unweighted Count	Weighted Percentage	Standard Error	Coefficient of Variation
Primary respondents only				
"Very easy"/ "Never" had problems	189	47.67%	3.92%	8.2%
"Somewhat easy"/ "Rarely" had problems	84	22.99%	2.12%	9.2%
"Neither difficult nor easy"/ "Sometimes" had problems	40	11.24%	2.15%	19.1%
"Somewhat difficult"/ "Often" had problems	35	10.44%	1.84%	17.6%
"Very difficult"/ "Almost always" had problems	28	7.65%	1.91%	25.0%
Total	376	100.00%		

Specialized Sampling Procedures

Incentive: The NFS Pilot offered incentives for completing each household interview (Initial and Final), with additional incentives for each household member for completing the Food Log each day and the Income Worksheet (when applicable). Table B.6a (the top panel) displays the incentives for 1 to 4 person households, but incentive amounts for larger households can be extrapolated from the information provided. For example, a 5-person household in the NFS Pilot could earn a maximum of \$210: \$189 (the 4-person maximum given in the table) plus \$21 (\$3 times 7 days, the maximum an additional person could earn for Food Log completion).

We plan to assess a change in the incentive plan between the NFS Pilot and the Field Test. Table B.6a summarizes the revised incentive scheme that we plan to test for effectiveness in the Field Test (see the bottom panel), compared to the incentive scheme used in the NFS Pilot (see the top panel). To achieve the intended overall response rate, it is critical to maximize response rate to both the mail and in-person screeners. The Field Test is

adding two additional incentives at the screening stage to increase initial response rates. The first is a \$5 unconditional incentive to be mailed with the mail screener and the second is a \$5 incentive given to the household upon completion of the in-person screening interview. There will also be an opportunity to earn incentives for each of the interviews and for completing the Food Log each day. After the household completes the Debriefing Interview and returns any loaned equipment, we will mail the household a check for the total earned incentive. In addition, the Field Test incentive is adjusted to account partially for inflation and is therefore higher than the amount used for the NFS Pilot and FoodAPS-1.

For the incentives experiment randomization to either the control or experimental conditions will occur at the time of address sample selection. In particular each sampled address will be randomly assigned, following a Bernoulli trial, to either the control condition or the experimental condition so that roughly 50% of the sampled addresses fall into each category for the incentives experiment. All individuals within each sampled address (i.e., household) that is determined to be eligible for the study and agrees to participate will be assigned to the same condition that the address was assigned to at the time of sample selection. For example if an address was assigned to the control group during sample selection and if the household at that address is determined to be eligible and agrees to participate, then all members of that household will also be assigned to the control group.

Given that there are multiple stages of recruitment, each of which may result in a household or individual refusing to participate, when evaluating the incentives experiment ERS will assess the extent to which, if any, there are any differences across distributions of underlying characteristics of the households or individuals. These characteristics may include sample frame information or be information that was observed or reported during data collection. If there are discrepancies, ERS will identify those when reporting on the incentives experiment and as appropriate or warranted make statistical adjustments in the analysis to control for potential confounding.

For the analysis, there are two main comparisons that will be done. First, while a mail screener was not administered in the NFS Pilot, the in-person

weighted screener response rates from the Field Test (with a pre-paid mail screener monetary incentive and conditional in-person screener incentive) can be compared to that of the NFS Pilot (no monetary incentive). Second, as described further below, the attrition rate of food log reporting during the week will be compared among the Field Test experimental groups.

Table B.6a. Incentive Plan for the Field Test

HH Size	Mail Screener	In-Person Screener	Initial Interview + Training	Food Log ^a	Final Interview/Debriefing Interview+Bonus	Income Wksh ^t ^b	Profile Queue (\$2 pp)	Max Total (Full Survey)
NFS Pilot								
1	n/a	\$5	\$50	\$21	\$50	\$5	n/a	\$131
2	n/a	\$5	\$50	\$42	\$50	\$5	n/a	\$152
3	n/a	\$5	\$50	\$63	\$50	\$5	n/a	\$173
4 ^c	n/a	\$5	\$50	\$84	\$50	\$5	n/a	\$194
Proposed for Field Test								
1	\$5	\$5	\$40	\$35/55	\$16	\$2	\$2	\$105/125
2	\$5	\$5	\$40	\$70/110	\$16	\$4	\$4	\$144/184
3	\$5	\$5	\$40	\$105/165	\$16	\$6	\$6	\$183/243
4 ^c	\$5	\$5	\$40	\$140/220	\$16	\$8	\$8	\$222/302

^a \$3 (NFS Pilot) per person per day/\$5 per person per day (Field Test Control condition)/\$5 per person per day for the first three days and \$10 per person per day for the rest of the four days (Field Test Experimental Condition) for a completed daily report.

^b \$5 per household in NFS Pilot; \$2 per person in the Field Test

^c For larger households, add \$21 (NFS Pilot) or \$39 (Field Test Control Condition)/\$59 (Field Test Experimental Condition) for each additional person to the 4 person HH maximum

Completing the Food Log each day of the data collection week is also very critical for the success of FoodAPS. As a result, the Field Test will include one incentive experiment that varies the amount of incentive for Food Log completion to reduce the attrition of food log reporting during the week. Previous research demonstrates the effectiveness of incentives at increasing response rates (Singer, Van Hoewyk, Gebler, Raghunathan, and McGonagle 1999) and having positive effects on data quality, attrition, and item nonresponse (Bonke and Fallesen 2010; Martin, Abreau, and Winters 2001; Singer 2002; and Singer, Van Hoewyk and Maher 1998). Chromy and Horvitz (1978) found strong evidence that supports a variable incentive protocol. These studies provide motivation for our experiment, which will vary the amount of the incentive offered for completing each day’s food acquisition (as shown in Table B.6b).

The control condition of the experiment offers \$5 per person for completing each day’s Food Log, yielding a maximum of \$35 per person for seven-day reporting. By contrast, the experimental condition offers \$5 per person per day for the first three days of food log reporting. Starting on Day 4, the daily incentive amount is increased to \$10. The maximum amount of incentive for completing all seven day’s food acquisition is \$55 per person. This experiment tests whether doubling the amount of incentive at Day 4 would decrease the trend of underreporting food acquisitions during the latter half of the seven-day data collection period observed in FoodAPS-1 and NFS pilot.

Table B.6b. Incentive Experiment for the Field Test

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Max Total (7-day Food Log)
Control	\$5	\$5	\$5	\$5	\$5	\$5	\$5	\$35
Experimental	\$5	\$5	\$5	\$10	\$10	\$10	\$10	\$55

The purpose of the experiment is to examine which incentive schema is cost efficient and provides more complete 7-day Food Logs for FoodAPS. This will inform the Full Survey. The measures of interest will be at the person level. The minimum detectable difference for 80% power and a 0.05 level of significance is given in Table B.7. For example, if the percentage of 7-day Food Logs completed is 75.0 percent in the control group, with a person-level sample size of about 1,290 among 430 households (assuming three persons per household) adjusted for design effect, the sample size per group will be large enough to detect a completion rate difference of 7.2 percentage points, e.g., a treatment group completion percent of 82.2.

Table B.7. Minimum Detectable Differences for the Incentive Field Test, n=430 Households

Estimated Percentage	Minimum Detectable Difference for Person Measures
75	7.2
90	4.7

The current Field Test sample size of 430 completed households is sufficient to support the proposed experiment by itself, without comparing to other data.

Defining sampling rates/income prediction modeling: As noted earlier, we plan to improve the sampling rates for screener subsampling further by building a predictive model to estimate the final income domain based on screener and American Community Survey (ACS) information, as discussed in Section B.2. The sampling rates used in FoodAPS-1 and the NFS Pilot are based only on screener income subsampling stratum. However, other screener variables (e.g., number of household members) and ACS variables such as tract-level education or income percentages may allow us to estimate the household's true income category more precisely, and both reduce misclassification and weight variation.

Any Use of Periodic (Less Frequent Than Annual) Data Collection Cycles to Reduce Burden

Data collection for the Field Test will be conducted over a six-month period. The first couple months of data collection will focus on recruiting sampled households via a mail screener. Field data collection for the Field Test will be conducted over a four-month period. The use of periodic data collection cycles to reduce burden is not applicable for the Field Test as each eligible household will be invited for an initial in-person interview, recording food acquisitions over the course of one week, completion of a Profile Questionnaire and an Income Worksheet during that week, and then a final in-person debriefing interview at the end of the seven-day data collection week.

B.3 Methods to Maximize Response Rates and to Deal With Issues of Nonresponse

Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield “reliable” data that can be generalized to the universe studied.

Planned methods to maximize response rates in the Field Test and deal with issues of nonresponse are described below. As mentioned in Section B.1, the approaches planned for the Field Test are designed to achieve a slightly higher rate than in the NFS Pilot. However, they are balanced with continued evidence of response rates declines since 2017.

Maximizing Response Rates

Overall, we plan to maximize response rates in the Field Test by explaining the importance and potential usefulness of the study findings in the FoodAPS-2 Field Test introductory letter, by providing interviewers with the tools to gain the households’ cooperation, by explaining the incentive structure, and by implementing a series of follow-up reminders and targeted call-backs to under-reporters. We expect an in-person screener response rate of 36% (the same as the 36% for the NFS Pilot); an overall completion rate of 71% for the Initial Interview conditioning on being screened in as eligible; and a completion rate of 92% for the debriefing interview. The product of the response rates arrives at an overall response rate of 24%. The following procedures will be used to maximize response rates:

- Implement standardized training for field data collectors.
- Mail an introductory letter stating the importance of the study and their participation and the incentives they will receive upon full participation.

- Provide an unconditional incentive of \$5 for the mail screener and a conditional incentive of \$5 for the in-person screener.
- Make multiple visits to a sampled address without reaching someone before considering whether to treat the case as “unable to contact.”
- Make a refusal conversion attempt to convert households that refuse to participate after the interviewer’s initial visit to complete the Household in-person Screener.
- Provide a toll-free number for respondents to call to verify the study’s legitimacy or to ask other questions about the study.
- In accordance with the Confidential Information Protection and Statistical Efficiency Act ("CIPSEA") enacted in 2002, assure respondents that the information they provide will be kept confidential.
- Provide in-person interviewing for the Screener, Initial Interview, training, and debriefing interview.
- Provide optional YouTube videos and other resources to respondents that reinforce material presented during training on how to use the FoodLogger app.
- Offer telephone interviews as an alternative mode for recording FoodLogger information, for households identified as having limitations/barriers with respect to the technology needed to participate otherwise.
- Loan study equipment (barcode UPC scanner, laptop, iPhone, and/or MiFi device) to households determined by the field interviewer to need the equipment in order to complete the FoodLogger.
- Provide multiple reminder email messages, text updates, and/or interactive voice response (IVR) calls throughout the data collection week informing respondents of the incentives they have earned and motivating them to complete their FoodLogger.
- Provide incremental incentives based on respondents’ reporting behavior to encourage complete reporting of food acquisition.
- Streamline the FoodLogger to take full advantage of GPS and Google functionalities and to integrate various lookup databases.

- Allow access to the FoodLogger through computers or smartphones.
- Provide smartphone application of the FoodLogger that has the full functionality of passively tracking GPS locations, scanning barcodes of food items, and taking pictures of food items and receipts, increasing convenience for respondents.

Nonresponse Bias Analysis

The magnitude of nonresponse bias in a survey estimate depends on the response rate and the extent to which the respondents and nonrespondents differ on the outcome of interest. In a nonresponse bias analysis conducted on the FoodAPS-1 sample by Petraglia, et al. (2016), a 42 percent response rate was reported, and the respondents differed significantly from nonrespondents on several socio-economic characteristics. A main difference was that higher response rates were found to be associated with SNAP participation and lower income. However, these differences were largely reduced through the weighting process. In addition, the weighting variables were correlated with food insecurity, total amount spent on food-at home (FAH) events, total amount spent on food-away-from-home (FAFH) events, and number of free events, suggesting that the weighting adjustments should also have reduced bias in these outcome estimates. Overall, the analysis did not indicate that nonresponse bias is a concern, although the extent of bias remaining after weighting adjustments is unknown.

There has been a significant downward trend in response rates which confounds direct comparison of the Field Test rate to the FoodAPS-1 rates. Thus, we will use both FoodAPS-1 and NFS Pilot response rates as reference points to compare Field Test response rates and caveat the findings accordingly. We will compare response rates from the point of screening to the completion of the debriefing interview. We will estimate and compare response rates for the in-person Screener, Initial Interview, Food Log (overall and by day), and debriefing interview to the response rates from FoodAPS-1. For the Field Test, we will conduct a nonresponse bias analysis, such as computing response rates by subgroups to determine if there are any indications of nonresponse bias. We will use area characteristics for screener

response rates, and in addition household characteristics collected in the screener for interview response rates conditional on completed screeners. An objective of the Field Test is to be able to use that experience to estimate a likely response rate for a full rollout of the data collection. The regression model will provide insights into the household characteristics that are associated with response. The nonresponse patterns revealed from the Field Test can help to anticipate nonresponse patterns in the Full Survey when assigning sampling rates to achieve sample size goals by domain. Nonresponse follow-up strategies could also be improved. Paradata collected during the Field Test on number of contacts needed until completion and the association with survey outcomes can be used to adjust contact protocols. For example, would reducing the standard number of contacts change survey estimates, or would increasing the number of contacts reduce bias due to nonresponse? The response propensity model that informs responsive design could include new covariates that are identified from the Field Test. In addition, the nonresponse adjustments in the weighting process can be informed by observing associations between auxiliary data (e.g., screener and tract-level data from the ACS) and survey outcomes, as well as response indicators.

B.4 Test of Procedures or Methods to be Undertaken

Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of tests may be submitted for approval separately or in combination with the main collection of information.

The development of FoodLogger is a result of collaborative effort between Westat, ERS, Census and NASS. Weekly meetings have been held to discuss developments of the concepts and the functionalities of the FoodLogger. The

Census Bureau is independently evaluating the functionality and usability of the smartphone application by conducting two rounds of formative usability testing. The purposes of the usability testing are to test the functions (what can the app do?) and the usability (how easy is it to use the app?) of the app. Said differently, the usability testing will assess the user experience and ensure respondents enter food acquisition data effectively, and with satisfaction.

The first round of usability testing was completed on May 24, 2021 and the key findings and recommendations for improvements and modifications to the FoodLogger were presented by Census to ERS and Westat on June 17, 2021. The final analysis report from the first round, which includes the full set of recommendations, is included as Attachment U. The second round of usability testing was completed on July 23, 2021 and the key findings and recommendations for improvements and modifications to the FoodLogger were presented by Census to ERS and Westat on August 2, 2021. The final analysis report from the second round, which includes the full set of recommendations, is included as Attachment X.

A consistent finding across each round of usability testing was that a majority of the participants rated each of the critical tasks as “Extremely Easy” or “Easy” to perform in the FoodLogger (see Table 5 of Attachment U and X for a summary of the self-reported easiness of task performance for the first and second rounds, respectively). This provides overwhelming evidence that the FoodLogger is usable and is easy to use when entering food acquisition data. That said, Westat has made modifications to the FoodLogger based on results of each round of usability testing.

During the first round of usability testing, the following usability issues were designated as a high priority: (1) making all options for reporting (e.g., barcode scan, PLU entry, and type food description for both individual food items and “combo meals”) available for all “food event” types; (2) improve respondent instructions and data entry process for barcode scanning and PLU entry to mitigate broken workflows; and (3) align response options for food item weight, volume, size, and package questions to correspond with

appropriate measurements and measurement units and ensure they are mutually exclusive.

Modifications to the Food Log questionnaire and the FoodLogger to address items (2) and (3) have been implemented and are currently being evaluated via a second and final round of usability testing. These modifications include improved respondent training on data entry procedures, changing the color and font size of the error messages a respondent receives when a barcode is incorrectly scanned or a PLU code is not found in one of the linked food item databases. Improved decision rules and database querying instructions were programmed into the FoodLogger so when a barcode, PLU code, or food item matches a linked food item database, the questions on package, weight, volume, and size are skipped. This is because those food item details are contained in the linked databases. This has the added benefit of reducing respondent burden and improving data quality. In instances where food items are not found in one of the linked databases, “other” response options were added to questions on packaging, weight, volume, and size questions. It is also worth noting that part of the purpose of the Field Test is to collect data on a larger and representative sample of the target population and ERS will use the findings from the Field Test to make additional improvements to these questions and their sequencing within the Food Log.

With respect to issue (1), while there are certainly acquisition scenarios for which respondents may acquire an item typically classified as FAFH (e.g., a prepared or ready-to-eat meal) at a food place typically classified as a FAH establishment (e.g., grocery store), and vice-versa, ERS believes that the decision to implement this significant structural change to the programming of the FoodLogger will be best informed by the data collected via the Field Test. The current structure was informed by and designed to balance respondent burden and data quality. The current structure intends to minimize the number of paths a respondent can follow within the FoodLogger thereby mitigating the potential occurrence of measurement errors. If the findings from the Field Test suggest that the prevalence of these types of acquisition scenarios is high, then such a structural change to the FoodLogger will be made prior to conducting the Full Survey.

During the second round of usability testing the following issues, distinct from those identified during the first round, were designated as major findings or high priority usability issues: (1) older respondents struggled the most with respect to the technology, including FoodLogger installation; (2) barcode scanning and the questionnaire sequencing from unmatched barcodes; (3) type-ahead feature may be burdensome when trying to find an exact match to the name of the food item to be reported; (4) combo meal entry; and, (5) uploading a receipt.

With respect to issue (1) it is important to note that one of the limitations of the usability testing is that it was conducted virtually via MS Teams because of the COVID-19 Pandemic. The data collection procedures for the Field Test include an in-person training on how to use the FoodLogger via a series of comprehensive videos of survey concepts and procedures (Attachments V1-V6) and an assessment by the interviewer as to whether the respondent should proceed with the survey using the FoodLogger or, if they struggle with the technology, another collection mode. Both the in-person training and the assessment by the interviewer will mitigate the occurrence of similar technological issues in the Field Test. In addition, although not part of the required in-person training two optional videos (Attachments V7 and V8) providing additional training on concepts relating to grocery store and fast food restaurant purchases will be made available to all respondents to view at their discretion or as needed.

To address issue (2) error messages for erroneous or unmatched scans will be displayed in the FoodLogger in red font. This will be a different color from all the other text on the screen making an unsuccessful scan more salient and prominent to the user. It will also be stressed in training that not all barcodes will be identified by the FoodLogger or the barcode scanner (Attachment V3). The implications of a successful or unsuccessful scan will also be reiterated in training. If the scan is successful, the respondent will have fewer follow-up questions to answer about the food item. If the scan is unsuccessful, the respondent will need to type a name or description of the food item and answer follow-up questions about the item.

For item (3), the recommendation was to populate the type-ahead list with only generic food item names and not brand names where appropriate. This recommendation is contrary to the level of detail that FoodAPS requires for nutrition coding and other critical calculations, like the Healthy-Eating-Index. Nutrition information for a food item is tied very closely to specific products. ERS believes that keeping this level of detail in the procedures for the Field Test will provide an opportunity to reevaluate this reporting requirement and make empirically supported revisions to the requirements for the Full Survey as appropriate.

To resolve the combo meal entry usability issue identified in the second round, the recommendation was to reconcile the data entry order and ask the respondent to enter all individual food items in a combo meal before entering the cost. This recommendation goes against concept of a “combo meal” which is a group of separate food items offered together for one price. The in-person training for the Field Test (Attachment V4) will stress the concept of a combo meal and instruct the respondent on how to appropriately enter these details within the FoodLogger.

Finally, the second round found that only one of the five participants were able to upload a receipt without major issues. The recommendation was to provide detailed training on uploading an electronic receipt. This is exactly the procedure to be executed in the Field Test. The training will stress how to upload a receipt via the available methods (Attachment V4). Respondents will also be encouraged to save hard copies of all receipts and Westat will have a help-desk call-in number and project email that the respondents can utilize if they have difficulty on or questions about certain tasks, procedures, and survey concepts.

An experiment is built in the FoodLogger that asks respondents to take a picture of a food/drink item on a random sample of 30% of FAFH events from food places that are not recognized in an external database. This experiment allows us to achieve two objectives: 1) to gain a better understanding of the burden associated with taking pictures of food items acquired and 2) to assess to what extent, if any, those pictures can be used for food item identification and nutrition coding.

B.5 Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data

Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.

The sampling plans were reviewed by Wendy Van de Kerckhove, Elizabeth Petraglia, Ting Yan, Jill DeMatteis, Tom Krenzke, Erika Bonilla, David Cantor, Janice Machado, and Laurie May at Westat. In addition, Jeffrey Gonzalez, Linda Kantor and Elina Page of ERS, Brady West from the University of Michigan Survey Research Center (U-M SRC), Lin Wang from the Human Factors Research Group, Center for Behavioral Science Methods at the U.S. Census Bureau, Joseph Rodhouse and Darcy Miller from the National Agricultural Statistics Service (NASS), and John Kirlin of Kirlin Analytic Services have reviewed components of this supporting statement. This supporting statement was revised per comments from the Westat team, ERS, NASS, and U-M SRC. All data collection and analysis will be conducted by Westat.

Name	Affiliation	Telephone number	E-mail
Jeffrey Gonzalez	Project Officer, ERS	202-694-5398	jeffrey.gonzalez@usda.gov
Brady West	Research Associate Professor, University of Michigan Survey Research Center	734-647-4615	Bwest@umich.edu
Laurie May	Corporate Officer, Westat	301-517-4076	LaurieMay@westat.com
Janice Machado	Project Director, Westat	301-294-2801	JaniceMachado@westat.com
David Cantor	Vice President, Westat	301-294-2080	DavidCantor@westat.com
Erika Bonilla	Study Manager,	301-610-	ErikaBonilla@westat.com

Name	Affiliation	Telephone number	E-mail
	Westat	4879	
Tom Krenzke	Vice President, Westat	301-251-4203	TomKrenzke@westat.com
Jill DeMatteis	Vice President, Westat	301-517-4046	JillDematteis@westat.com
Ting Yan	Senior Statistician, Westat	301-250-3302	TingYan@westat.com
Wendy Van de Kerckhove	Senior Statistician, Westat	240-453-2785	WendyVandeKerckhove@westat.com
Elizabeth Petraglia	Senior Statistician, Westat	240-314-7535	ElizabethPetraglia@westat.com
Linda Kantor	Economist, ERS	202-694-5392	Linda.Kantor@usda.gov
Elina T. Page	Economist, ERS	202-694-5032	Elina.t.page@usda.gov
Joseph Rodhouse	Mathematical Statistician, NASS	202-692-0289	Joseph.Rodhouse@usda.gov
Darcy Miller	Mathematical Statistician, NASS	202-690-2652	Darcy.Miller@usda.gov
Lin Wang	Research Psychologist, U.S. Census Bureau	301-763-9069	Lin.wang@census.gov
John Kirlin	President, Kirlin Analytic Services	540-786-1042	johnkir516@hotmail.com

References

- Bonke, J., & Fallesen, P. (2010). The Impact of Incentives and Interview Methods on Response Quantity and Quality in Diary- and Booklet-Based Surveys. *Survey Research Methods*, Vol. 4 No. 2, 91-101.
- Bowman, S.A., Clemens, J.C., Shimizu, M., Friday, J.E., and Moshfegh, A.J. (2018). Food Patterns Equivalents Database 2015-2016: Methodology and User Guide [Online]. Food Surveys Research Group, Beltsville Human Nutrition Research Center, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, Maryland. September 2018. Available at: <http://www.ars.usda.gov/nea/bhnrc/fsrg>
- Chromy, J. (1979). Sequential Sampling Selection Methods. Proceedings of the Survey Research Methods Section: American Statistical Association, Vol. 2.
- Chromy, J. R., & Horvitz, D. G. (1978). The use of monetary incentives in national assessment household surveys. *Journal of the American Statistical Association*, 73(363), 473-478.
- Kalton, G. (2009). Methods for oversampling rare subpopulations in social surveys. *Survey Methodology*, 35(2), pp. 145-141.
- Kalton, G. and Anderson, D.W. (1986). Sampling rare populations. *Journal of the Royal Statistical Society, Series A*, 149, pp. 65-82.
- Kalton, G., Kali, J. and Sigman, R. (2014), Handling Frame Problems When Address-Based Sampling is Used for In-Person Household Surveys, *Journal of Survey Statistics and Methodology*, 2, 283-304.
- Krenzke, T., and Kali, J. (2016). *Review of the FoodAPS 2012 Sample Design*. Written for the Economic Research Service, U.S. Department of Agriculture. Washington, D.C.
- Link, M.W. (2010, November). Address Based Sampling: What Do We Know So Far? Webinar, American Statistical Association.
- Mercer, A., Caporaso, A., Cantor, D. and Townsend, R. (2015). How Much Gets You How Much? Monetary Incentives and Response Rates in Household Surveys. *Public Opinion Quarterly* 79 (1):105-29.
- Petraglia, E., Van de Kerckhove, W, and Krenzke, T. (2016). *Review of the Potential for Nonresponse Bias in FoodAPS 2012*. Prepared for the

Economic Research Service, U.S. Department of Agriculture.
Washington, D.C., accessible at
<https://www.ers.usda.gov/media/9070/potentialnonresponsebias.pdf>.

Martin, E., Abreu, D., Winters, F. (2001) Money and Motive: Effects of Incentives on Panel Attritions in the Survey of Income and Program Participation. *Journal of Official Statistics*, Vol. 17 No. 2, 267-284

Singer, E., Van Hoewyk, J. and Maher, M. (1998). "Does the Payment of Incentives Create Expectation Effects?" *Public Opinion Quarterly* 62:152-64.

Singer, E. Van Hoewyk, J., Gebler, N., Raghunathan, T., and McGonagle, K. (1999). The Effect of Incentives on Response Rates in Interviewer-Mediated Surveys. *Journal of Official Statistics*, Vol.15, No. 2, pp. 217-230

Singer, E. (2002). *The Use of Incentives to Reduce Nonresponse in Household Surveys*. in *Survey Nonresponse*, eds. Robert M. Groves, Dan A. Dillman, John L, Eltinge, and Roderick J. A. Little, pp. 163-77.