

National Household Food Acquisition and Purchase Survey (FoodAPS)

User's Guide to Survey Design, Data Collection, and Overview of Datasets

**October 29, 2014
Last Revised: June 20, 2016**

The OMB clearance number for FoodAPS is 0536-0068. The data were collected by the U.S. Department of Agriculture under authority of U.S.C, Title 7, Section 2026 (a)(1).

Further information about the entire data collection, including instructions on how to request access to the data, may be found at <http://www.ers.usda.gov/foodaps>.

For further information contact: FoodAPS@ers.usda.gov

Table of Contents

Chapter 1: Introduction	1
Chapter 2: Sample Design	2
2.1: Household Definition and Membership	2
2.2: Target Groups	3
2.3: First Stage - Primary Sampling Units (PSUs)	3
2.4: Second Stage - Secondary Sampling Units (SSUs)	4
2.5: Third Stage – Addresses.....	5
2.5.1 <i>Sampling frames</i>	5
2.5.2 <i>Sampling addresses</i>	6
Chapter 3: Data Collection Process	7
3.1: Data Collection Protocols and Instruments.....	7
3.2: The Data Collection Week	8
3.3: Incentives	9
3.4: Food at home (FAH) and Food away from home (FAFH)	10
3.5: Survey Response Rates	11
Chapter 4: FoodAPS Data Sets	11
Chapter 5: SNAP Administrative Data Matching.....	14
Chapter 6: Analytic Notes	15
6.1: Accounting for the complex sampling design.....	15
6.1.1 <i>Construction of sampling weights</i>	16
6.1.2 <i>Applying weights and calculating variances</i>	18
6.2: Factors Affecting Data Quality	19
6.2.1 <i>Underreporting of Food Acquisitions</i>	19
6.2.2 <i>Observational Effects</i>	21
6.3: Non-Response Bias	22
Appendix A – List of Data Files and Documentation	25
Appendix B– Merging Data files	28
Appendix C– Examples of Variance Estimation.....	30
Appendix D– Further Details on Sampling and Weighting	38
Appendix E– Summary of Revisions.....	42

Chapter 1: Introduction

The National Household Food Acquisition and Purchase Survey (FoodAPS) collected information on foods purchased or otherwise acquired and the prices and nutrient characteristics of those foods for a nationally-representative sample of U.S. households. Data on factors expected to affect food acquisition decisions, such as household size and composition, demographic characteristics, income, participation in Federal food assistance programs and dietary restrictions were also collected.

The FoodAPS data collection was sponsored by the U.S. Department of Agriculture (USDA) and managed by USDA's Economic Research Service (ERS) with support from USDA's Food and Nutrition Service (FNS). Due to special interest in the food acquisition patterns of households participating in these programs, the survey oversampled low-income households, both those receiving SNAP benefits and those not receiving SNAP benefits. The survey is weighted to be representative of all non-institutionalized households in the continental United States.

The data collection instruments and protocols for FoodAPS were designed by Mathematica Policy Research (Mathematica) under contract to ERS. Mathematica fielded the survey from April 2012 through mid-January 2013. Each sampled household participating in the survey was asked to provide information on the foods all members acquired over a seven-day period. The household's primary respondent also participated in two in-person interviews and up to three telephone interviews. Data were collected from 4,826 households.

To increase USDA's and users' full understanding of the FoodAPS sample design, data collection protocols, sample weights, and data quality, ERS contracted with Westat, Inc. to conduct an independent review of the entire FoodAPS survey. The User's Guide references some of this work and its results. Reports from the review will be posted to the FoodAPS section of ERS's website when they become available.

FoodAPS was conducted under the protections and regulations of the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA). The protections

accrue to respondents because CIPSEA requires that the collected data be used only for statistical purposes (e.g., not turned over for use by program administrators or legal authorities), and it promises high levels of data protection against disclosure of identifying information. To help ensure that these protections are taken seriously and followed, penalties for a violation of CIPSEA procedures can result in a fine of up to \$250,000 and/or five years in prison.

At this time, FoodAPS data are not being released publicly due to concerns about disclosure risks and protecting respondent confidentiality. Researchers requesting access to FoodAPS data files will be required to sign a CIPSEA pledge of confidentiality and be subject to the Act's penalties. They must also agree that they will not disseminate any research results based on FoodAPS data until ERS has reviewed the results for risk of disclosure of confidential information and then authorized their release. Instructions for requesting access to the restricted data and conditions on use are available in the FoodAPS section of ERS's website (<http://www.ers.usda.gov/foodaps>).

The OMB clearance number for FoodAPS is 0536-0068. The data were collected by USDA under authority of U.S.C, Title 7, Section 2026 (a)(1).

Source documentation for this *User's Guide* includes technical memoranda provided to ERS by Mathematica and Westat.

Chapter 2: Sample Design

2.1: Household Definition and Membership

Because of its focus on food acquisitions and the role of food and nutrition assistance programs like SNAP on household food demand, the concept of a FoodAPS "household" is generally broader than used in most general surveys.

The FoodAPS household is defined as all persons who live together and share food and who expect to be present at the sampled address during at least part of the data collection week.

Plans to travel during the week did not preclude a household member from participating, as they were asked to fill out their food books and keep food receipts while on the road. Family members not at home during the data collection week (e.g., students away at school, military personnel stationed elsewhere) were not included as household members. The goal was to match food acquisitions as closely as possible with the people at the sampled residence during the week.

2.2: Target Groups

The survey had four target groups, defined in terms of participation in the Supplemental Nutrition Assistance Program (SNAP) and total reported household income:

- Households receiving SNAP benefits, with a target sample size of 1,500;
- Non-SNAP households with income less than the poverty guideline, with a target sample size of 800;
- Non-SNAP households with income greater than or equal to 100 percent and less than 185 percent the poverty guideline, with a target sample size of 1,200; and
- Non-SNAP households with income greater or equal to 185 percent of the poverty guideline, with a target sample size of 1,500.

The sample of households was selected through a multi-stage sample design with coverage in the contiguous United States.

2.3: First Stage - Primary Sampling Units (PSUs)

Before sampling, PSUs were defined as counties or groups of contiguous counties. In forming PSUs, metropolitan statistical area (MSA) boundaries, defined by the Office of Management and Budget, were used (some MSAs were split into multiple PSUs, but in no case was part of one MSA joined to part of another MSA to form a PSU). The number of PSUs formed was 948.

After the PSUs were formed, a stratified sample of 50 PSUs was selected using probability proportional to size (PPS) selection. The measure of size (MOS) for each PSU was a composite of four estimates derived from the 2006-2008 American Community Survey (ACS) Public Use Microdata Sample (PUMS) files: the number of SNAP households in the PSU; and the number of non-SNAP households from the three income groups: below 100 percent of the poverty guideline; between 100 and 184 percent of the poverty guideline, and equal to or greater than 185 percent of the poverty guideline. The composite measure reflected the number of households in each of the four target groups within each PSU and the relative overall sampling rate of households within the PSU (details available upon request).

The FoodAPS sample includes two strata: one large PSU sampled with certainty (i.e., guaranteed to be in the sample of PSUs) and the remaining 49 non-certainty PSUs. The non-certainty PSUs were selected using PPS with implicit stratification based on the metropolitan status of the PSU and its FNS regions. Metropolitan status was either “metro” (the entire PSU was within one or more MSAs,); “non-metro” (not in any MSA); or “mixed” (included counties in an MSA and counties not in an MSA). FNS divides the 50 States, the District of Columbia, and U.S. territories into seven administrative regions. The 50 selected PSUs include 34 metro PSUs, 10 non-metro PSUs, and 6 mixed PSUs. The number of selected PSUs in each FNS region varies from 4 to 11.

2.4: Second Stage - Secondary Sampling Units (SSUs)

Within each of the 50 sampled PSUs, eight secondary sampling units (SSUs) were selected using PPS. Each SSU comprised a Census Block Group (CBG), or a group of contiguous block groups if the CBG was expected to contain fewer than 50 survey-eligible households (Only four of the 400 SSUs contained more than one CBG). The MOS for selecting the SSUs was a composite MOS constructed in the same way as the MOS for PSUs (details available upon request). 205 of 400 selected SSUs were located within the boundaries of incorporated cities; with the remainder in towns, villages, boroughs, and unincorporated areas.

2.5: Third Stage – Addresses

2.5.1 Sampling frames

To more efficiently sample SNAP and non-SNAP households, a sampling frame for each SSU was constructed from a commercial list of addresses, known as an Address-Based Sampling (ABS) list, paired with a list of SNAP addresses. The ABS list was obtained from the United States Postal Service Delivery Sequence File. The SNAP address list was obtained from 22 of the 27 States with selected PSUs.¹ This SNAP list contained addresses for households receiving SNAP benefits in February 2012. SNAP household addresses were matched to the ABS list. Any addresses that matched were identified as SNAP addresses and any addresses in the ABS list that did not match the SNAP list were identified as non-SNAP addresses. Any addresses that appeared in the SNAP list, but not in the ABS list, were retained in the final sampling frame and identified as SNAP addresses. Each address appeared only once on the final composite frame. This approach was used for 315 SSUs.

For the remaining 85 SSUs, it was not possible to construct a composite sampling frame where SNAP and non-SNAP addresses were differentiated. Not all States provided the SNAP address list, and in some SSUs, the ABS list contained a large number of addresses that were not useful for locating households (e.g., P.O. boxes, Rural Delivery). For these remaining SSUs, the sampling frame was constructed using one of the following methods:

- **Field listing.** In the 14 SSUs where the ABS list a large number of addresses that were not useful for locating households (e.g., P.O. boxes, Rural Delivery), addresses in the sampled SSU were listed by Mathematica field staff. The listed addresses comprised the sampling frame in that SSU.²

¹ The SNAP agencies in the 27 States where the selected PSUs were located were asked to provide a list of addresses for all SNAP recipients in the selected PSUs. Four States were unable to provide a list of SNAP addresses, and one State provided the list too late to be included in the sampling process.

² Even though SNAP addresses were obtained for these 14 field-listed SSUs, it was not possible to match the SNAP address list due to differences in how the addresses were listed during field listing.

- **ABS list only.** For the 71 SSUs in the five States where agencies were unable to provide addresses of SNAP recipients in the PSUs, the ABS list was the sampling frame.

The SNAP Address list was expected to become less useful over time in finding SNAP households as they moved or left the program. Thus, the screener used a direct question about SNAP participation to help determine a household's target group.

Five of the 400 SSUs selected had no households that were screened, found eligible for the survey, and willing to participate in the survey. Thus, the final sample of 4,826 households comes from 395 sampled SSUs.

2.5.2 Sampling addresses

In order to meet the target sample sizes for the target groups in Section 2.2, Mathematica randomly divided the initial sample for both the SNAP frame and the non-SNAP frame into 70 equal size replicates. The sample was then supplemented with 41 replicates from the non-SNAP frame, each of the same expected size as the initial replicates in the non-SNAP frame. Decisions were made about which replicates to release at the SSU level, separately for the SNAP and non-SNAP sample. Because some target groups were easily located, particularly non-SNAP households above 185 percent of the poverty guideline, the replicates in SSUs with high proportions in these groups were selected at a lower rate. Consequently, attention was focused on reaching the target goals for target groups that were harder to locate. Only 20,084 addresses out of the total sample of 42,143 addresses were released to the field.

FoodAPS used a two-phase sampling approach for conducting the screener interview as a way to reduce the potential of non-response bias. The first phase included all 20,084 addresses released to the field to be screened for the study. If a sampled address had a dwelling unit that appeared to be occupied, field interviewers made at least eight attempts at different times of the day and different days of the week to contact the residents. If no contact was made, the address was classified as "effort ended" and removed from the interviewer's Phase 1 list.

In mid-October 2012, 985 addresses that were classified as “effort ended” formed the sample frame for Phase 2. From the Phase 2 frame, 138 cases (14 percent) were randomly selected and re-released to the field for up to 10 additional contact attempts. The effort resulted in 12 completed cases that were added to the 4,814 completed cases from Phase 1.

Chapter 3: Data Collection Process

3.1: Data Collection Protocols and Instruments

A screening interview determined whether the household at a sampled residence was eligible to participate in FoodAPS based on two criteria. First, the screening interview determined whether the sampled housing unit was in scope for the survey. The sampled address had to be the usual residence for the household and not, for example, a vacation home. Also, the sampled housing unit could not be an institutional group quarters. Second, information collected on the screener—household composition, current participation in SNAP, and a preliminary report of monthly income—was used to classify the household into one of four target groups. Because the most-difficult-to-locate households were those with income less than the poverty guideline but not receiving SNAP, this group was always classified as eligible for the study. Other households were subsampled at differing rates. (Section 2.2 describes the four target groups.) For example, only roughly 1 in 4 households not on SNAP and with income above 185 percent of the poverty guideline was classified as eligible to complete the study.

If eligible, the screener identified the main food shopper or meal planner in the household and invited him or her to participate in the week-long data collection. If the person agreed to participate, he or she became the household’s primary respondent (PR).

The PR was asked to complete two in-person interviews and to call the study’s telephone center for three brief telephone interviews regarding food acquisition events over the course of one week. Each household member age 11 years and older was

asked to track and report all food acquisitions during the one-week period. The PR and other Adult food books contained Daily List pages, which documented summary information for each acquisition for each of the seven days of food reporting. The PR food book included both Blue pages to report details for “food at home” and Red pages to report “food away from home” acquisitions. Youth books, for children 11 to 17 years old, included only Red pages to report food-away-from-home acquisitions.³ The PR was responsible for recording food acquisitions by members under 11 years old.

Households were asked to scan barcodes on foods, save their receipts from stores and restaurants, and write information in their food books. For food-at-home acquisitions, the scanned barcodes were intended to be the primary source of item-level descriptions, while the receipts were intended to provide the price or expenditure information for each item. The Food Book (Blue) pages would provide the rest of the information and saved receipts would be used to verify this information and/or fill-in missing information from the Blue page. For food-away-from-home acquisitions, the phone calls were intended to be the main source of item-descriptions, details about the event, and price/expenditure information. The Red pages were reviewed to identify and capture any information that had not been reported during a phone call.

Copies of all data collection instruments are posted on the FoodAPS section of ERS’s website at <http://www.ers.usda.gov/foodaps>.

Instruments were available in English, Spanish, Vietnamese, and Korean, and interviews were conducted in these languages depending on respondent preference.

3.2: The Data Collection Week

A typical survey week for a participating household included the following:

- Initial Household Visit

³ The terms “food at home” and “food away from home” are defined in Section 3.4.

- Completed *Household Screener*
- PR provided consent and completed the CAPI-based⁴ *Initial Household Interview*
- PR received training on use of food books and scanner; PR was responsible for training other household members, as needed
- Each day from Day 1 to Day 7, household members or the PR were asked to
 - Record food acquisitions in their food books⁵
 - Fill out the Meals and Snacks form to indicate which meals and snacks each household member ate each day
- Each member age 16 and older was asked to fill out the *Income Worksheet*
- Days 2, 5, 7
 - PR called the Survey Operations Center (SOC) to report food acquisitions
- Final Household Visit (after end of week)
 - PR completed CAPI-based *Final Household Interview*
 - Interviewer reviewed completeness of food books, issued base incentive check (\$100) and gift cards, and collected all food books and the scanner
 - PR was asked to complete a *Respondent Feedback Form*

3.3: Incentives

FoodAPS offered a \$5 unconditional incentive to all households contacted during the screening interviews. Once found eligible to participate in the study, respondents were informed of additional incentives to encourage both initial agreement to participate in the study and continued participation throughout the food reporting week. The following incentives were provided at the end of the data collection week.

- The primary respondent received a \$100 check (the base incentive).

⁴ CAPI stands for Computer-Aided Personal Interviewing.

⁵ FoodAPS used three different food books. The PR used the *Primary Respondent Book*; other adults in the household each used an *Adult Food Book*; and children ages 11-17 each used a *Youth Food Book*. The PR was instructed to record foods acquired by all children under 11 years old in their *Primary Respondent Book*.

- The primary respondent received a \$10 gift card for each of the three scheduled telephone calls to the SOC that were initiated by the primary respondent.
- Children age 11-14 years received a \$10 gift card if they filled out their Youth Food Books or otherwise reported food acquisitions to the primary respondent.
- Members 15 and older received a \$20 gift card if they filled out their Youth or Adult Food Book or otherwise reported food acquisitions to the primary respondent.

3.4: Food at home (FAH) and Food away from home (FAFH)

When filling out their food books, participants were asked to distinguish between "food and drinks brought into the home" and "meals, snacks, and drinks you got outside the home," with the former being recorded on Blue pages within the Primary and Adult food books and the latter being recorded on Red pages within each member's food book. The intention was that the FAH pages would collect information about foods and drinks that are brought home and used to prepare meals for consumption at home or elsewhere (for example, food used to make a sandwich that you bring to work) and that FAFH pages would collect information about foods and drinks that are obtained and consumed away from home, and prepared foods that are brought home or delivered (e.g., pizza).

However, such a clean distinction was not always made by the respondents and foods of both types appear in the FAH and FAFH data sets. Users may wish to use the PLACETYPE and PLACENAME variables as well as item descriptions to help characterize the acquisition event.

It is also important to note that although the terms "food at home" and "food away from home" and the acronyms FAH and FAFH are used throughout this and other FoodAPS codebooks, they were not used on the data collection instruments or during interactions with respondents.

3.5: Survey Response Rates

Based on a weighted average of responses in Phases 1 and 2, the FoodAPS screener response rate was 72.2 percent, and the overall study response rate (OSRR) was 45.6 percent. The OSRR is equivalent to the RR3 response rate of the American Association for Public Opinion Research. More details on response rates are available upon request.

Chapter 4: FoodAPS Data Sets

The data collected by FoodAPS is provided in a number of different data files. Appendix A provides a list of the data and documentation files, along with their release date. Table B1 summarizes how to merge the data files, and Figure B1 shows the overall structure of the FoodAPS data sets.

Household and individual characteristics are provided in two data files: a household-level and an individual-level file based on the two interview survey instruments (the initial and final interview). The household data file **faps_household** includes 4,826 records. The person-level variables from the initial and final household interviews are contained in data file **faps_individual**. This file has 14,317 records and may be linked to **faps_household** using HHNUM.

Details on acquisitions are split into event-level files and item-level files for both FAH and FAFH acquisitions. The event-level files have one record per acquisition and characterize the acquisition place, participants, expenditure, and payment type. Each of the 15,998 records in the file **faps_fahevent** represents one food-at-home (FAH) acquisition. For the file **faps_fafhevent**, each of the 39,120 records represents a food-away-from-home (FAFH) acquisition recorded on a Red page in an individual's food book or described to the interviewer when the primary respondent made one of three calls during the reporting week to the Survey Operations Center. These event files may be linked to **faps_household** using HHNUM and to **faps_individual** using the variables HHNUM and the person identifier (PNUM in **faps_individual** and WHOGOTPNUM or BOOKPNUM in **faps_fahevent** and **faps_fafhevent**).

FAH and FAFH item-level information, including details on the type of foods, size and amount, and expenditure, are provided in the **faps_fahitem** and **faps_fafhitem** files. It is important to note that each record in the item files does not necessarily represent a unique product. Because the receipt sometimes guided how items were entered into the database, the purchase of multiple units of the same item, such as two boxes of a specific cereal, may appear in the data two different ways. If the barcodes were scanned or the receipt recorded the purchase on two separate lines (one for each box), the **faps_fahitem** data will include two line records, one for each box of cereal. However, if the barcodes were not scanned and the receipt recorded the two boxes on only one line, the **faps_fahitem** data will include only one record for the two boxes. FAFH items are recorded as reported by respondents and so, just as with FAH items, each record does not necessarily represent a unique food item. Items can be linked to the event record using EVENTID, which is unique across all FAH and FAFH events across all households.

Some barcodes in the FoodAPS data were matched to a proprietary data set created by Information Resources, Inc. (IRI)—a market research company—to obtain item descriptions and other information. An additional FAH item-level data file containing IRI categorization variables (such as Department, Aisle, Category, and Type) is available to researchers who work on a USDA project and are thus able to sign a Third Party Agreement with IRI to use the IRI-level data.

Information on the nutrient characteristics of each acquired FAH and FAFH food item is contained in separate FAH and FAFH nutrient files. The nutrient file contains the macro and micro nutrients and the Food Pattern Equivalents per 100 grams of each identifiable food. The FAH nutrients are in the **faps_fahnutrient** data file, and the FAFH nutrients are in the **faps_fafhnutrients** data file. An overview of the nutrient coding process is provided in *FoodAPS Nutrient Coding Overview*.

The file **faps_screeners** contains data from 13,445 completed screening interviews. Data from this file may be linked to household-level data from the initial and final

household interviews using the common variable HHNUM, which is unique across all FoodAPS households.

All weight variables (main weights and replicate weights) and variables necessary for Taylor series and jackknife variance estimation are provided in the **faps_hhweights** data file. The main household weights and strata and pseudo-psu identifiers necessary for Taylor series variance estimation are attached to both the **faps_household** and **faps_individual** data files. The strata and pseudo-psu identifiers should be used for variance estimation. Users can also utilize the replicate weights that have been constructed for variance estimation. See section 6.1.2 for more details about how to apply the weights in estimation and how to estimate variances.

The **faps_meals** data file contains individual-level data from the Meals & Snacks form that household members filled out during the study week.

The **faps_places** data file contains one record for every unique geocoded place that appears in the **faps_household**, **faps_fahevent**, or **faps_fafhevent** data files.

The **faps_access** data file provides distance and count estimates of the food retailers and restaurants in the areas surrounding each household.

The **faps_snapelig_hh** data files provide estimates of each household's eligibility for SNAP under four different estimation situations and assumptions about the number of SNAP units in each household. The **faps_snapelig_unit** and **faps_snapelig_ind** data files provide additional information about the eligibility estimations.

Geography Component (GC) data includes two main components: information about the food retail environment and related policy data, and the store-level food basket costs data. The food environment and policy data can be linked to FoodAPS households by Census block group, tract, or county. The food basket costs data contain weekly store-level estimated prices for 29 food groups corresponding to the Thrifty Food Plan (TFP) for each store in the IRI store located in the FoodAPS sampled PSUs and neighboring counties. The documentation for both of these components are provided on the FoodAPS section of ERS's website: <http://www.ers.usda.gov/foodaps>, with

supplementary documentation available at NORC. Users who wish to gain access to the GC data can do so when requesting access to the FoodAPS data at NORC.

The data file **faps_hhgeodata** contains information to link Geography Component and other geographically-based data to other FoodAPS files by the common variable HHNUM or by the Census FIPS identifier. Access to the Geography Component and **faps_hhgeodata** files is restricted to only those projects with approval to link such data.

ERS has also developed a new identification variable to allow linkages across multiple directories of food retail locations. A master directory of food retailers from TDLinx, STARS (SNAP-authorized stores), and IRI was created and merged with data on the places visited or the stores identified as usual shopping locations by FoodAPS respondents. In this master directory, a unique identifier for each food retailer was created and is called Temp_ERS_ID. Additionally, files that link the Temp_ERS_ID to each of the source datasets are available to researchers with the required permissions at NORC.

Chapter 5: SNAP Administrative Data Matching

At the beginning of the Initial Interview, the primary respondent was asked to provide consent for FoodAPS to obtain information from State agencies about the household's participation in government food programs. If the primary respondent did not give consent at the interview, the request was repeated at the end of the Final Interview. All but 122 FoodAPS households provided consent.

An attempt was made to match all consenting FoodAPS households to State agency SNAP administrative files after the survey was completed to identify or confirm households that had received SNAP benefits either prior to or during their data collection week. In addition, household records were matched to the Anti-fraud Locator using EBT Retailer Transactions (ALERT) system data, which tracks use of the program's electronic benefits transfer (EBT) card. When a direct link via a SNAP identification was not possible, households were matched using probabilistic matching.

The household interview file contains three variables indicating SNAP participation status: participation as reported in the initial interview (SNAPNOWREPORT); participation based on the match to administrative data (SNAPNOWADMIN), and reported SNAP participation status that is revised per the match to administrative data (SNAPNOWHH). Details about the matching process are provided in separate documentation.

Primary respondents were asked if anyone in the household received SNAP benefits, and if so, when SNAP was last received. Of the 4,826 households in the sample, 1,461 reported that they or someone in the household received SNAP. Based on the date of last receipt, only 1,365 received benefits within the past 30 days at the time of the initial interview. During the Administrative data match, a household was considered to be a current participant if they had last received SNAP benefits within 36 days of the final day of the survey week. Among the 1,581 households determined to be SNAP participants through the combination of the self-report and the administrative match, 74.6 percent had participation confirmed by a data match; 16.8 percent were not confirmed; and 8.6 percent failed to report SNAP at the Initial Interview but were identified as SNAP participants by the matches. Among the 3,243 non-participants, 0.5 percent (16 households) reported SNAP to the survey but ADMIN and/or ALERT data indicated that participation was not active at the time of the survey.

The reported SNAP participation status that is revised per the match to administrative data (SNAPNOWHH) was utilized to construct the final sampling weights (see section 6.1.1).

Chapter 6: Analytic Notes

This section provides analytical information that is common to all the FoodAPS data. Information specific to particular datasets is provided in each data file's codebook.

6.1: Accounting for the complex sampling design

Each household is given a final sampling weight that makes the sample nationally representative of all non-institutionalized households in the contiguous United States. The remainder of this section summarizes how the weights were constructed and provides users with information on how to apply them in analyses, as well as how to obtain standard errors that account for the complex sampling design.

6.1.1 Construction of sampling weights

The household weights were constructed in three stages. In the first stage, the weights accounted for the differences in the probability of selection across households and then were adjusted to account for unit nonresponse. An initial base weight was formed as the reciprocal of the probability of being included in the sampled PSU, SSU, household, and selected replicates. Among the principal adjustments to the base weight were: an adjustment for unknown occupancy status; an adjustment for nonresponse to the screening interview; using the reciprocal of the estimated probability of being included in the study based on a screener-based assignment to a group; and an adjustment for nonresponse to the full study. Appendix D contains more information on the construction of the base weights and weighting steps.

The second stage of the weighting process involved post-stratifying the weights from the first stage to agree with external estimates of the number of households in the United States and the distribution by specific demographic and economic characteristics using a raking process (iterative proportional fitting). The characteristics selected for this post stratification were a combined race/ethnicity variable (Hispanic, Non-Hispanic White only, Non-Hispanic Black only, and Non-Hispanic all other), annual income (<\$15,000; \$15,000 to \$49,999; and \$50,000 and over)⁶, receipt of SNAP (SNAPNOWHH=1, see chapter 5),⁷ household size (1 person; 2 - 4 persons; 5 or more

⁶ The measure of household income used during post-stratification is 12 times either reported income or, for individuals that did not report or refused to report income from one or more income sources, the average of five imputations of monthly income (INCHHAVG).

⁷ A household is treated as a reported SNAP household if anybody within the household was reported as receiving SNAP benefits. This does not mean that everybody within the sampled unit was receiving

persons), number of children in the household (0; 1; 2; 3; 4 or more), and presence of at least one person age 60 or older in the household.⁸ External estimates for Hispanic status, race, income, household size, and number of children were obtained from the 2013 Current Population Survey Annual Social and Economic Supplement. Estimates for SNAP receipt and inclusion of a person age 60 and older were obtained from the 2012 American Community Survey, adjusted to agree with the other control totals based on the 2013 Current Population Survey.

The final stage of the weighting process involved trimming the weights to reduce the variability of the weights and the overall design effect. The trimming procedure was developed for the National Assessment of Education Progress and described in Potter (1990)⁹. Four groups were defined by SNAP status and household income relative to the 2012 poverty thresholds¹⁰ for this process, and each group was assigned a separate

SNAP. For program administration, a SNAP “administrative unit” generally contains everyone in a residence who lives together and purchases and prepares meals together. This rule may exclude some household members from the SNAP administrative unit. It is also possible for multiple SNAP administrative units to reside at the same address. Thus, there is not necessarily a one-to-one correspondence between individuals living in a FoodAPS SNAP household and participation within a single SNAP administrative unit. Similarly, a FoodAPS household receiving SNAP benefits may include non-participants.

⁸ These characteristics were selected for post-stratifying the weights, in part, because an analysis of the risk of non-response bias suggested that they were correlated with both response rates and one or more key variables summarizing household food acquisitions. See Section 6.3 on Non-response Bias.

⁹ Potter, Francis J. “Methods for Extreme Weights in Sample Surveys.” In *The Institute of Statistics, Monograph Series No. 1875T*. Chapel Hill, NC: University of North Carolina at Chapel Hill, Department of Biostatistics, 1990.

¹⁰ Previous steps of sampling and estimation used the poverty guidelines for 2012, but poverty thresholds were used for trimming. Poverty guidelines depend only on the size of the household, but the official poverty thresholds depend on the number of children and whether the head is age 65 or older, in addition to depending on the number of household members. For the trimming step, non-SNAP households were classified according to total reported income relative to the poverty threshold, unless the household reported zero income in the final interview. The 420 households with zero reported income were classified using an estimate of their income, which was the mean in the PSU among households of similar size or the mean among households of similar size across all PSUs (when there were not enough households of the same size within the PSU). This step was done before imputed values of missing income at the individual level were available.

maximum weight.¹¹ The four groups were: (1) SNAP households (SNAPNOWHH=1); (2) non-SNAP households with annual income less than the poverty threshold; (3) non-SNAP households with income above the poverty threshold but less than 185 percent of the poverty threshold; and (4) non-SNAP households with income greater than or equal to 185 percent of the poverty threshold. Post-stratification was repeated after trimming.

6.1.2 *Applying weights and calculating variances*

Each household is given a final sampling weight, HHWGT. The weights were constructed for the household, but they can be applied to individual-level analysis. Software such as SUDAAN, STATA, SAS, and R can be used to estimate sampling errors by the Taylor series (linearization) method using the HHWGT along with the stratum variable, TSSTRATA, and the (pseudo) PSU variable, TSPSU. The variables necessary for Taylor series variance estimation are attached to both the **faps_household** and **faps_individual** data files and examples are provided in the remainder of this section. Users can also estimate variance using the Jackknife Repeated Replication (JRR) (or simply “jackknife”) technique, which requires use of the replicate weights that are provided in the **faps_hhweights** data file. Appendix C provides examples of how to obtain estimates using both methods in Stata, SAS, and R. Stata, SAS, and R support basic and many advanced statistical procedures using the replicate weights, and users may prefer to conduct all of their analyses using them except for some specialized applications where the jackknife might not be applicable.¹² The advantage of the replicate weights is that they incorporate the uncertainty in the nonresponse adjustments and the benefits of the poststratification. In particular, users generally will see somewhat lower variance estimates when using the replicate weights for characteristics correlated with household income. This will be particularly the case

¹¹ In the context of survey weighting, weight trimming usually sets a maximum allowed weight to prevent any one observation from having too much influence on the overall estimate. The maximum weights selected were 80,000; 136,000; 211,000; and 295,000 for the four groups, respectively. After the final post-stratification, some survey weights may have exceeded the maximum set during trimming.

¹² Users are cautioned, however, when using replicate weights in quantile regression, as not much is known about the performance of the jackknife for quantile estimation.

when analyzing characteristics nationally instead of by target group. Information for computing Taylor series estimates has been made available as an alternative for specific situations where the use of replicate weights is not supported.

6.2: Factors Affecting Data Quality

FoodAPS data are obtained from a sample of the population and therefore any estimates using these data may differ from the actual population values because of sampling variability or other factors. Nonresponse to particular survey questions can reduce data quality but was generally low in FoodAPS. Codebooks identify the counts and unweighted percentages of item nonresponse for every variable in the data files. The remainder of this section summarizes three factors that may affect the quality of FoodAPS data: under-reporting of food acquisitions; observational effects; and general nonresponse to the survey.

6.2.1 Underreporting of Food Acquisitions

A form of non-response that is difficult to assess is member cooperation in reporting all food acquisitions throughout the data collection week. This non-response could manifest itself in three different ways: (1) complete refusal to provide information on food acquisitions by a member even when other household members were providing information; (2) member refusal to provide information on food acquisitions on some days but not the entire week; and (3) member under-reporting of food acquisitions even on days when other acquisitions were being reported.

Each time a primary respondent called the SOC to report household members' food acquisitions, the interviewer probed to confirm that members not identified during the call had no food acquisitions to report for each day. From this information a series of daily "status" variables for each member was created with the following possible values: (1) member reported at least one food acquisition for that day; (2) primary respondent confirmed that member did not have any food acquisitions to report for the day; (3) primary respondent reported that member refused to provide information on food

acquisitions that day; and (4) no food acquisitions were reported for the member that day but interviewer unable to confirm that no acquisitions had been made. The latter situation could occur if either no reporting call was made, a call ended before the interviewer could confirm that no acquisitions had been made, or if the primary respondent did not know whether the member had made any acquisitions. Some initially missing information might have been captured in a subsequent phone call, during the final interview, or during review and data entry of information provided on Red pages of the food books.

Table 1 shows the distributions of the status variable values by day of the data collection week. There is a clear drop-off in members reporting food acquisitions by day of the data collection week.¹³ Furthermore, the number of members refusing to provide information remained relatively steady throughout the week (averaging less than 4 percent of all members). Most of the change in status over the week, therefore, is the climb in the number of members with no reported acquisitions but for whom the absence of acquisitions could not be confirmed by the primary respondent or by reviewing food books. Researchers can investigate these data further using the `faps_reportstatus` dataset.

Table 1: Member Food Reporting Status by Day of Week (unweighted counts)

STATUS	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Total
(1) Food acquisition reported	7,823	7,042	6,765	6,450	6,124	5,925	5,780	45,909
(2) PR confirmed no food acquisition	5,048	5,757	5,749	5,927	6,157	5,896	5,726	40,260
(3) Member refused	529	569	574	589	587	531	497	3,876
(4) No acquisition but not confirmed	917	949	1,229	1,351	1,449	1,965	2,314	10,174
Total	14,317	14,317	14,317	14,317	14,317	14,317	14,317	100,219

¹³ Preliminary analysis also shows that the observed decline in days with food acquisitions is less when one controls for day of calendar week (e.g., Sunday, Monday...).

In general, possible underreporting of food acquisitions by individuals on days when acquisitions were reported is the most difficult form of potential underreporting to assess, and ERS has no information on this at this time.

6.2.2 *Observational Effects*

There is a risk that participating households may have changed their food acquisition patterns as a result of participating in the study. One risk is an income effect arising from the incentives provided, especially among the low-income households that were oversampled for FoodAPS. That is why incentives were provided at the end of the data collection week. Although it is possible that some household members may have changed their food acquisitions *in anticipation of* receiving the incentive at the end of the study, that risk seems much lower than if members already had the incentives in hand at the beginning of the week. In addition, with the possible exception of the base incentive of \$100, the gift card values were not believed to be high enough to materially affect food acquisitions prior to their distribution.¹⁴

Another possible observational effect is household members changing their food acquisition patterns as a result solely of participating in the study. A myriad of reasons could explain such an effect, including: wanting to show that healthful foods were being purchased; wanting to avoid having to scan purchased groceries after returning home; doing more shopping trips as family events so children could scan the groceries; not buying lots of food for a big celebration occurring the following week in the (mistaken) belief that such a large purchase would skew the results of the study; and so on.

To gauge the possible prevalence of such changes in food acquisition patterns, interviewers gave primary respondents a feedback form after the final interview was completed. All but 117 of the 4,826 households submitted the form. Table 2 shows the unweighted responses to the question, *“During the past week, did you (or other household members) change the way you got food because you were taking part in the*

¹⁴ The base incentive of \$100 was selected after a pilot study in 2011 demonstrated that response rates with a \$100 incentive were higher than when a \$50 incentive was offered.

study?” For each listed item, fewer than 4 percent of the households indicated that they had changed their behavior. Furthermore, the results suggest some offsetting behaviors when considering the entire sample. For example, whereas 173 of the responding households reported eating out less often because of the survey, 111 households reported eating out more often. (To reduce response burden, no attempt was made to ascertain the magnitude of reported changes.)

Overall, reported observational effects seem limited. Responses to all feedback questions will be released in a separate dataset (**faps_feedback**).

Table 2: Feedback Form Responses (unweighted)

ACTION	Checked	Not checked
Ate out more often	111	4,552
Ate out less often	173	4,490
Did more food shopping	118	4,545
Did less food shopping	170	4,493
Bought specific items to scan	55	4,611
Changed, avoided specific items so you wouldn't have to scan them	23	4,643
Other changes	5	4,661
No, did not change	4,235	431

6.3: Non-Response Bias

Non-response bias in a survey occurs when the answers or responses of respondents differ from those of non-respondents. As overall survey response rates decline, the *risk* of non-response bias increases. The difficulty in measuring non-response bias is that usually little is known about the answers of the non-respondents.

Non-response bias is multi-dimensional. A sample may not exhibit non-response bias along some dimensions (e.g., the gender or age of respondents or their accessibility to supermarkets), while exhibiting non-response bias in other dimensions (e.g., respondent race and ethnicity or household size or income).

Mathematica conducted a preliminary analysis of non-response bias using a two-step process. The first step examined correlations between unweighted rates of FoodAPS non-response and local-level variables drawn from the American Community Survey at the PSU, county, and SSU levels. The second step examined correlation between the area-based variables that were correlated to rates of non-response and four key study measures: (1) whether the household made any food acquisitions during the week; (2) the total number of food acquisitions; (3) the total number of free acquisitions; and (4) the total amount paid during food acquisitions. The risk of non-response bias is highest when rates of non-response are correlated with local measures that, in turn, are related to key study outcomes.

Forty-three (43) area-based variables have a significant ($p \leq .05$) association with study non-response at one or more levels of contact (household determination, screener complete, main study complete). The metropolitan status of PSUs is associated with household determination and study completion. At the county level, a number of economic and demographic (age and race) characteristics are associated with non-response at each step. At the Census tract level, economic, housing (vacancies, renter status) and demographic characteristics (age, race, education, whether Hispanic) are significantly associated with non-response at one or more levels. Finally, several household-level characteristics are associated with screener completion.

Examining the relationships between these 43 variables (and others not correlated with non-response) and the four key study measures:

- 25 variables are associated with both non-response and the level of one or more key study variables;
- 18 of the variables correlated with non-response are not significantly correlated with any of the four study variables; and
- Seven variables are associated with study variables, but are not associated with non-response.

Some of the 25 variables associated with both non-response and key study variables are:

- PSU metropolitan status;
- median household income, percent of families below poverty threshold, percentages of households receiving Social Security benefits or with earnings;
- percentages of individuals in county over 65 years old or are Black/African-American;
- percentages of individuals in the census tract over 16 years old or are Black/African-American;
- percentages of homes in the census tract that are not single units or are vacant;
- percentages of households in the census tract that rent or have earnings, and
- race, language spoken, household size, and SNAP status of screened household.

Household and individual characteristics identified by this analysis were used to post-stratify the sample and adjust sampling weights to reduce expected bias arising from non-response (see section 6.1.1).

Appendix A – List of Data Files and Documentation

Data/documentation name	format(s)	Posted date (revised)
<u>Household-level data</u>		
faps_household	SAS and Stata	October 31, 2014 (12/3/2014) (1/28/2015) (5/7/2015) (6/15/2015) (9/23/2015) (9/30/2015) (5/26/2016)
Household Codebook	Word	October 31, 2014 (12/3/2014) (1/28/2015) (5/7/2015) (5/22/2015) (6/15/2015) (9/23/2015) (9/30/2015) (5/26/2016)
faps_hhweights	SAS and Stata	May 26, 2016
faps_hhgeodata	SAS and Stata	October 27, 2014
HH Geodata Codebook	Word	October 27, 2014
faps_access	SAS and Stata	February 5, 2015
Access Codebook	Word	February 5, 2015
faps_snapelig_hh	SAS and Stata	June 1, 2015
faps_snapelig_unit	SAS and Stata	June 1, 2015
faps_snapelig_ind	SAS and Stata	June 1, 2015
SNAP Eligibility Estimation Codebook	Word	June 1, 2015
<u>Individual-level data</u>		
faps_individual	SAS and Stata	October 31, 2014 (1/28/2015) (5/26/2016)
Individual Codebook	Word	October 31, 2014 (1/28/2015) (5/26/2016)
faps_reportstatus	SAS and Stata	January 28, 2015
faps_meals	SAS and Stata	November 25, 2014
Meals Codebook	Word	November 25, 2014

Data/documentation name	format(s)	Posted date (revised)
<u>Event and Item-level data</u>		
faps_places	SAS and Stata	May 7, 2015
Places Codebook	Word	May 7, 2015
faps_fahevent	SAS and Stata	January 28, 2015 (4/16/2015) (5/7/2015) (5/11/2015) (1/25/2016)
FAH Events Codebook	Word	January 28, 2015 (4/16/2015) (5/7/2015) (5/11/2015) (1/25/2016)
faps_fahitem	SAS and Stata	February 5, 2015 (4/16/2015) (5/11/2015) (1/25/2016)
FAH Items Codebook	Word	February 5, 2015 (4/16/2015) (5/11/2015) (1/25/2016)
faps_fahnutrient	SAS and Stata	January, 25, 2016
FAH Nutrient Codebook	Word	January, 25, 2016
faps_fahitem_iri	SAS and Stata	February 5, 2015 (4/16/2015) (5/11/2015) (1/25/2016)
FAH Item IRI Codebook	Word	February 5, 2015 (4/16/2015) (5/11/2015) (1/25/2016)
faps_fafhevent	SAS and Stata	January 28, 2015 (4/16/2015) (5/7/2015)
FAFH Events Codebook	Word	January 28, 2015 (4/16/2015) (5/7/2015)
faps_fafhitem	SAS and Stata	April 16, 2015 (1/25/2016)
FAFH Items Codebook	Word	April 16, 2015 (1/25/2016)
faps_fafhnutrient	SAS and Stata	January, 25, 2016
FAFH Nutrient Codebook	Word	January, 25, 2016
FoodAPS Nutrient Coding Overview	Word	January, 25, 2016
<u>Supplementary documentation</u>		
FAFH Supplementary Documentation	Word	April 16, 2015
Places Supplementary Documentation	Word	May 7, 2015

Data/documentation name	format(s)	Posted date (revised)
<u>Geography Component (GS) data</u>		
BGMain_v6	SAS and Stata	October 30, 2015
BGRestaurant_v2	SAS and Stata	October 30, 2015
Tract_v5	SAS and Stata	October 30, 2015
County_v6	SAS and Stata	October 30, 2015
GCCodebookRev_7_11_14.doc	Word	On FoodAPS webpage
VariableList_GC_7_11_14.doc	Word	On FoodAPS webpage
basketprices	Stata	February 25, 2016 (3/30/2016)
basketprices_raw	Stata	February 25, 2016 (3/30/2016)
Construction of Construction of Weekly Store-Level Food Basket Costs - Documentation	Word	February 25, 2016
Store Linker Files		
PlaceID_IRI_TempERSID	Excel	March 21, 2016
PlaceID_TDLinx_TempERSID	Excel	March 21, 2016
PlaceID_TempERSID	Excel	March 21, 2016

Appendix B– Merging data files

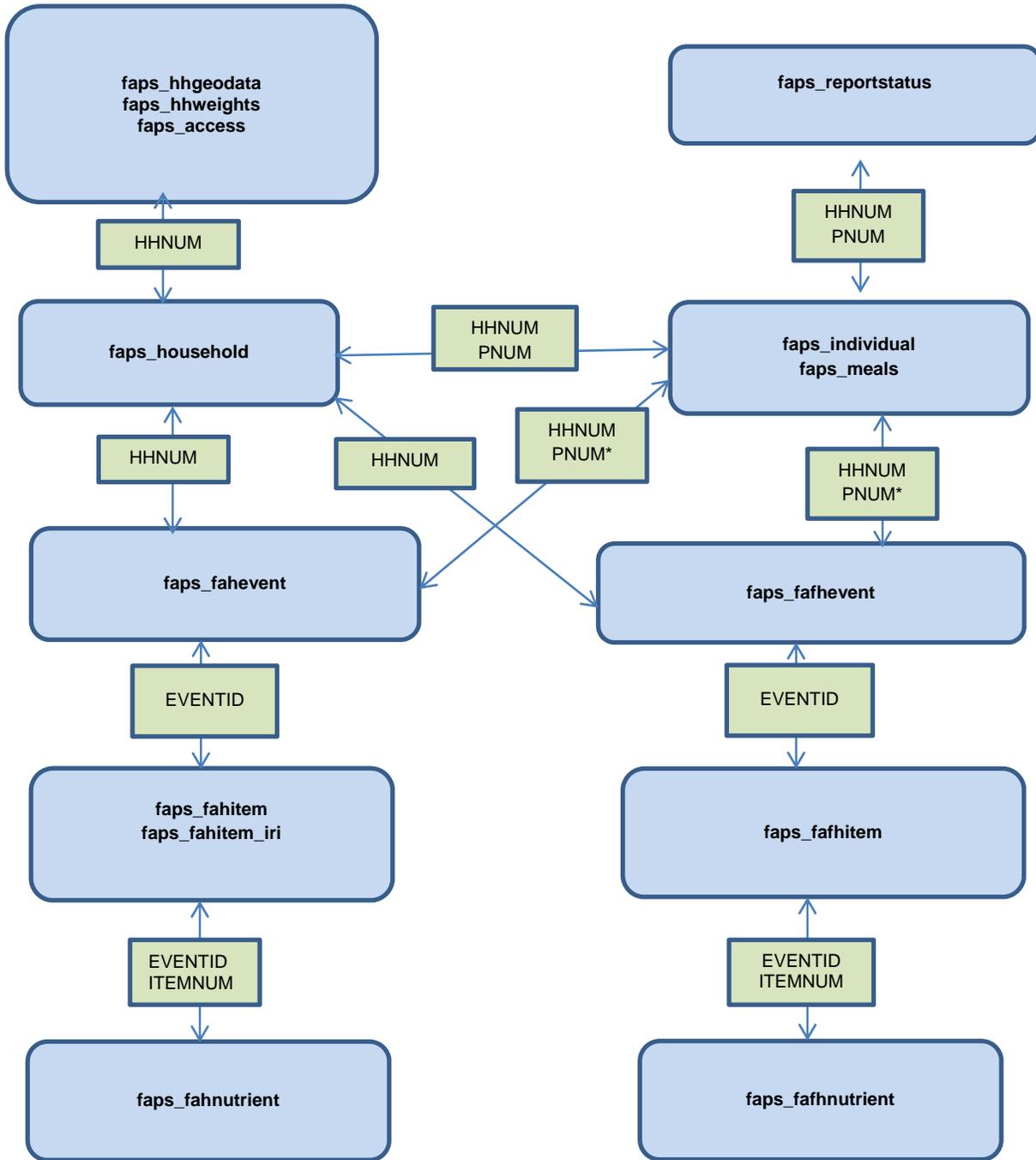
This section summarizes how to merge the various data files together. Table A1 presents the information in tabular format and Figure A1 in a graphical format.

Table B1. Summary of datasets and linkages

Data type dataset	Link to household-level data by	Link to individual-level data by	Link to event-level data by	Link to item-level data by
<i>Household level data</i>				
faps_household faps_hhweights faps_hhgeodata faps_access	HHNUM	HHNUM	HHNUM	HHNUM
<i>Individual-level data</i>				
faps_individual faps_reportstatus faps_meals	HHNUM	HHNUM and PNUM	HHNUM and PNUM*	indirectly through link to event data
<i>Event-level data</i>				
faps_fahevent	HHNUM	HHNUM and PNUM*	n/a	EVENTID
faps_fafhevent	HHNUM	HHNUM and PNUM*	n/a	EVENTID
faps_places			PLACEID	n/a
<i>Item-level data</i>				
FAH faps_fahitem faps_fahitem_iri faps_fahnutrient	HHNUM	indirectly through link to event data	EVENTID (to faps_fah only)	EVENTID and ITEMNUM (within FAH item files only)
FAFH faps_fafhitem faps_fafhnutrient	HHNUM	indirectly through link to event data	EVENTID (to faps_fafh only)	EVENTID and ITEMNUM (within FAH item files only)

*In the event level data PNUM is contained in WHOTGOTPNUM (who obtained the foods for that event) or BOOKPNUM (whose book the event was recorded in). To merge event-level data to individuals, the user must rename either WHOGOTPNUM or BOOKPNUM to PNUM (or create PNUM based on one of these variables) to merge to individual-level interview data using HHNUM and PNUM.

Figure B1: FoodAPS Data Structure and Matching Variables



Note: The variables in the green boxes are the common identifiers used to merge datasets linked by arrows.

*In the event level data PNUM is contained in WHOTGOTPNUM (who obtained the foods for that event) or BOOKPNUM (whose book the event was recorded in). To merge event-level data to individuals, the user must rename either WHOGOTPNUM or BOOKPNUM to PNUM (or create PNUM based on one of these variables) to merge to individual-level interview data using HHNUM and PNUM.

Appendix C– Examples of Variance Estimation

As noted in Section 6.1.2, the weight **HHWGT** provided on **faps_household** and **faps_hhweight** supports estimation of characteristics for the contiguous United States consistent with the complex sample design. Variance codes **TSSTRATA** and **TSPSU** are also provided to support variance estimation through Taylor series linearization. In addition, the file **faps_hhweights** provides 57 replicate weights, **HHWGT1** – **HHWGT57**, which can be used to form variance estimates based on the jackknife replication method. For the most part, the two approaches typically produce similar variance estimates, but using the replicate weights carries the advantage of reflecting the variance impact of the estimation steps to determine **HHWGT**.

SAS, Stata, and R are among the software systems that can use this information for analyzing the survey. This appendix demonstrates how to estimate variances using the replicate weights and through the Taylor series linearization. In the examples below, the results from the three systems agree within numerical precision. The example uses data obtained from the household file, but by merging **faps_hhweights** with other survey files, variances can be obtained for any of the estimates.

In all three systems, there are important restrictions on how to estimate variances for subdomains, particularly when using Taylor series linearization. A correct approach is illustrated for each system.

C.1 Jackknife Replication Method

Proper use of the replicate weights requires specifying the jackknife coefficients or multipliers for each replicate weight. These are provided in Table C1 and the examples indicate how these are specified in SAS, Stata, and R. Most coefficients are 0.5, which corresponds to paired PSUs treated as a pseudo-stratum. A triplet of PSUs is associated with replicates 21-23, and replicates 40-47 are associated with 8 SSUs in the certainty PSU. The general formula for these coefficients is $(n_h - 1)/n_h$, where n_h is the number of units in the stratum.

Table C1. Jackknife coefficient multipliers

HHWGT#	JK coefficient/multiplier
hhwgt1-20	0.5
hhwgt21-23	0.66666667
hhwgt24-39	0.5
hhwgt40-47	0.875
hhwgt48-57	0.5

C.1.1 SAS

The example SAS code for replicate weights (after merging replicate weights hhwgt1-hhwgt57 from faps_hhweights) follows:

```
proc surveymeans data=temp2 varmethod= JACKKNIFE ;
weight hhwgt;
repweights hhwgt1-hhwgt57 / jkcoefs = 0.5 0.5 0.5 0.5 0.5 0.5 0.5
0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5
0.66666667 0.66666667 0.66666667 0.5 0.5 0.5 0.5 0.5 0.5
0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5
0.875 0.875 0.875 0.875 0.875 0.875 0.875 0.875
0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ;
var ADLTFSSRAW;
domain rural;
run;
```

Note the use of “domain” to produce estimates for subdomains. In general, analysts should analyze subsets of their data this way instead of subdividing the data before using proc surveymeans or other related procedures for surveys. Also note that any of this code, including the long statement for repweights, can be cut and pasted from this document for use in other applications.

Using “options linesize=80;” the output list file (condensed) is as follows:

The SURVEYMEANS Procedure

Data Summary

Number of Observations 4826
 Sum of Weights 122459424

Variance Estimation

Method Jackknife
 Replicate Weights TEMP2
 Number of Replicates 57

Statistics

Variable	Label	N	Mean
ADLTFSRAW	Adult food security score - 30-day measure	4826	1.009350

Statistics

Variable	Std Error of Mean	95% CL for Mean	
ADLTFSRAW	0.030772	0.94773030	1.07096952

Domain Statistics in rural

fara:
rural

tract	Variable	Label	N
0	ADLTFSRAW	Adult food security score - 30-day measure	3515
1	ADLTFSRAW	Adult food security score - 30-day measure	1311

Domain Statistics in rural

fara:
rural

tract	Variable	Mean	Std Error of Mean	95% CL for Mean	
0	ADLTFSRAW	1.123068	0.043061	1.03683979	1.20929566
1	ADLTFSRAW	0.786520	0.069926	0.64649553	0.92654468

C.1.2 Stata

The example Stata code follows:

```
svyset [pweight=hhwgt], jkrweight(hhwgt1-hhwgt57, ///
    stratum(1 1 2 2 3 3 4 4 5 5 6 6 7 7 8 8 9 9 10 10 ///
        11 11 11 12 12 13 13 14 14 15 15 16 16 ///
        17 17 18 18 19 19 20 20 20 20 20 20 20 20 ///
        21 21 22 22 23 23 24 24 25 25) ///
    multiplier(0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ///
        0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ///
        0.5 0.5 0.5 0.66666667 0.66666667 0.66666667 ///
        0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ///
        0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ///
        0.875 0.875 0.875 0.875 0.875 0.875 0.875 ///
        0.875 0.875 0.5 0.5 0.5 0.5 0.5 0.5 ///
        0.5 0.5 0.5 0.5)) ///
    vce(jackknife) mse dof(26) singleunit(missing)

svy: mean adltsrsw
svy: mean adltsrsw, over(rural)
```

Note the use of “over” to produce estimates for subdomains. In general, analysts should analyze subsets of their data this way instead of subdividing the data before calling “svyset”. The subpop option (“svy, subpop(conditional statement):”) can also be used to produce estimates for subdomains. Also note any of this code, including the long statement for multiplier, can be cut and pasted from this document for use in other applications.

The results are:

```
. svy: mean rfoodsecscore
(running mean on estimation sample)
Jackknife replications (57)
-----+----- 1 -----+----- 2 -----+----- 3 -----+----- 4 -----+----- 5
.....
.....

Survey: Mean estimation

Number of strata =      25      Number of obs   =      4,826
                               Population size = 122,459,424
                               Replications   =       57
                               Design df     =       26

-----+-----
|               |           Jknife *
|               |           Mean   Std. Err.   [95% Conf. Interval]
-----+-----+-----
adltsrsw        |           1.00935   .0307719   .9460974   1.072602
-----+-----
```

```
. svy: mean adltsrsw, over(rural)
(running mean on estimation sample)
```

```

Jackknife replications (57)
-----+--- 1 -----+--- 2 -----+--- 3 -----+--- 4 -----+--- 5
.....
.....

```

Survey: Mean estimation

```

Number of strata =      25      Number of obs   =      4,826
Population size = 122,459,424
Replications    =          57
Design df      =          26

```

```

0: rural = 0
1: rural = 1

```

-----+-----				
Over	Mean	Jknife * Std. Err.	[95% Conf. Interval]	
-----+-----				
adlthsraw				
0	1.123068	.0430609	1.034555	1.211581
1	.7865201	.0699261	.642785	.9302552
-----+-----				

C.1.3 R

The example R code follows:

```
require(survey)

# replicate weights

design.rw <- svrepdesign(repweights = "hhwt[+0-9]",
  weights = household$hhwt, data = household, type = "JKn", scale=1,
  rscales=c(rep(0.5, 20), rep(0.66666667, 3), rep(0.5, 16),
    rep(0.875, 8),rep(0.5, 10)),
  combined.weights=TRUE, mse=TRUE)

svymean(~adltsraw, design=design.rw)
svyby(~adltsraw, ~rural, design=design.rw, svymean)
rural.rw <- subset(design.rw, rural==1)
svymean(~adltsraw, design=rural.rw)
```

Note the use of “svyby” to produce estimates for subdomains. Another approach to analyze a subset of the data is to produce the survey design object (e.g., design.ts) for the whole data set first, then apply subset() to extract the subdomain of interest.

The results are:

```
> design.rw <- svrepdesign(repweights = "hhwt[+0-9]",
+   weights = household$hhwt, data = household, type = "JKn", scale=1,
+   rscale=c(rep(0.5, 20), rep(0.66666667, 3), rep(0.5, 16),
+     rep(0.875, 8),rep(0.5, 10)),
+   combined.weights=TRUE, mse=TRUE)
>
> svymean(~adltsraw, design=design.rw)
      mean      SE
adltsraw 1.0093 0.0308
> svyby(~adltsraw, ~rural, design=design.rw, svymean)
      rural adltsraw      se
0      0      1.1230677 0.04306089
1      1      0.7865201 0.06992609
> rural.rw <- subset(design.rw, rural==1)
> svymean(~adltsraw, design=rural.rw)
      mean      SE
adltsraw 0.78652 0.0699
>
```

C.2 Taylor Series Linearization

Examples of the Taylor series linearization are provided for SAS, Stata, and R.

C.2.1 SAS

SAS users can use this syntax:

```
proc surveymeans ;
  cluster tspsu ;
  strata tsstrata ;
  var `varlist' ;
  weight hhwt ;
  *domain rural;
run ;
```

* the 'domain' option should be used when estimates for a subpopulation are desired; if using PROC SURVEYFREQ, the domain (subpopulation) variable should be the first variable listed in a tables statement (e.g., tables rural*var1*var2;).

C.2.2 Stata

Stata users can use the svyset command as follows to compute weighted estimates and corresponding estimates of standard errors for the weighted estimates:

```
svyset tspsu [pweight=hhwt], strata(tsstrata)
svy: mean `var'
```

Estimates for subpopulations should use the 'over' option, or the 'subpop' option as indicated below using the rural/urban distinction as an example:

```
svy: mean `var' , over(rural)
svy, subpop(if rural==1): mean `var'
```

C.2.3 R

R users can use the contributed package `survey` and use this syntax to create a survey design object for further analysis:

```
require(survey)
design.ts <- svydesign(ids=~tspstu, strata=~tsstrata,
  nest=TRUE, weights=~hhwgt, data=household, mse=TRUE)
svymeans(~var1 + var2 + var3 , design=design.ts)
```

To get estimates for subpopulations, users should use the `svyby` syntax:

```
svyby(~variable, ~rural, design=design.ts, svymean)
```

Appendix D– Further Details on Sampling and Weighting

Sections 2.2, 2.3, and 2.4 provided an account of the first stages of sampling. To review, first the counties in the 48 contiguous States and the District of Columbia were divided into 948 PSUs. Of these, 50 PSU were included in the FoodAPS study, with 1 certainty PSU. The other 49 PSUs were selected with probability proportional to a measure of size using systematic sampling from a list of the remaining 947 PSUs.

As a minor detail, 9 of the sampled PSUs comprised 5 or more counties. They were subdivided into 2 or 3 groups of counties, and one county group was selected to represent the original PSU. Except for an adjustment to the weights, the sampled groups of counties were treated as equivalent to the other PSUs.

Sampled PSUs were then divided into secondary sampling units (SSUs) typically consisting of a single block group but occasionally two or more adjacent block groups. Within each PSU, eight SSUs were selected, yielding 400 sampled SSUs in total.

Section 2.5 describes the third stage of sampling, namely, the sampling of addresses within sampled SSUs. In the majority of SSUs, the ABS list was first unduplicated from the SNAP list, so that each housing unit would appear in only one list. Sample addresses were then selected from the updated ABS list and the SNAP lists at different rates. In 14 SSUs, listing was required. SNAP lists were unavailable in five states, so the sample there was drawn only from the ABS sample.

Section 2.5.2 describes the division of the initial sample into 70 equal size replicates, followed by a supplemental sample of 41 more replicates from the non-SNAP frame. In an attempt to achieve the goals for the target groups in section 2.2, Mathematica varied the number of replicates they released by SSU.

At any point in the sampling, the probabilities of selection varied substantially. When properly constructed, survey weights can compensate for the effect of unequal sampling probabilities. The most common strategy for weighting probability samples of a finite population is to begin by computing the base weight, that is, the reciprocal of the probability of selection, for each sample unit. Because of the complexities of sampling

for the 2012 FoodAPS, a total of eight probabilities are needed to compute the base weights, which are the reciprocals of the product of the following probabilities:

1. The probability of selecting the PSU (=1 in the case of the certainty PSU).
2. The probability of selecting the county group in the 9 sample PSUs that were divided because of their large size (=1 in the remaining 41 PSUs.)
3. The probability of selecting the SSU within the PSU/county group.
4. The probability of selecting the tertiary sampling unit (TSU) within the sampled SSU, in SSUs that were listed (=1 in all other SSUs).
5. The probability of sampling the address within the SSU/TSU.
6. The probability of selecting the unit within large drop points. Some sample addresses were drop points rather than individual housing units. The post office delivers mail at a drop point for two or more units to a common location. In other words, the housing units share the same postal address.
7. The proportion of replicates released for this SSU and type of frame (SNAP, ABS).
8. The probability of being sampled for the Phase 2 sample for those in the Phase 2 frame (=1 otherwise).

Note that after the unconditional probability of selection of the PSU, all of the remaining probabilities are conditional upon the unit in the previous step being included in the sample.

Base weights were produced for 19,237 addresses, that is, all of the addresses except the Phase 2 frame that were not sampled for Phase 2 follow up.

Subsequent weighting steps compensated for nonresponse and excluded cases that were ineligible. The steps were:

1. An adjustment for units with undetermined eligibility. Vacant and otherwise unoccupied housing units were ineligible, but occupancy status was unknown for some units. The adjustment distributed the weights of the units with unknown

occupancy among the occupied and vacant units. The vacant units were then dropped, leaving 15,968 occupied units in the sample.

2. An adjustment for units not completing the screener questionnaire. Group quarters and unknown status were dropped as ineligible, leaving 15,796 units. Of these, 12,300 completed the screener and received positive weights at the end of this step.
3. An adjustment for 44 units that were not classified into a group. The weight for these cases was distributed among the classified units. As part of this step, 186 seasonal units and 1 with unknown status were removed. As a result, 12,069 cases received positive weights at the end of this step.
4. An adjustment for the quota group subsampling. As noted in section 6.1.1, screening information became the basis for assigning each case to a quota group. One quota group was fully included, and another group was subsampled at approximately 1 in 4. Although in principle an exact determination of the required probability could have been reconstructed from precise records, a modeling approach was developed to estimate the probability and in turn to weight the sample. As part of this step, weights were adjusted to include 80 cases that should have been dropped as a result of the quota group subsampling, but they were nonetheless included among the completed cases.¹⁵ At the end of this step, 7,650 cases have positive weights.
5. An adjustment for nonresponse for 1,277 cases who completed the screening interview and were eligible for the study, but who declined to participate further. At the end of this step, 6,373 cases have positive weights.
6. An adjustment for 1,547 cases who were eligible to complete the study, but who did not do so, leaving 4,826 completed cases with positive weights.

¹⁵ These 80 residential units were eligible for the survey in all respects except the quota group subsampling, which was designed to more efficiently use survey resources by not interviewing all the households in easier-to-locate quota groups. Once the 80 units were erroneously included, there was no reason not to keep them in the sample with sample weights appropriately adjusted.

The weights from the last step were the input to the first raking procedure described in section 6.1.1. As noted there, the first raking was followed by a weight trimming procedure and a final raking.

Appendix E– Summary of Revisions

November 25, 2014

- Meals and snack form data posted and added to Appendix A list.

January 28, 2015

- Definition of target groups (section 2.2) was revised for clarity and correctness.
- Explanation of the construction of sampling weights was expanded for clarity and completeness.

February 5, 2015

- Access data added to list of posted data.
- FAH item data added to list of posted data.

April 15, 2015

- FAFH item data added to list of posted data
- Dates for updated files added.
- Reference to jackknife replicate weights removed while these weights undergo further review
- Section 6.3 edited

May 6, 2015

- Places data added to list of posted data
- Dates for updated files added.

May 11, 2015

- One event from FAH event data dropped.
- Dates for updated files added.

May 27, 2015

- The number of FAH and FAFH event records (chapter 4) was corrected.
- A brief description of the **faps_places** file was added to chapter 4.

May 29, 2015

- The SNAP eligibility data files and codebook were added to chapter 4 and the list of data files (appendix A).

June 12, 2015

- Update to Household data and codebook noted

September 23, 2015

- Update to Household data and codebook noted

September 30, 2015

- Update to Household data and codebook noted

October 27, 2015

- The description of the FAH and FAFH data (section 3.4) was revised for accuracy and clarity.
- Other sections of the text were revised to align versions of the document.

January 25, 2016

- Posting of FAH and FAFH nutrient data and related documentation (codebooks and Nutrient Coding Overview) list in Appendix A, list of Datafiles. Chapter 4 was edited to reflect the fact that the nutrient data files were posted and are no longer forthcoming.

February 19, 2016

- Added information about the Food Basket Price data (part of the Geography Component data) that was posted to NORC to chapter 4 and listed the data files in Appendix A.

May 26, 2016

- The summary of the results of the match to SNAP administrative data in determining SNAP participation in chapter 5 was corrected.
- Revised Appendices A and B to accurately reflect the data files available at NORC.
- Revised chapter 6 to account for changes in the construction of the sampling weights; added instructions for weighted mean and variance estimation in R.
- Added Appendix C which provides examples of how to estimate means and variances using jackknife and the replicate weights for SAS, Stata, and R users.
- Appendix D added to provide additional details on the sampling and weighting process.
- This appendix (summary of revisions) became Appendix E.

June 20, 2016

- The document was revised to be consistent with the version on the secure enclave.