United States Department of Agriculture

Economic Research Service

November 2016

National Household Food Acquisition and Purchase Survey (FoodAPS)

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

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).

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

Suggested citation:

National Household Food Acquisition and Purchase Survey (FoodAPS): User's Guide to Survey Design, Data Collection, and Overview of Datasets. U.S. Department of Agriculture, Economic Research Service, November 2016.

In accordance with Federal civil rights law and U.S. Department of Agriculture (USDA) civil rights regulations and policies, the USDA, its Agencies, offices, and employees, and institutions participating in or administering USDA programs are prohibited from discriminating based on race, color, national origin, religion, sex, gender identity (including gender expression), sexual orientation, disability, age, marital status, family/parental status, income derived from a public assistance program, political beliefs, or reprisal or retaliation for prior civil rights activity, in any program or activity conducted or funded by USDA (not all bases apply to all programs). Remedies and complaint filing deadlines vary by program or incident.

Persons with disabilities who require alternative means of communication for program information (e.g., Braille, large print, audiotape, American Sign Language, etc.) should contact the responsible Agency or USDA's TARGET Center at (202) 720-2600 (voice and TTY) or contact USDA through the Federal Relay Service at (800) 877-8339. Additionally, program information may be made available in languages other than English.

To file a program discrimination complaint, complete the USDA Program Discrimination Complaint Form, AD-3027, found online at How to File a Program Discrimination Complaint and at any USDA office or write a letter addressed to USDA and provide in the letter all of the information requested in the form. To request a copy of the complaint form, call (866) 632-9992. Submit your completed form or letter to USDA by: (1) mail: U.S. Department of Agriculture, Office of the Assistant Secretary for Civil Rights, 1400 Independence Avenue, SW, Washington, D.C. 20250-9410; (2) fax: (202) 690-7442; or (3) email: program.intake@usda.gov.

USDA is an equal opportunity provider, employer, and lender.

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 Sampling frames 5
2.5.2 Sampling addresses 6
Chapter 3: Data Collection Process 7
3.1: Data Collection Protocols and Instruments 7
3.2: The Data Collection Week 9
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 Datasets 11
4.1: Restricted Use Files 11
4.2: Public Use Files 15
4.2.1 Personal Identifiers and Geographic Identifiers 16
4.2.2 Data Coarsening 17
4.2.3 Swapping 18
Chapter 5: SNAP Administrative Data Matching 19
Chapter 6: Analytic Notes 20
6.1: Accounting for the Complex Sampling Design 21
6.1.1 Construction of Sampling Weights 21
6.1.2 Applying Weights and Calculating Variances 23
6.2: Factors Affecting Data Quality 24
6.2.1 Underreporting of Food Acquisitions 24
6.2.2 Observational Effects 26

6.3: Non-Response Bias 27
Appendix A – List of RUF and PUF Data Files and Documentation 30
Table A1. RUF data files and documentation 30
Table A2. PUF data files and documentation 32
Appendix B – Merging Data Files 33
Table B1. Summary of datasets and linkages 33
Figure B1: FoodAPS Data Structure and Matching Variables 34
Appendix C – Variables Suppressed or Coarsened on PUFs 35
Table C1. Variables suppressed on PUFs 35
Table C2. Variables coarsened (recoded, top-coded, or derived) on PUFs 45
Appendix D – Examples of Variance Estimation 49
D.1 Jackknife Replication Method 49
D.1.1 SAS 50
D.1.2 Stata 52
D.1.3 R 54
D.2 Taylor Series Linearization 55
D.2.1 SAS 55
D.2.2 Stata 55
D.2.3 R 56
Appendix E – Further Details on Sampling and Weighting 57
Appendix F – Summary of Revisions for RUFs 61

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 Supplemental Nutrition Assistance Program (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 7-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 5 years in prison.

Two versions of the FoodAPS data have been released: public use files (PUF) and restricted use files (RUF). Researchers requesting access to the FoodAPS RUF 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 <u>and</u> 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 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 4 of the 400 SSUs contained more than 1 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). A total of 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 Free 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 lists a large number of addresses that were not useful for locating households (e.g., P.O. boxes, Rural Free Delivery), addresses in the sampled SSU were listed by

[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.

Mathematica field staff. The listed addresses comprised the sampling frame in that SSU.[2]

- **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

[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.

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 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 households that were most difficult to locate 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 one in four 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 1-week period. The PR and other Adult food books contained Daily List pages, which documented summary information for each acquisition for each of the 7 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] The terms "food at home" and "food away from home" are defined in Section 3.4.

3.2: The Data Collection Week

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

- Initial Household Visit
  - Completed *Household Screener*
  - PR provided consent and completed the *Initial Household Interview* based on Computer-Aided Personal Interviewing (CAPI)
  - 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 books[4]
  - 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

[4] 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*.

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).
- 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 datasets. 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 70.9 percent, and the overall study response rate (OSRR) was 41.5 percent. The OSRR is equivalent to the Response Rate 3 (RR3) of the American Association for Public Opinion Research. More details on response rates are available upon request.

Chapter 4: FoodAPS Datasets

The data collected by FoodAPS is provided in a number of different data files. The restricted use files (RUFs) contain the complete set of variables collected in the survey, but some variables may be confidential or sensitive. As a solution, a set of public use files (PUFs) were created by implementing statistical disclosure control (SDC) approaches on the RUFs.

4.1: Restricted Use Files

Table A1 in appendix A provides a list of the RUFs 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 datasets.

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 the variable 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 FAH acquisition. For the file **faps_fafhevent**, each of the 39,120 records represents a 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 the variable EVENTID, which is unique across all FAH and FAFH events across all households.

Some barcodes in the FoodAPS data were matched to a proprietary dataset 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 micronutrients and the Food Pattern Equivalents per 100 grams of each identifiable food. The FAH nutrients are in the **faps_fahnutrients** 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 and 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 is 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 Federal Information Processing Standard (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.

4.2: Public Use Files

Over the past decade, concerns about the disclosure of information related to individual survey respondents have increased dramatically. The main challenge for FoodAPS is a challenge present for all surveys: balancing the needs of data users to have data with the highest possible degree of accuracy and fidelity with the need to comply with Federal confidentiality laws. Laws have been passed since the Privacy Act of 1974 to further ensure the protection of confidential data. The most recent of these is the Confidential Information Protection and Statistical Efficiency Act (CIPSEA),[5] which requires that statistical agencies protect the confidentiality of all those responding to their sponsored surveys.

Several types of data were collected and derived during the FoodAPS sampling, data collection, and weighting processes. These variables were reviewed to determine their disclosure risk levels. The confidentiality analysis used a three-step process to reduce disclosure risk: (1) determining the disclosure risk arising from existing external data, (2) coarsening the data, and (3) perturbing the data. Westat conducted the risk analysis, coarsening, and data perturbation procedures to produce the PUFs that are included in the FoodAPS data dissemination products.

Note that due to the data perturbation that occurred in the PUFs, the results from the PUFs and the RUFs can differ slightly. The RUFs contain non-coarsened, non-perturbed data and the PUFs contain coarsened, perturbed data. However, no changes were made to the nutrient data files (**faps_fahnutrients** and **faps_fafhnutrients**) for their release in the PUF.

[5] Title V of the E-Government Act of 2002 is the Confidential Information Protection and Statistical Efficiency Act (CIPSEA). CIPSEA protects from improper disclosure any data collected under a confidentiality pledge that are intended for statistical purposes. The concern of data users is that the integrity of the data may be compromised in order to make the data safe for release by applying statistical disclosure control treatments.

Table A-2 in appendix A provides a list of the PUFs and documentation files. Merging PUFs can be done in the same way as described for merging RUFs in appendix B, although there are fewer PUFs available.

The approach for creating the PUFs included categorizing or top-coding variables with some risk of data disclosure and suppressing variables with a very high risk of disclosure. These actions were based on extensive initial disclosure risk analyses, which included the following steps:

- Identifying personal identifiers, geographic information, and contextual variables (variables that can indirectly identify a geographic area);
- Evaluating the existence of other publicly available files;
- Evaluating the disclosure risk associated with release of the sampling and variance estimation variables;
- Evaluating the disclosure risk associated with release of key variables (i.e., visible variables) through extensive frequency tables and various risk metrics; and
- Exhaustive review of each data item in the RUFs to determine whether any of the data presented a non-negligible risk of individual disclosure.

Careful attention was given to the interview items and to combinations of indirect identifying items. Even a very limited amount of detail among indirect identifying variables—such as income, age, household size, race/ethnicity, geographic location, and birthplace—can increase the chance that an individual can be identified.

4.2.1 Personal Identifiers and Geographic Identifiers

Any information that might be used to directly identify sample persons and/or sample locations was suppressed from the PUFs and RUFs. This information included names, addresses, and telephone numbers. The inclusion of any geographic detail has a large impact on the level of disclosure risk. Review of the available variables indicated that there were several geographic variables, such as the PSU and segment identifiers that

should not be available in the PUFs. Only Census region (four levels), tract-level urban-rural designation (two levels), and Census Core-Based Statistical Area (CBSA) flag (two levels) were retained on the PUFs; all other geographic indicators were suppressed.

4.2.2 Data Coarsening

In general, data coarsening includes several types of procedures that decrease disclosure risk by reducing the amount of information released. Coarsening approaches include removing direct identifiers, limiting geographic detail, categorizing continuous variables, performing top-coding,[6] and recoding values into broader categories. Suppression was also performed by removing variables from the file that could lead to a high risk of identifying individuals. After the overall re-identification risk was measured, the results of frequencies and multi-way cross-tabulations were used to guide the coarsening process. First, one-way tabulations were reviewed to determine the categorical variables that would need to be recoded because of the small number of responses in one or more categories. Next, multi-way tabulations were reviewed to identify variables that had problematic categories when used in combination with other variables.

Suppressed Variables. A number of variables were suppressed due to low frequency counts (i.e., frequencies of less than 15) because the variables revealed too much detail about the sample design units and geographical location. Most variables related to the sample design units were suppressed in response to disclosure concerns about being able to indirectly identify the location of the respondents. Informed by a risk assessment, the variance strata and variance unit variables were retained on the PUFs to facilitate variance estimation. All open-ended variables (e.g., “other, specify” responses) were also suppressed from the PUFs to prevent the possibility of revealing geography or an individual identity from the responses. A complete list of variables that were suppressed in the PUFs can be found in table C1 of appendix C.

[6] With top-coding, the largest values of a variable are replaced with an upper limit, reducing the appearance of outlier data.

Recoded Variables. The process of recoding categorical variables helps to minimize the risk of data disclosure. Variables were suppressed instead of recoded if recoding would have resulted in losing the meaning and usefulness of the data in order to provide adequate protection. The process for recoding categorical variables involved grouping different levels of the same variable to create categories with larger frequencies. This reduced the risk of identifying a respondent when an attack consists of combining survey variables. A list of the recoded variables is provided in table C2 of appendix C.

Continuous Variables. Top-coding was performed for some continuous variables, or ordinal variables with several categories. One approach considered was to categorize all continuous variables, since, with top-coding, there is still potential for bias on computations of the average for subgroups or for regression analysis. Categorization protects against bias and reduces disclosure risk. However, it was decided to use top-code cutoffs for some continuous variables because the analytical value of continuous versions of some variables outweighed the potential disclosure risk. The criterion was that the proportion of cases with values greater than the cutoff was 1 percent of the weighted sample size. The top-coding cutoff was used as the replacement value for the cases with values greater than the cutoff. A list of the top-coded variables and cutoff points are provided in table C2 of appendix C.

Derived Variables. New variables were derived from the suppressed variables in order to provide data useful to researchers while ensuring data confidentiality. A complete list of the derived variables can be found in table C2 of appendix C.

4.2.3 Swapping

To ensure that a data intruder could never identify an individual, the risk of data disclosure was further reduced for the PUFs by using a data swapping technique. Data swapping reduces risk by modifying microdata. The swapping approach, as discussed further in Krenzke et al. (2014),[7] was applied to the individual sample. The steps

[7] Available at: http://support.sas.com/resources/papers/proceedings14/1603-2014.pdf

included (1) estimating the disclosure risk for the file and for each record, (2) selecting target records to be swapped, (3) selecting swapping partners, and (4) swapping the data.

In data swapping as applied to FoodAPS, a probability sampling of records were selected. Some records with the highest risk were selected with certainty for the swapping. Other records were selected with probability proportionate to the estimated risk of the record. Swapping partners were then selected with regards to its closeness in terms of a set of variables of interest. Then variables were swapped between the two records. The sampling rate for swapping was assigned to protect the confidentiality of the data without affecting the usability of the dataset. This method is an effective way of keeping as much valuable data as possible while not identifying any research participants.

Swapping preserves the unweighted frequencies, means, and variances; however, it may affect the weighted distributions and multivariate relationships. Checks were conducted to examine the distributions of and the relationships between variables pre- and post-swapping.

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, 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),[8] receipt of SNAP

[8] 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).

(SNAPNOWHH=1, see chapter 5),[9] household size (1 person; 2-4 persons; 5 or more 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.[10] 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).[11] Four groups were defined by SNAP status and household income relative to the 2012 poverty thresholds[12] for this process, and each group was assigned a separate

---

[9] 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 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.

[10] 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.

[11] 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.

[12] 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

maximum weight.[13] 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 D 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.[14] The advantage of the replicate weights is that they incorporate the uncertainty in the nonresponse adjustments and the benefits of the post-stratification. In particular, users generally will see somewhat lower variance estimates when using the replicate weights

---

the same size within the PSU). This step was done before imputed values of missing income at the individual level were available.

[13] 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.

[14] 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.

for characteristics correlated with household income. This will particularly be the case 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: underreporting 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 underreporting 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 dropoff in members reporting food acquisitions by day of the data collection week.[15] 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 (RUF only).

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

[15]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.[16]

Another possible observational effect is household members changing their food acquisition patterns as a result solely of participating in the study. Myriad 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*

[16] 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.

household members) change the way you got food because you were taking part in the 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 are available in the PUF (in the **faps_household_puf** dataset) and the RUF (in the **faps_feedback** dataset).

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
- 7 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 who are Black/African-American;
- percentages of individuals in the census tract over 16 years old or who 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 RUF and PUF Data Files and Documentation

Table A1. RUF data files and documentation

Data/documentation name	format(s)	Date first posted (last revised)
Household-level data		
faps_household	SAS and Stata	October 31, 2014 (9/2/2016)
Household Codebook	Word	October 31, 2014 (9/2/2016)
faps_hhweights	SAS and Stata	May 26, 2016
jkwgt_factors	SAS and Stata	October 25, 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
Individual-level data		
faps_individual	SAS and Stata	October 31, 2014 (5/26/2016)
Individual Codebook	Word	October 31, 2014 (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
Event and Item-level data		
Places Supplementary Documentation	Word	May 7, 2015
faps_places	SAS and Stata	May 7, 2015
Places Codebook	Word	May 7, 2015
faps_fahevent	SAS and Stata	January 28, 2015 (1/25/2016)
FAH Events Codebook	Word	January 28, 2015 (1/25/2016)
faps_fahitem	SAS and Stata	February 5, 2015 (1/25/2016)
FAH Items Codebook	Word	February 5, 2015 (1/25/2016)
faps_fahitem_iri	SAS and Stata	February 5, 2015 (1/25/2016)
FAH Item IRI Codebook	Word	February 5, 2015 (1/25/2016)

Data/documentation name	format(s)	Date first posted (last revised)
Event and Item-level data (continued)		
FAFH Supplementary Documentation	Word	April 16, 2015
faps_fafhevent	SAS and Stata	January 28, 2015 (5/7/2015)
FAFH Events Codebook	Word	January 28, 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)
Nutrient data		
FoodAPS Nutrient Coding Overview	Word	January, 25, 2016 (10/25/2016)
faps_fahnutrient	SAS and Stata	January, 25, 2016 (10/25/2016)
FAH Nutrient Codebook	Word	January, 25, 2016 (10/25/2016)
faps_fafhnutrient	SAS and Stata	January, 25, 2016 (10/25/2016)
FAFH Nutrient Codebook	Word	January, 25, 2016 (10/25/2016)
Geography Component (GS) data		
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 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

Table A2. PUF data files and documentation

Data/documentation name	format(s)	Posted date
Household-level data		
faps_household_puf*	SAS, Stata, CSV	November 9, 2016
Household Codebook_PUF	Word	November 9, 2016
faps_access_puf	SAS, Stata, CSV	November 9, 2016
Access Codebook_PUF	Word	November 9, 2016
jkwgt_factors	SAS, Stata, CSV	November 9, 2016
Individual-level data		
faps_individual_puf	SAS, Stata, CSV	November 9, 2016
Individual Codebook_PUF	Word	November 9, 2016
faps_meals_puf	SAS, Stata, CSV	November 9, 2016
Meals Codebook_PUF	Word	November 9, 2016
Event and Item-level data		
faps_fahevent_puf	SAS, Stata, CSV	November 9, 2016
FAH Events Codebook_PUF	Word	November 9, 2016
faps_fahitem_puf	SAS, Stata, CSV	November 9, 2016
FAH Items Codebook_PUF	Word	November 9, 2016
faps_fafhevent_puf	SAS, Stata, CSV	November 9, 2016
FAFH Events Codebook_PUF	Word	November 9, 2016
faps_fafhitem_puf	SAS, Stata, CSV	November 9, 2016
FAFH Items Codebook_PUF	Word	November 9, 2016
Nutrient data		
FoodAPS Nutrient Coding Overview**	Word	November 9, 2016
faps_fahnutrient**	SAS, Stata, CSV	November 9, 2016
FAH Nutrient Codebook**	Word	November 9, 2016
faps_fafhnutrient**	SAS, Stata, CSV	November 9, 2016
FAFH Nutrient Codebook**	Word	November 9, 2016

* Household PUFs also contain some variables from the feedback form file and the household-level SNAP-eligibility estimation file.
** The FAH and FAFH nutrient data files are the same in the PUF and RUF.

Appendix B – Merging Data Files

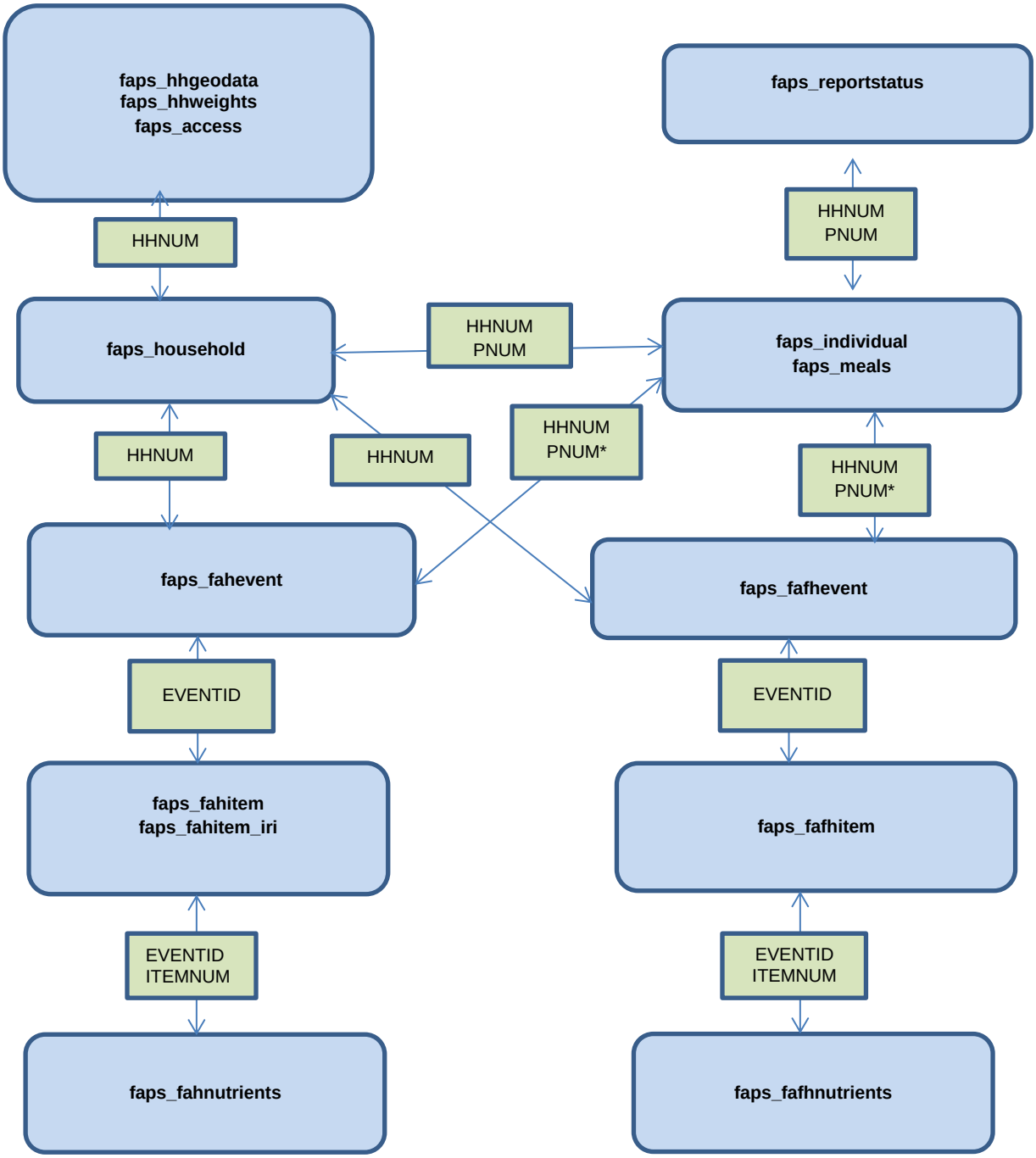
This section summarizes how to merge the various data files together. Table B1 presents the information in tabular format and figure B1 in a graphical format.

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
Household level data				
faps_household faps_hhweights faps_hhgeodata faps_access	HHNUM	HHNUM	HHNUM	HHNUM
Individual-level data				
faps_individual faps_reportstatus faps_meals	HHNUM	HHNUM and PNUM	HHNUM and PNUM*	indirectly through link to event data
Event-level data				
faps_fahevent	HHNUM	HHNUM and PNUM*	n/a	EVENTID
faps_fafhevent	HHNUM	HHNUM and PNUM*	n/a	EVENTID
faps_places			PLACEID	n/a
Item-level data				
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_fafhnutrients	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 – Variables Suppressed or Coarsened on PUFs

Table C1. Variables suppressed on PUFs

Dataset	Variable Name	Variable Label
Household	INITIALINTRVID	Identification code of interviewer conducting initial interview
Household	INITIALLANG	Language in which initial interview was conducted
Household	FINALINTRVID	Identification code of interviewer conducting final interview
Household	FINALLANG	Language in which final interview was conducted
Household	PRDISPHONE	Primary respondent has difficulty using phone b/c of disability
Household	PRDISWRITING	Primary respondent has difficulty writing b/c of disability
Household	PRDISCOGNITIVE	Primary respondent has difficulty with memory/concentration/making decisions
Household	PRDISVISION	Primary respondent has vision/other problem making it hard to read
Household	SURVASSIST	Someone assisted primary respondent with data reporting
Household	LODGERS	Number of lodgers living at residence
Household	BOARDERS	Number of boarders living at residence
Household	HHSIZEBIRTH	Child was born into the household in past 3 months
Household	HHSIZEOTHCHILD	Step-, foster, or adopted child entered the household in past 3 months
Household	HHSIZESEPARATION	Separation or divorce occurred in the household in past 3 months
Household	HHSIZEDEATH	Death occurred in the household in past 3 months
Household	HHSIZEMARRIAGE	Marriage occurred in the household in past 3 months
Household	HHSIZEPARTNER	New partner entered the household in past 3 months
Household	HHSIZEOTH	Household sized changed in past 3 months for another reason
Household	HHSIZEOTHSP	Reason why household size changed in past 3 months
Household	FARMWORKERHH	Anyone in household is a migrant or seasonal worker (y/n)
Household	LIQASSETS2000	Household has $2,000 or more in liquid assets (y/n)
Household	LIQASSETS3000	Household has $3,000 or more in liquid assets (y/n)
Household	EXPOPMEDICAL60	Out-of-pocket medical expenses last month for those 60 and older or disabled
Household	EXPMEDICAL60COM	Out-of-pocket medical expenses reported as part of another expense
Household	EXPAMT1	Amount paid for rent or mortgage last month
Household	EXPFREQ1	Frequency of reported rent or mortgage payment
Household	EXPAMT2	Amount paid for homeowner's or renter's insurance last month
Household	EXPFREQ2	Frequency of reported homeowner's or renter's insurance payment

Dataset	Variable Name	Variable Label
Household	EXPCOM2	Homeowner's or renter's insurance payment reported as part of another expense
Household	EXPAMT3	Amount paid for property tax last month
Household	EXPFREQ3	Frequency of reported property tax payment
Household	EXPCOM3	Property tax payment reported as part of another expense
Household	EXPAMT4	Amount paid for public transport last month
Household	EXPFREQ4	Frequency of reported public transport payment
Household	EXPCOM4	Public transport payment reported as part of another expense
Household	EXPAMT5	Amount paid for electricity last month
Household	EXPFREQ5	Frequency of reported electricity payment
Household	EXPCOM5	Electricity payment reported as part of another expense
Household	EXPAMT6	Amount paid for heating fuel last month
Household	EXPFREQ6	Frequency of reported heating fuel payment
Household	EXPCOM6	Heating fuel payment reported as part of another expense
Household	EXPAMT7	Amount paid for sewer and garbage collection last month
Household	EXPFREQ7	Frequency of reported sewer and garbage collection payment
Household	EXPCOM7	Sewer and garbage collection payment reported as part of another expense
Household	EXPAMT8	Amount paid for health insurance last month
Household	EXPFREQ8	Frequency of reported health insurance payment
Household	EXPAMT9	Amount paid for health insurance copays last month
Household	EXPFREQ9	Frequency of reported health insurance copays
Household	EXPCOM9	Health insurance copays reported as part of another expense
Household	EXPAMT10	Amount paid to doctors or hospitals last month
Household	EXPFREQ10	Frequency of reported doctor or hospital payments
Household	EXPCOM10	Doctor or hospital payments reported as part of another expense
Household	EXPAMT11	Amount paid for prescription drugs last month
Household	EXPFREQ11	Frequency of reported prescription drug payments
Household	EXPCOM11	Prescription drug payments reported as part of another expense
Household	EXPAMT12	Amount paid for child care last month
Household	EXPFREQ12	Frequency of reported child care payment
Household	EXPCOM12	Child care payment reported as part of another expense
Household	EXPAMT13	Amount paid for child support last month
Household	EXPFREQ13	Frequency of reported child support payment

Dataset	Variable Name	Variable Label
Household	EXPCOM13	Child support payment reported as part of another expense
Household	EXPAMT14	Amount paid for adult care last month
Household	EXPFREQ14	Frequency of reported adult care payment
Household	EXPCOM14	Adult care payment reported as part of another expense
Household	ADMINAMT1	Last SNAP amt, admin/caseload data, match1
Household	ADMINAMT2	Last SNAPSNAP amt, admin/caseload data, match2
Household	ADMINAMT3	Last SNAP amt, admin/caseload data, match3
Household	ADMINAMT4	Last SNAP amt, admin/caseload data, match4
Household	ALERTAMT1	Last SNAP amt, alert data, match1
Household	ALERTAMT2	Last SNAP amt, alert data, match2
Household	MATCHADMIN	Result of match to SNAP admin/caseload data
Household	MATCHALERT	Result of match to SNAP alert data
Household	SNAPLASTADMIN1	Date last SNAP, admin/caseload data, match1
Household	SNAPLASTADMIN2	Date last SNAP, admin/caseload data, match2
Household	SNAPLASTADMIN3	Date last SNAP, admin/caseload data, match3
Household	SNAPLASTADMIN4	Date last SNAP, admin/caseload data, match4
Household	SNAPLASTALERT1	Date last SNAP, alert data, match1
Household	SNAPLASTALERT2	Date last SNAP, alert data, match2
Household	SNAPSTATEGRP	Type of SNAP admin data from state
Household	PRIMSTORENAME	Name of household's primary food store
Household	PRIMSTORENAME_ERS	ERS edited primary store place name
Household	PRIMSTOREEDIT_SPACES	ERS removed spaces to edit place name
Household	PRIMSTOREEDIT_CASE	ERS edited place name case
Household	PRIMSTOREEDIT_NAME	ERS edited the place name
Household	PRIMSTORESOURCE	Source of information and method of cleaning name/address of primary store
Household	PRIMSTORETYPE	primary store's FoodAPS place type code
Household	PRIMSTORETYPE_ERS	ERS edited primary store type
Household	PRIMSTORETYPEREPORT	Reported store type of primary store, when not selected from drop-down list
Household	PRIMSTORETYPEREPORTSP	Specification of other reported store type for primary store
Household	PRIMSTORETRAVELOTH	Specified other means of getting to primary food store
Household	PRIMSTOREOTHREASONSP	Specified reason for using primary store - some cleaning done
Household	PRIMSTOREOTHREASONCODE	How primstoreothreasonsp was post-coded
Household	PRIMSTOREREASON_FLAG	Type of post-code for primstoreothreasoncode
Household	WHYNOTSUPERMKT	Why not use supermarket for most food shopping?

Dataset	Variable Name	Variable Label
Household	ALTSTORENAME	Name of household's alternate food store
Household	ALTSTORENAME_ERS	ERS edited alternate store place name
Household	ALTSTOREEDIT_SPACES	ERS removed spaces to edit place name
Household	ALTSTOREEDIT_CASE	ERS edited place name case
Household	ALTSTOREEDIT_NAME	ERS edited the place name
Household	ALTSTORESOURCE	Source of information and method of cleaning name/address of secondary store
Household	ALTSTORETYPE	Alternative store's FoodAPS place type code
Household	ALTSTORETYPE_ERS	ERS edited alternate store type
Household	ALTSTOREREASONSP	Specified other reason for shopping at alternate store
Household	SHOPANYOTHERSP	Specified other type of store for shopanyother
Individual	BOOKTYPE	Type of food booklet assigned to individual
Individual	BOOKTYPE_FLAG	Type of food booklet assigned imputed based on individual's age
Individual	GUESTTYPESP	Specifies how information in guesttype was determined
Individual	AGE	Individual's age in years
Individual	BIRTHMO	Month of birth
Individual	BIRTHYR	Year of birth
Individual	HISPGROUP	Reported Hispanic group (Spanish, Hispanic, or Latino)
Individual	RACEWHITE	Respondent reported individual is White
Individual	RACEBLACK	Respondent reported individual is Black/African American
Individual	RACEAIAKNAT	Respondent reported individual is American Indian or Alaska Native
Individual	RACEASIAN	Respondent reported individual is Asian
Individual	ASIANGROUP	Reported Asian group for individuals of reported to be Asian
Individual	RACEHIPI	Respondent reported individual is Native Hawaiian or Other Pacific Islander
Individual	RACEOTHER	Respondent reported individual is Other Race
Individual	RACEOTHERSP	Cleaned reported specification of race when reported “Other Race”
Individual	SPOUSEPNUM	Value of PNUM for respondent's spouse
Individual	WORKSTATUS	Individual's work status last week
Individual	WORKLASTWK	Individual worked at all last week at a job or business (Y/N)
Individual	JOBCHANGE	Individual changed jobs in the past 3 months (Y/N)
Individual	JOBCHANGEEARNINGS	Individual earns more, less, or same as before changing jobs
Individual	SCHLEVELSP	Specified other reason for not attending school
Individual	PREGNANT	Individual is currently pregnant (Y/N)
Individual	HEIGHT	Individual's reported height in inches

Dataset	Variable Name	Variable Label
Individual	WEIGHT	Individual's reported weight in pounds
Individual	BMIAGECAT	Individual's age category for determining body mass index (BMI) category
Individual	BIVHGT_FLAG	Biologically implausible value (BIV) for height
Individual	BIVWGT_FLAG	Biologically implausible value (BIV) for weight
Individual	BIVBMI_FLAG	Biologically implausible value (BIV) for BMI
Individual	OBTHRESHOLD	Weight threshold for being “obese” given individual's reported height
Individual	OBTHRESHOLDY	Individual weighs more or less than threshold for being “obese”
Individual	OWTHRESHOLD	Weight threshold for being “overweight” given individual's reported height
Individual	OWTHRESHOLDY	Individual weighs more or less than threshold for being “overweight”
Individual	INCEARNINDREPORTED	Individual's reported earnings last month w/o net versus gross adjustment, excluding imputed amounts
Individual	INCUNEMPINDREPORTED	Individual's reported unemployment insurance income last month, excluding imputed amounts
Individual	INCTRANSFERINDREPORTED	Individual's reported income last month from welfare, child support, and alimony payments, excluding imputed amounts
Individual	INCRETDISINDREPORTED	Individual's reported retirement and disability income last month, excluding imputed amounts
Individual	INCINVESTINDREPORTED	Individual's reported investment income last month, excluding imputed amounts
Individual	INCOTHERINDREPORTED	Individual's reported income last month from other sources
Individual	INCTOTINDAVG	Individual's total reported or avg imputed total gross income (over 5 imputations)
Individual	INCTOTINDAVG_FLAG	One or more individual's income components were imputed
Individual	INCTOTINDIMP1	Individual's reported or imputed total gross income, imputation 1
Individual	INCTOTINDIMP2	Individual's reported or imputed total gross income, imputation 2
Individual	INCTOTINDIMP3	Individual's reported or imputed total gross income, imputation 3
Individual	INCTOTINDIMP4	Individual's reported or imputed total gross income, imputation 4
Individual	INCTOTINDIMP5	Individual's reported or imputed total gross income, imputation 5
Individual	INCTOTINDREPORTED	Individual's reported total gross income from all sources, excluding imputed amounts
Individual	INCAMOUNT1	Amount, earnings from work, individual
Individual	INCFREQ1	Frequency, earnings from work, individual
Individual	INCTYPE1	Type (net or gross), earnings from work, individual
Individual	INCAMOUNT2	Amount, unemployment insurance, individual
Individual	INCFREQ2	Frequency, unemployment insurance, individual

Dataset	Variable Name	Variable Label
Individual	INCAMOUNT3	Amount, welfare, child support, alimony, individual
Individual	INCFREQ3	Frequency, welfare, child support, alimony, individual
Individual	INCAMOUNT4	Amount, retirement and disability income, individual
Individual	INCFREQ4	Frequency, retirement and disability income, individual
Individual	INC4TYPESP	Type(s) of retirement/disability income in incamount4
Individual	INCAMOUNT5	Amount, investment income, individual
Individual	INCFREQ5	Frequency, investment income, individual
Individual	INC5TYPESP	Type(s) of investment income received in incamount5
Individual	INCAMOUNT6	Amount, other income, individual
Individual	INCFREQ6	Frequency, other income, individual
Individual	INC6TYPESP	Type(s) of other income received in incamount6
Individual	INCCORRECT	Individual income reported so far is correct (Y/N)
Individual	INCMOREANY	Individual has more income to report (Y/N)
Individual	INCMOREAMT	Amount of additional monthly income reported
Individual	INCMOREEARN	Reported additional earnings or wages (Y/N)
Individual	INCMOREUNEMP	Reported additional unemployment insurance income (Y/N)
Individual	INCMORETRANSFER	Reported additional welfare, child support, and alimony income (Y/N)
Individual	INCMORERETDIS	Reported additional retirement and disability income (Y/N)
Individual	INCMOREINVEST	Reported additional investment income (Y/N)
Individual	HOMETENUREMO	Month in which individual moved to this address
Individual	HOMESINCEBIRTH	Individual has lived in current residence since birth (Y/N)
Individual	PRIORHOMETHISSTATE	Individual's previous home was in current State (Y/N)
Individual	PRIORHOMESTATE	Individual's previous State of residence
Individual	BORNTHISSTATE	Individual was born in current or previous State of residence (Y/N)
Individual	BIRTHSTATE	State where individual was born
Fahevent	PLACETYPE	Acquisition place type
Fahevent	PLACETYPE_ERS	ERS cleaned/standardized placetype
Fahevent	PLACEEDIT_FILLTYPE	ERS filled the missing place type
Fahevent	PLACENAME	Name of place where food was obtained
Fahevent	PLACENAME_ERS	ERS cleaned/standardized placename
Fahevent	SCANDATE	Date items were scanned by respondent
Fahitem	SCANDATE	Date items were scanned by respondent
Fafhevent	PLACETYPE	Recode–acquisition place type
Fafhevent	PLACETYPE_ERS	ERS edited place type
Fafhevent	PLACENAME	Name of place where food was obtained

Dataset	Variable Name	Variable Label
Fafhevent	PLACENAME_ERS	ERS edited place name
Fafhitem	MENUID	Unique id for top or menustat restaurant
Access	WALMART1	# SNAP-authorized Walmart stores within 0.25 mi of hh
Access	WALMART2	# SNAP-authorized Walmart stores within 0.5 mi of hh
Access	WALMART3	# SNAP-authorized Walmart stores within 1 mi of hh
Access	WALMART4	# SNAP-authorized Walmart stores within 2 mi of hh
Access	WALMART5	# SNAP-authorized Walmart stores within 5 mi of hh
Access	WALMART6	# SNAP-authorized Walmart stores within 10 mi of hh
Access	WALMART7	# SNAP-authorized Walmart stores within 15 mi of hh
Access	WALMART8	# SNAP-authorized Walmart stores within 30 mi of hh
Access	MCDONALDS1	# McDonald's restaurants within 0.25 mi of hh
Access	MCDONALDS2	# McDonald's restaurants within 0.5 mi of hh
Access	MCDONALDS3	# McDonald's restaurants within 1 mi of hh
Access	MCDONALDS4	# McDonald's restaurants within 2 mi of hh
Access	MCDONALDS5	# McDonald's restaurants within 5 mi of hh
Access	MCDONALDS6	# McDonald's restaurants within 10 mi of hh
Access	MCDONALDS7	# McDonald's restaurants within 15 mi of hh
Access	MCDONALDS8	# McDonald's restaurants within 30 mi of hh
Access	NEARFF_PLACEID	placeid of nearest fast-food restaurant
Access	NEARNONFF_PLACEID	placeid of nearest non-fast-food restaurant
Access	NEARMCD_PLACEID	placeid of nearest McDonald's restaurant
Feedback	FF4_7oth	FEEDBACK-Other Changes - Other Specify
Feedback	FLAG_FF4	FEEDBACK-FLAG: Postcode indicator for FF4_7oth
Snapelig_ind	SSICA	California household member receiving Supplemental Security Income (Y/N)
Snapelig_ind	DISABLED	Indicator of individual not working or not in school due to disability (Y/N)
Snapelig_ind	FTSTUDENT	Indicator of full-time, post-secondary student (Y/N)
Snapelig_ind	UNITID12	ID of estimated SNAP unit in model runs 1 and 2
Snapelig_ind	UNITID34	ID of estimated SNAP unit in model runs 3 and 4
Snapelig_unit	HHNUM	6-digit unique identifier for each household
Snapelig_unit	MODELRUN	Indicates whether unit created in models 1 and 2 or 3 and 4
Snapelig_unit	UNITID12	ID of estimated SNAP unit in model runs 1 and 2
Snapelig_unit	UNITID34	ID of estimated SNAP unit in model runs 3 and 4
Snapelig_unit	UNITSIZE12	Number of people in estimated SNAP unit created in model runs 1 and 2
Snapelig_unit	UNITSIZE34	Number of people in estimated SNAP unit created in model runs 3 and 4

Dataset	Variable Name	Variable Label
Snapelig_unit	DISAB12	Number of disabled persons in unit in model runs 1 and 2
Snapelig_unit	DISAB34	Number of disabled persons in unit in model runs 3 and 4
Snapelig_unit	ELDERLY12	Number of persons age 60 or over in unit in model runs 1 and 2
Snapelig_unit	ELDERLY34	Number of persons age 60 or over in unit in model runs 3 and 4
Snapelig_unit	NONCIT12	Number of non-U.S. citizens in unit in model runs 1 and 2
Snapelig_unit	NONCIT34	Number of non-U.S. citizens in unit in model runs 3 and 4
Snapelig_unit	EARN1	Total monthly earnings for unit in model run 1
Snapelig_unit	EARN2	Total monthly earnings for unit in model run 2
Snapelig_unit	EARN3	Total monthly earnings for unit in model run 3
Snapelig_unit	EARN4	Total monthly earnings for unit in model run 4
Snapelig_unit	SOCSEC12	Total monthly Social Security income for unit in model runs 1 and 2
Snapelig_unit	SOCSEC34	Total monthly Social Security income for unit in model runs 3 and 4
Snapelig_unit	SSI12	Total monthly Supplemental Security Income (SSI) for unit in model runs 1 and 2
Snapelig_unit	SSI34	Total monthly SSI for unit in model runs 3 and 4
Snapelig_unit	TANF12	Total monthly Temporary Assistance for Needy Families (TANF) income for unit in model runs 1 and 2
Snapelig_unit	TANF34	Total monthly TANF income for unit in model runs 3 and 4
Snapelig_unit	GROSSINC1	Total monthly gross SNAP income for unit in model run 1
Snapelig_unit	GROSSINC2	Total monthly gross SNAP income for unit in model run 2
Snapelig_unit	GROSSINC3	Total monthly gross SNAP income for unit in model run 3
Snapelig_unit	GROSSINC4	Total monthly gross SNAP income for unit in model run 4
Snapelig_unit	NETINC1	Total monthly net income for unit in model run 1
Snapelig_unit	NETINC2	Total monthly net income for unit in model run 2
Snapelig_unit	NETINC3	Total monthly net income for unit in model run 3
Snapelig_unit	NETINC4	Total monthly net income for unit in model run 4
Snapelig_unit	POVLINE12	Monthly poverty guideline for unit size and composition, model runs 1 and 2
Snapelig_unit	POVLINE34	Monthly poverty guideline for unit size and composition, model runs 3 and 4
Snapelig_unit	POVRATIO1	Total monthly gross SNAP income for unit divided by poverty line in model run 1
Snapelig_unit	POVRATIO2	Total monthly gross SNAP income for unit divided by poverty line in model run 2
Snapelig_unit	POVRATIO3	Total monthly gross SNAP income for unit divided by poverty line in model run 3
Snapelig_unit	POVRATIO4	Total monthly gross SNAP income for unit divided by poverty line in model run 4
Snapelig_unit	GROSSTEST1	Unit passed the gross income test in model run 1 (Y/N)

Dataset	Variable Name	Variable Label
Snapelig_unit	GROSSTEST2	Unit passed the gross income test in model run 2 (Y/N)
Snapelig_unit	GROSSTEST3	Unit passed the gross income test in model run 3 (Y/N)
Snapelig_unit	GROSSTEST4	Unit passed the gross income test in model run 4 (Y/N)
Snapelig_unit	NETTEST1	Unit passed the net income test in model run 1 (Y/N)
Snapelig_unit	NETTEST2	Unit passed the net income test in model run 2 (Y/N)
Snapelig_unit	NETTEST3	Unit passed the net income test in model run 3 (Y/N)
Snapelig_unit	NETTEST4	Unit passed the net income test in model run 4 (Y/N)
Snapelig_unit	ASSETTEST12	Unit passed the SNAP asset test in model runs 1 and 2 (Y/N)
Snapelig_unit	ASSETTEST34	Unit passed the SNAP asset test in model runs 3 and 4 (Y/N)
Snapelig_unit	PUREPA12	Unit is pure PA and SNAP-eligible in model runs 1 and 2 (Y/N)
Snapelig_unit	PUREPA34	Unit is pure PA and SNAP-eligible in model runs 3 and 4 (Y/N)
Snapelig_unit	CATELIG1	Unit is SNAP-eligible under BBCE in model run 1 (Y/N)
Snapelig_unit	CATELIG2	Unit is SNAP-eligible under BBCE in model run 2 (Y/N)
Snapelig_unit	CATELIG3	Unit is SNAP-eligible under BBCE in model run 3 (Y/N)
Snapelig_unit	CATELIG4	Unit is SNAP-eligible under BBCE in model run 4 (Y/N)
Snapelig_unit	ZEROBEN1	Unit is not SNAP-eligible in run 1 only because estimated benefit amount is zero
Snapelig_unit	ZEROBEN2	Unit is not SNAP-eligible in run 2 only because estimated benefit amount is zero
Snapelig_unit	ZEROBEN3	Unit is not SNAP-eligible in run 3 only because estimated benefit amount is zero
Snapelig_unit	ZEROBEN4	Unit is not SNAP-eligible in run 4 only because estimated benefit amount is zero
Snapelig_unit	ELIG1	Unit is SNAP-eligible in model run 1 (Y/N)
Snapelig_unit	ELIG2	Unit is SNAP-eligible in model run 2 (Y/N)
Snapelig_unit	ELIG3	Unit is SNAP-eligible in model run 3 (Y/N)
Snapelig_unit	ELIG4	Unit is SNAP-eligible in model run 4 (Y/N)
Snapelig_unit	BENEST1	Estimated monthly SNAP benefit for unit in model run 1
Snapelig_unit	BENEST2	Estimated monthly SNAP benefit for unit in model run 2
Snapelig_unit	BENEST3	Estimated monthly SNAP benefit for unit in model run 3
Snapelig_unit	BENEST4	Estimated monthly SNAP benefit for unit in model run 4
Snapelig_hh	SNAPADMINAMT	Amount of SNAP benefits on administrative file record
Snapelig_hh	COUNT_MATCHES	# current-month matches between household (HH) and State's SNAP admin file
Snapelig_hh	SUM_SNAPAMTS	Sum of SNAPADMINAMT from all matches to admin data for HH
Snapelig_hh	SNAPUNIT_HH	Household contains at least one SNAP unit (Y/N)
Snapelig_hh	SOCSEC_HH	Indicator of household receipt of Social Security income (Y/N)
Snapelig_hh	SSI_HH	Indicator of household receipt of Supplemental Security Income (Y/N)

Dataset	Variable Name	Variable Label
Snapelig_hh	TRANSFER_HH	Indicator of household receipt of welfare assistance (Y/N)
Snapelig_hh	ASSETCAT	Categorical measure of household's total liquid assets
Snapelig_hh	ELIGSTRING	Summary of eligibility estimated in runs 1 through 4

Table C2. Variables coarsened (recoded, top-coded, or derived) on PUFs

Dataset	Original Variable	New Variable	New Variable Definition
Household	INITIALDATE	INITINTRVMON	Month of the initial interview
Household	STARTDATE	STARTMON	Month of start date
Household	FINALDATE	INITFINALDAYS	Elapsed days between initial and final interviews
Household	GUESTS	NUMGUESTS	Combine categories into none, one, and more than one
Household	LODGERS	GUESTSPAY	Yes/No indicating if any boarders or lodgers
Household	BOARDERS		
Household	HHSIZEBIRTH	HHSIZECHILD	Combine hhsizebirth and hhsizeothchild into one indicator Yes/No
Household	HHSIZEOTHCHILD		
Household	HHSIZESEPARATION	HHSIZECHANGEOTH	Combine hhsizeseparation, hhsizedeath, hhsizemarriage, hhsizepartner, hhsizeoth into one indicator Yes/No
Household	HHSIZEDEATH		
Household	HHSIZEMARRIAGE		
Household	HHSIZEPARTNER		
Household	HHSIZEOTH		
Household	INCFAMREPORTED	INCFAMREPORTED_R	Top-code at 99th percentile
Household	INCFAMAVG	INCFAMAVG_R	Top-code at 99th percentile
Household	INCHHREPORTED	INCHHREPORTED_R	Top-code at 99th percentile
Household	INCHHAVG	INCHHAVG_R	Top-code at 99th percentile
Household	INCHHIMP1–INCHHIMP5	INCHHIMP1_R–INCHHIMP5_R	Top-code at 99th percentile
Household	PCTPOVGUIDEHH	PCTPOVGUIDEHH_R	Top-code at 99th percentile
Household	JOBCHANGENUM	JOBCHANGECAT	Combine categories into 1, and more than 1
Household	EARNMORENUM	EARNMORENUM_R	Yes/No indicator
Household	EARNLESSNUM	EARNLESSNUM_R	Yes/No indicator
Household	LIQASSETS2000	LIQASSETS	Combine liqassets2000 and liqassets3000, categorize as: < $2000, $2000 to < $3000, at least $2000, >= $3000
Household	LIQASSETS3000		
Household	AUTO	ANYVEHICLE	Whether anybody in household owns or leases a vehicle (Yes/No)
Household	AUTONUM	VEHICLENUM	Top-code at 4
Household	EXPRENTMRTG	EXPRENTMRTG_R	Top-code at 99th percentile
Household	EXPHOMEINS	EXPHOMEINS_R	Top-code at 99th percentile
Household	EXPPROPTAX	EXPPROPTAX_R	Top-code at 99th percentile
Household	EXPPUBTRANS	EXPPUBTRANS_R	Top-code at 99th percentile
Household	EXPELECTRIC	EXPELECTRIC_R	Top-code at 99th percentile
Household	EXPHEATFUEL	EXPHEATFUEL_R	Top-code at 99th percentile
Household	EXPWASTEDISP	EXPWASTEDISP_R	Top-code at 99th percentile
Household	EXPHEALTHINS	EXPHEALTHINS_R	Top-code at 99th percentile
Household	EXPCOPAY	EXPCOPAY_R	Top-code at 99th percentile
Household	EXPDOCTOR	EXPDOCTOR_R	Top-code at 99th percentile
Household	EXPRX	EXPRX_R	Top-code at 99th percentile
Household	EXPCHILDCARE	EXPCHILDCARE_R	Top-code at 99th percentile
Household	EXPCHILDSUPPORT	EXPCHILDSUPPORT_R	Top-code at 99th percentile
Household	EXPADULTCARE	EXPADULTCARE_R	Top-code at 99th percentile

Dataset	Original Variable	New Variable	New Variable Definition
Household	SNAPLASTAMT	SNAPLASTAMT_R	1="0<amount<=16", 2="16<amount<50", 3 - 17 are categories each with a range of amount 50, 18="800<=amount<1200"
Household	PRIMSTORETYPE	PRIMSTORETYPE	Use the edited values in primstoretype_ers to update primstoretype
Household	PRIMSTORETYPE_ERS		
Household	ALTSTORETYPE	ALTSTORETYPE	Use the edited values in altstoretype_ers to update altstoretype
Household	ALTSTORETYPE_ERS		
Household	ILLNESSWHO	ANYILLNESS	Yes/No indicator
Household	NGUESTBRKFSTSUN	NGUESTBRKFSTSUN_R	Top-code at 99th percentile
Household	NGUESTBRKFSTMON	NGUESTBRKFSTMON_R	Top-code at 99th percentile
Household	NGUESTBRKFSTTUE	NGUESTBRKFSTTUE_R	Top-code at 99th percentile
Household	NGUESTBRKFSTWED	NGUESTBRKFSTWED_R	Top-code at 99th percentile
Household	NGUESTBRKFSTTHU	NGUESTBRKFSTTHU_R	Top-code at 99th percentile
Household	NGUESTBRKFSTFRI	NGUESTBRKFSTFRI_R	Top-code at 99th percentile
Household	NGUESTBRKFSTSAT	NGUESTBRKFSTSAT_R	Top-code at 99th percentile
Household	NGUESTLUNCHSUN	NGUESTLUNCHSUN_R	Top-code at 99th percentile
Household	NGUESTLUNCHMON	NGUESTLUNCHMON_R	Top-code at 99th percentile
Household	NGUESTLUNCHTUE	NGUESTLUNCHTUE_R	Top-code at 99th percentile
Household	NGUESTLUNCHWED	NGUESTLUNCHWED_R	Top-code at 99th percentile
Household	NGUESTLUNCHTHU	NGUESTLUNCHTHU_R	Top-code at 99th percentile
Household	NGUESTLUNCHFRI	NGUESTLUNCHFRI_R	Top-code at 99th percentile
Household	NGUESTLUNCHSAT	NGUESTLUNCHSAT_R	Top-code at 99th percentile
Household	NGUESTDINNERSUN	NGUESTDINNERSUN_R	Top-code at 99th percentile
Household	NGUESTDINNERMON	NGUESTDINNERMON_R	Top-code at 99th percentile
Household	NGUESTDINNERTUE	NGUESTDINNERTUE_R	Top-code at 99th percentile
Household	NGUESTDINNERWED	NGUESTDINNERWED_R	Top-code at 99th percentile
Household	NGUESTDINNERTHU	NGUESTDINNERTHU_R	Top-code at 99th percentile
Household	NGUESTDINNERFRI	NGUESTDINNERFRI_R	Top-code at 99th percentile
Household	NGUESTDINNERSAT	NGUESTDINNERSAT_R	Top-code at 99th percentile
Household	NGUESTSNACKSUN	NGUESTSNACKSUN_R	Top-code at 99th percentile
Household	NGUESTSNACKMON	NGUESTSNACKMON_R	Top-code at 99th percentile
Household	NGUESTSNACKTUE	NGUESTSNACKTUE_R	Top-code at 99th percentile
Household	NGUESTSNACKWED	NGUESTSNACKWED_R	Top-code at 99th percentile
Household	NGUESTSNACKTHU	NGUESTSNACKTHU_R	Top-code at 99th percentile
Household	NGUESTSNACKFRI	NGUESTSNACKFRI_R	Top-code at 99th percentile
Household	NGUESTSNACKSAT	NGUESTSNACKSAT_R	Top-code at 99th percentile
Individual	INITIALDATE	INITINTRVMON	Month of the initial interview
Individual	FINALDATE	INITFINALDAYS	Elapsed days between initial and final interviews
Individual	GUESTTYPE	GUEST	Indicator variable GUEST = 1 if GUESTTYPE in (1, 2) and 0 otherwise

Dataset	Original Variable	New Variable	New Variable Definition
Individual	RELATION	RELATION_R	Combine categories into: Respondent, Spouse/Unmarried partner, Child/Stepchild/Adopted child, Grandchild, Parent, Brother/Sister, Other members
Individual	AGEGROUP	AGE_R	Categorized age: 0, 1, 2, 3, 4, 5, 6-7, 8-11, 12-15, 16-17, 18, 19, 20-35, 36-59, 60-65, 66-70, >=71, Adult >18 years (imputed)
Individual	AGEGROUP_FLAG	AGE_R_FLAG	Rename to AGE_R_FLAG
Individual	RACECAT	RACECAT_R	White; Black; American Indian or Alaska Native; Asian or Native Hawaiian or Other Pacific Islander; other race; multiple race
Individual	EDUC	EDUCCAT	< 1st to 10th; 11th and 12th w/ no diploma; high school grad or GED; some college or assoc degree; bachelor's degree; master's or higher
Individual	WORKSTATUS	EMPLOYMENT	Combine WORKSTATUS and WORKLASTWK: Working at a job or business; With a job or business but not at work; Looking for work, with no work last week; Looking for work, with some work last week; Not working at a job or business
Individual	WORKLASTWK		
Individual	WORKCOMMUTETIME	WORKCOMMUTETIME_R	Top-code at 99th percentile
Individual	SCHLEVEL	SCHLEVEL_R	Kindergarten; Elementary school; Primary school; Middle school or Junior high; High school; Some other school; On school break; Summer vacation; Home-schooled; Dropped out, disabled, or other reason not in school; Not old enough
Individual	CHILDCARE	CHILDCARE_R	Yes/No indicator
Individual	FOODALLERGYWHEAT	ALLERGYWHEAT (HOUSEHOLD FILE)	Yes/No indicator, combining the data from all household members
Individual	FOODALLERGYMILK	ALLERGYMILK (HOUSEHOLD FILE)	Yes/No indicator, combining the data from all household members
Individual	FOODALLERGYEGG	ALLERGYEGG (HOUSEHOLD FILE)	Yes/No indicator, combining the data from all household members
Individual	FOODALLERGYFISH	ALLERGYFISH (HOUSEHOLD FILE)	Yes/No indicator, combining the data from all household members
Individual	FOODALLERGYSHELLFISH	ALLERGYSHELLFISH (HOUSEHOLD FILE)	Yes/No indicator, combining the data from all household members
Individual	FOODALLERGYCORN	ALLERGYCORN (HOUSEHOLD FILE)	Yes/No indicator, combining the data from all household members
Individual	FOODALLERGYPEANUT	ALLERGYPEANUT (HOUSEHOLD FILE)	Yes/No indicator, combining the data from all household members

Dataset	Original Variable	New Variable	New Variable Definition
Individual	FOODALLERGYOTHNUT	ALLERGYOTHNUT (HOUSEHOLD FILE)	Yes/No indicator, combining the data from all household members
Individual	FOODALLERGYSOY	ALLERGYSOY (HOUSEHOLD FILE)	Yes/No indicator, combining the data from all household members
Individual	FOODALLERGYOTH	ALLERGYOTH (HOUSEHOLD FILE)	Yes/No indicator, combining the data from all household members
Individual	BMI	BMI	Keep 2 decimal places
Individual	BMIPCT	BMIPCT	Keep 2 decimal places
Individual	INCEARNINDAVG	INCEARNINDAVG_R	Yes/No indicator
Individual	INCUNEMPINDAVG	INCUNEMPINDAVG_R	Yes/No indicator
Individual	INCTRANSFERINDAVG	INCTRANSFERINDAVG_R	Yes/No indicator
Individual	INCRETDISINDAVG	INCRETDISINDAVG_R	Yes/No indicator
Individual	INCINVESTINDAVG	INCINVESTINDAVG_R	Yes/No indicator
Individual	INCOTHERINDAVG	INCOTHERINDAVG_R	Yes/No indicator
Individual	HOMETENUREYR	YEARSRESIDENCE	Pre-1993, 1993-1996, 1997-2001, 2002-2005, 2006-2009, 2010-2012
Individual	HOMESINCEBIRTH	USBORN	Combine HOMESINCEBIRTH and BORNTHISSTATE into USBORN (an indicator variable)
Individual	BORNTHISSTATE		
Fahevent	PLACETYPE	PLACETYPE	Use placetype_ers to fill the missing value of placetype
Fahevent	PLACETYPE_ERS		
Fahevent	STARTDATE	STARTMON	Month of startdate
Fafhevent	PLACETYPE	PLACETYPE	Use placetype_ers to fill the missing value of placetype
Fafhevent	PLACETYPE_ERS		
Fafhevent	PLACENAME	CHAIN	Use placename_ers to fill in missing value of placename, then create flag for top 30 fast-food restaurants and top 30 casual dining restaurants (Yes/No)
Fafhevent	PLACENAME_ERS		
Fafhevent	STARTDATE	STARTMON	Month of startdate
Meals	STARTDATE	STARTMON	Month of startdate

Appendix D – 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.

D.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 D1. 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

A text file (jkn_factors.txt) is also available that contains the factors needed for computing variances with the stratified jackknife replicates.

D.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  0.5
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
0.875  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 ADLTFSRAW;
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

                                Std Error
              Variable            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                                Std Error
  tract   Variable       Mean          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
 ---------------------------------------------------------------------------
```

D.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 0.66666667 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 ///
          0.875  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 adltfsraw
svy: mean  adltfsraw, 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 that 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
..................................................    50
.......

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]
--------------+-----------------------------------------------
adltfsraw     |    1.00935   .0307719      .9460974    1.072602
--------------------------------------------------------------


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

```
Jackknife replications (57)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
.......

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

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

D.1.3 R

The example R code follows:

```
require(survey)

# replicate weights

design.rw <- svrepdesign(repweights = "hhwgt[+0-9]",
   weights = household$hhwgt, 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(~adltfsraw, design=design.rw)
svyby(~adltfsraw, ~rural, design=design.rw, svymean)
rural.rw <- subset(design.rw, rural==1)
svymean(~adltfsraw, 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 dataset first, then apply subset() to extract the subdomain of interest.

The results are:

```
> design.rw <- svrepdesign(repweights = "hhwgt[+0-9]",
+    weights = household$hhwgt, 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(~adltfsraw, design=design.rw)
                  mean     SE
adltfsraw        1.0093 0.0308
> svyby(~adltfsraw, ~rural, design=design.rw, svymean)
    rural  adltfsraw           se
0     0     1.1230677   0.04306089
1     1     0.7865201   0.06992609
> rural.rw <- subset(design.rw, rural==1)
> svymean(~adltfsraw, design=rural.rw)
                 mean     SE
adltfsraw        0.78652 0.0699
>
```

D.2 Taylor Series Linearization

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

D.2.1 SAS

SAS users can use this syntax:

```
proc surveymeans ;
cluster tspsu ;
strata tsstrata ;
var 'varlist' ;
weight hhwgt ;
*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;”).

D.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=hhwgt], 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'
```

D.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=~tspsu, 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 E – 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 PSUs 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, nine of the sampled PSUs comprised five or more counties. They were subdivided into two or three 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 followup.

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 one in four. 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.[17] 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.

[17] 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 F – Summary of Revisions for RUFs

November 25, 2014

- Meals and snacks 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 website version

September 1, 2016

- Revised chapter 4 to include the descriptions about RUFs and PUFs, as well as the disclosure control treatments in the PUFs
- Added appendix C to provide the list of variables suppressed or coarsened in the PUFs.
- The original appendices C, D, and E were renumbered to D, E, and F.

September 2, 2016

- A revision to the faps_household restricted-use data file was made and posted; appendix A was updated accordingly