United States Department of Agriculture

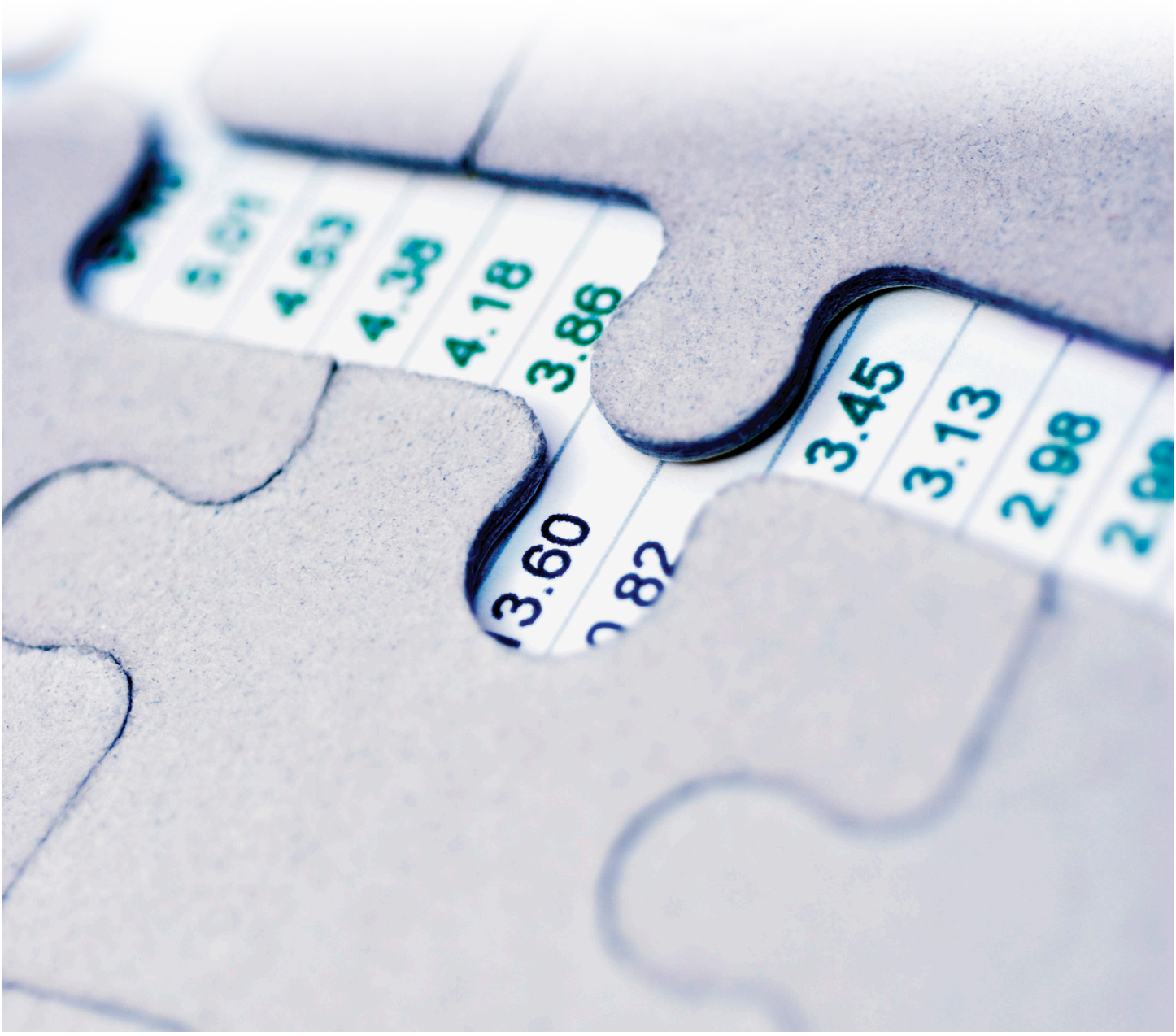
Economic Research Service

Technical Bulletin Number 1947

January 2018

USDA's National Household Food Acquisition and Purchase Survey: Methodology for Imputing Missing Quantities To Calculate Healthy Eating Index-2010 Scores and Sort Foods Into ERS Food Groups

Lisa Mancino, Jessica E. Todd, and Benjamin Scharadin



United States Department of Agriculture

Economic Research Service
www.ers.usda.gov

Recommended citation format for this publication:

Mancino, Lisa, Jessica E. Todd, and Benjamin Scharadin. *USDA's National Household Food Acquisition and Purchase Survey: Methodology for Imputing Missing Quantities To Calculate Healthy Eating Index-2010 Scores and Sort Foods Into ERS Food Groups,* TB-1947, U.S. Department of Agriculture, Economic Research Service, January 2018.

United States Department of Agriculture

Economic Research Service

Technical Bulletin Number 1947

January 2018

USDA's National Household Food Acquisition and Purchase Survey: Methodology for Imputing Missing Quantities To Calculate Healthy Eating Index-2010 Scores and Sort Foods Into ERS Food Groups

Lisa Mancino, Jessica E. Todd, and Benjamin Scharadin

Abstract

We use data from the U.S. Dept. of Agriculture's 2012-13 National Household Food Acquisition and Purchase Survey (FoodAPS) to develop a methodology for imputing missing quantities for item-level data. To check the quality of these imputations, we compare the overall nutritional quality of 7 days of household-level acquisitions reported in the FoodAPS survey to 2 days of individual-level dietary intakes as reported in the 2011-12 National Health and Nutritional Examination Survey (NHANES). The comparison shows that while there are some slight differences in the Healthy Eating Index-2010 (HEI-2010) component densities, both surveys show that Americans acquire too few fruits, vegetables, and whole grains and obtain too much salt and too many refined grains and empty calories. We also describe the method used to classify food items into the ERS food groups developed for FoodAPS.

Keywords: FoodAPS, missing quantities, Healthy Eating Index, ERS food groups

Acknowledgments

The authors thank peer reviewers Lorenzo Nicolas Almada (Georgia State University), Shu Wen Ng (University of North Carolina), Young Jo (USDA, Economic Research Service (ERS)), and TusaRebecca Pannucci (USDA, Center for Nutrition Policy and Promotion) for their comments. Jean Buzby and Constance Newman (ERS) provided helpful guidance and review. Thanks also to Courtney Knauth and Curtia Taylor of ERS for editing and design assistance.

Contents

United States Department of Agriculture

A report summary from the Economic Research Service *January 2018*

USDA's National Household Food Acquisition and Purchase Survey: Methodology for Imputing Missing Quantities To Calculate Healthy Eating Index-2010 Scores and Sort Foods Into ERS Food Groups

Lisa Mancino, Jessica E. Todd, and Benjamin Scharadin

What Is the Issue?

USDA's National Household Food Purchase and Acquisition Survey (FoodAPS) collected detailed information about the types of food households acquire over the course of 1 week, between April 2012 and January 2013. One objective of the survey was to evaluate the nutritional quality of household food acquisitions as measured against the 2010 Healthy Eating Index (HEI-2010) and whether economic and sociodemographic factors affect the nutritional quality of these acquisitions. Total grams or calories can be used to normalize expenditures for comparing food quality across households or venues; however, complete quantity information is needed for research questions not related to nutritional quality. Unfortunately, information on the quantities acquired is missing for some items. Rather than have individual researchers grapple with how to handle these missing quantities, ERS has added imputed quantities to the FoodAPS data. Also, the ability to study specific categories of food is useful for research on policies that affect those categories, such as subsidies for fruits and vegetables or taxes on certain beverages. Therefore, ERS developed a food classification scheme for the data that groups foods in relation to their main ingredient, quality, and likely price premiums for convenience and processing. Researchers now have more than one choice when aggregating the FoodAPS items into groups for analysis.

What Did the Study Find?

The HEI-2010 is calculated by considering the proportions in which foods are consumed or are included in a shopping basket or pantry. Thus, knowing the quantity of each item acquired is essential. When quantity information is missing, the researcher must either drop the item from HEI-2010 calculations or impute a quantity so it can be included. ERS researchers developed a methodology to impute missing quantities based on all available information about each item, including the item description, the type of store where it was purchased, the geographic location of the purchasing household, the amount paid for the item (when available), and the size of the household. A dataset with imputed quantities has been posted to the FoodAPS public use files on the ERS website. This allows researchers to include all FoodAPS items with nutrient information in calculations of HEI-2010 scores and component densities and will facilitate comparisons across various studies.

We compute and compare household-level HEI-2010 scores over the 7 days of FoodAPS to individual-level HEI-2010 scores from 2 days of dietary recall data from the 2011-12 National Health and Nutrition Examination Survey (NHANES). In both cases, we use the simple algorithm to generate household-level (or respondent-level, in the case of NHANES) HEI-2010 component densities and then estimate means. The comparison of means indicates that while there are some slight differences in HEI-2010 component densities, both surveys show that Americans acquire too few fruits, vegetables, and whole grains, too many refined grains and empty calories, and too much salt. The Stata programs used to calculate HEI-2010 scores for FoodAPS acquisitions and data with imputed quantities are provided to users on the ERS website.

Also in this study, ERS developed a food classification system for FoodAPS, based on the 2010 *Dietary Guidelines for Americans* as well as on other aspects of food items, like the convenience and processing that consumers consider when purchasing food. This classification system differs from others, such as the USDA's *What We Eat in America* food categories, because it allows data users to separate foods not only by main ingredient and nutritional characteristics, but also by level of convenience and form, such as fresh, canned, or frozen. Each food item was classified into 1 of 82 ERS food groups using all available information about the item.

How Was the Study Conducted?

Information about food items, the stores from which they were obtained, and household characteristics available in the FoodAPS data were used to impute quantities when missing. The SAS programs developed by the U.S. Department of Health and Human Services, National Cancer Institute for HEI-2010 scores were adapted to FoodAPS using Stata. Detailed item-level information was used to classify foods into 82 ERS food groups. Stata 12.2 was used for all quantity imputations. Sorting food items into food groups was performed using a series of programs in both SAS 9.0 and Stata.

USDA's National Household Food Acquisition and Purchase Survey: Methodology for Imputing Missing Quantities To Calculate Healthy Eating Index-2010 Scores and Sort Foods Into ERS Food Groups

Introduction

To examine the ways in which economic and demographic factors influence the nutritional quality of the American diet, USDA's Economic Research Service (ERS) and Food and Nutrition Service (FNS) cosponsored the *National Household Food Acquisition and Purchase Survey (FoodAPS).*[1] The FoodAPS data are comprehensive, and researchers can readily examine a number of economic and sociodemographic characteristics, such as the relationship among Supplemental Nutrition Assistance Program (SNAP) participation, food insecurity, and household composition.

Assessing the nutritional quality of household food acquisitions requires additional steps, however. First, there needs to be a standard way to define nutritional quality. For this, we use the *Dietary Guidelines for Americans* (USDA and HHS, 2010), which forms the basis for all Federal nutrition guidance, and the companion 2010 Healthy Eating Index (HEI-2010) to measure alignment with the 2010 *Dietary Guidelines* (Guenther et al., 2013). Two overarching themes of the guidelines are to maintain a calorie balance over time and to focus on choosing nutrient-dense foods and beverages. Thus, the HEI-2010 score uses a density approach to set standards, such as servings per 1,000 calories or as a percentage of calories. The total HEI-2010 score ranges from 0 to 100. The scores increase with relative increases in dietary constituents that are encouraged, such as total fruit, total vegetables, and whole grains, and decrease with relative increases in dietary constituents that are recommended in moderation, such as added sugars, sodium, and extra calories. This means that to accurately calculate HEI-2010 scores in FoodAPS, dependable information is needed on the dietary constituents in each food item and the total amount of the item acquired. While total grams or calories can be used to normalize overall expenditures for comparison across households or venues, quantity information is useful for research questions not related to nutritional quality.

Unfortunately, information is missing on quantities for a number of items. This report describes ERS's extensions to the FoodAPS data to include imputed quantities, rather than leaving individual researchers to grapple with how to handle these missing amounts. Posting the imputed quantities to the public-use files allows all users to calculate HEI-2010 scores in a consistent manner and facilitates comparison of research results across multiple studies.

In addition to analyzing overall nutritional quality, researchers may also wish to study the intermediate steps that consumers take—selecting which combination of foods and which forms for the items—to make up their diets. Again, there are multiple ways that researchers could choose to group food items. Each FoodAPS item was linked to a USDA What We Eat In America (WWEIA) food category, which groups food items by the main ingredient. In the WWEIA categories, mixed dishes are grouped by primary ingredients, and there is no distinction given to the form in which foods are originally obtained. We developed another food classification scheme for food items in FoodAPS. The report describes this food group classification scheme, which provides researchers an additional way in which to aggregate foods for analysis. Researchers can also use this classification scheme to identify particular subsets of foods that they may wish to disaggregate in their analyses.

[1] For more information on these data, see https://www.ers.usda.gov/foodaps.

Missing Quantity Imputations for HEI-2010

ERS chose the 2010 Healthy Eating Index (HEI-2010) as the main measure of nutritional quality with which to evaluate the foods reported in FoodAPS. Developed by the USDA's Center for Nutrition Policy and Promotion and the National Cancer Institute (NCI), the HEI-2010 summarizes how well a set of foods conforms to the *Dietary Guidelines for Americans, 2010* (USDA and HHS, 2010; Guenther et al., 2013). Although originally developed to evaluate the nutritional quality of food consumed as part of a 24-hour dietary recall, the HEI-2010 measure is also appropriate for assessing the quality of any set of foods, including the national food supply, a community's food environment, store purchases, and menu offerings (Schap, Kuczynski, and Hiza, 2017). As a research tool, the HEI has been used to assess the diet quality of the U.S. population and subpopulations, evaluate the effect of nutrition interventions, and assess various aspects of the food environment (Guenther et al., 2014).

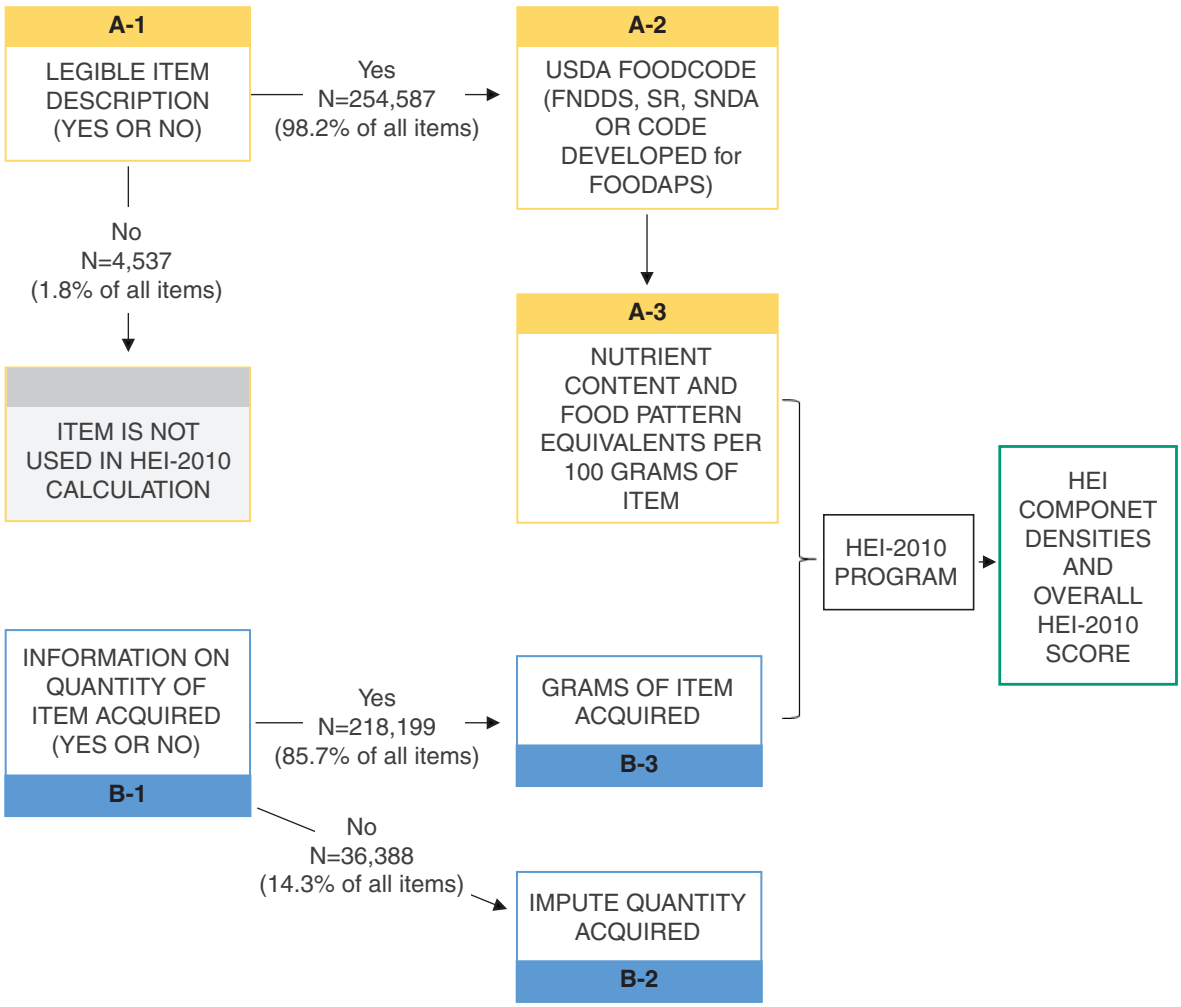
HEI-2010 scores range from 0 to 100 and incorporate 12 components, 9 of which are adequacy components (e.g., whole fruit, whole grains, and total vegetables) and the remaining three are moderation components (empty calories, sodium, and refined grains). To account for differences in individual total caloric needs, components are measured using a density approach to set standards, such as per 1,000 calories or as a percentage of calories. In order to construct an HEI-2010 score for a set of foods, information from the Food Pattern Equivalent Database (FPED) (USDA, ARS, 2016) for each item must be known or estimated. The FPED contains either cup or ounce equivalents of specific food groups such as fruit, whole grains, and dairy or other dietary components such as added sugars, tracked in the HEI-2010 score. Nutritionists at the USDA have calculated FPED values for foods that were reported consumed in the National Health and Nutrition Examination Survey (NHANES) (Bowman et al., 2014). Each USDA food code also has information on other nutrients, independent of FPED values, such as total energy, total fat, saturated fat, and sodium.

Food items reported in FoodAPS were linked to nutrient and FPED values through a number of different processes, depending on how the item was reported by respondents and where it was reported to have come from. The *FoodAPS Nutrient Coding Overview*, which is posted on the ERS FoodAPS website, outlines the processes used for items from stores (food at home, FAH), and at eating places and other places (food away from home, FAFH). In general, items were matched to a USDA nutrient food code, either from the USDA Food and Nutrient Database for Dietary Studies (FNDDS), or from the USDA National Nutrient Database for Standard Reference (SR). For simplicity, we refer to these as FNDDS/SR food codes (USDA, ARS, 2016). Items were matched to FNDDS/SR food codes using the item description and information about the form in which it was purchased (e.g., prepared, canned, or dry/unprepared). For example, a snack-pack of tapioca pudding (a food item) would be matched to food code "13230500, pudding, ready-to-eat, tapioca." Limiting factors in this process included many incomplete or nonspecific item descriptions, a lack of food codes with FPED values for dry or unprepared forms of foods, and a lack of food codes for prepared foods in their purchased form (e.g., a salad with dressing, a combination platter). In total, there were 259,124 food items included in the FoodAPS survey. Unfortunately, some item descriptions were too incomplete to be identified and it was not possible to assign them food codes or FPED values. Of the 259,124 total items, 254,587, or roughly 98 percent, were assigned food codes.

Once an item was assigned a food code, nutrient and FPED information were appended to that item (fig. 1). However, all the FNDDS/SR nutrient and FPED data provide information for 100 grams of

that food code, not the actual amount acquired. To accurately calculate the HEI scores and component densities, we need to know the amount acquired. While the FoodAPS survey requested that respondents give the quantity for all foods they reported, there were still a non-trivial number of items, primarily in the food-at-home (FAH) item data, that were missing quantities. Among the 254,587 items with FPED values, 218,199 (almost 86 percent) had information on the quantity acquired.[2] ERS imputed quantities for the remaining 36,388 (14 percent) of items with nutrient and FPED information (fig. 1).

Figure 1
Information required to calculate Healthy Eating Index-2010 scores and component densities



Source: USDA, Economic Research Service.

[2] There were considerably fewer items in the food-away-from-home (FAFH) files with missing quantities. The contractor for the FoodAPS data collection obtained or imputed quantity information for all FAFH items that were assigned a foodcode during their processing of the data. However, during post processing, we were able to identify foodcodes based on item descriptions for 188 items, which resulted in 188 FAFH items with missing quantities.

For some insights into the nature of missing quantities, we first compare the distribution of items with missing and nonmissing quantities across broad place categories—big grocery stores, other grocery stores, all other FAH stores, own production, and other assistance (table 1a). Within FAH outlets, 75 percent of items acquired at big grocery stores had reported quantities. In comparison, 83 percent of items from all other FAH places, such as convenience stores, gas stations, and dollar stores, had items with reported quantities because most items sold in these venues, like snack chips, loaves of bread, soda, or milk, have UPC codes, which provide information on package size. Conversely, items from grocery venues, such as farmers markets and specialty stores, had a significantly lower share of items with reported quantities (62 percent), likely reflecting the fact that these items are more likely to be sold in bulk or less likely to have UPC codes. Not surprisingly, items from own production had the lowest share of reported quantities, 49 percent, but the total number of such items was small.

Table 1a
Share of Food at Home (FAH) items with reported quantities by place of acquisition

Event location (number of items)	Share of items with reported quantities
Big grocery (123,117)	75%
Other grocery (3,983)	62%
All other FAH stores (16,008)	83%
Own production (924)	49%
Other assistance (1,337)	90%

FoodAPS = National Household Food Acquisition and Purchase Survey.
Source: USDA, Economic Research Service estimates using FoodAPS data.

We do a similar comparison using the What We Eat in America (WWEAI) 4-digit food category codes (USDA, ARS, 2016). The WWEAI category codes sort each of the FoodAPS food codes into broader food categories. The first digit of the food category classifies foods into very broad groups, such as milk and dairy, protein foods, and grains. Using these codes, we grouped items into 13 categories: grains; dark green, red, and orange vegetables and legumes; all other vegetables; whole fruit; 100 percent juice; dairy; meat; other protein; prepared food; beverages; desserts; salty snacks; and all other items (table 1b). Not surprisingly, items associated more with convenience stores or FAFH, such as prepared foods, beverages, and salty snacks, have a relatively high share of items with known quantities. Conversely, items such as vegetables and meat are more likely to have missing quantities, likely reflecting the fact that these items also make up the bulk of acquisitions from farmers markets and own production (e.g., gardening, hunting, or fishing).

Table 1b

Share of FoodAPS items with reported quantities by broad food groups, based on 4-digit What We Eat in America (WWEIA) codes

WWEIA 4-digit category aggregation (number of items)	Share of items with reported quantities
Grains (21,428)	86%
Dark green, red, and orange veg, and legumes (11,882)	80%
All other vegetables (11,947)	72%
Whole fruit (14,609)	81%
100% juice (4,506)	94%
Dairy (21,440)	89%
Meat (21,220)	64%
Other proteins (8,073)	80%
Prepared foods (50,165)	94%
Beverages (40,347)	94%
Desserts (22,595)	84%
Salty snacks (10,872)	88%
All other items (30,112)	83%

FoodAPS = National Household Food Acquisition and Purchase Survey.
Source: USDA, Economic Research Service estimates using FoodAPS data.

For the quantity imputations, the method used depended on whether there was expenditure information on that item and whether there was price information on suitable comparison items. The preferred imputation method used the average price paid for an item with the same FNDDS/SR food code in the same primary sampling unit (PSU), with the same type of rural or urban designation of the household's location, and of the same store type (e.g., supercenter or convenience store). Given the average price for that item, we then imputed a quantity using the total expenditures on the items with missing quantities. Quantities were imputed using this method for a total of 16,946 items, or roughly 47 percent of the items with missing quantities. When there was no suitable comparison in the PSU, the prices were averaged over the FNDDS/SR food code; the State, rural, or urban designation of the household's location; and store type. Quantities for another 3,767 items, or 10 percent of missing quantities, were imputed this way.

When there was not a match for the exact food code, the item's WWEIA 4-digit food category code was used for calculating average prices. The first digits of these codes represent the broadest categories, such as milk and dairy or protein foods. The next two digits refine from there, including, for

instance, flavored milk and cheese within the milk and dairy category or meat, poultry, and seafood within protein foods. The 4-digit codes are the most refined WWEIA grouping in the data; for example, the subcategory for poultry includes chicken, whole pieces, chicken patties, nuggets and tenders, and turkey, duck, other poultry (table 2). Another 9,858, or 27 percent, of item quantities were imputed this way. Another 2,490 of item quantities were imputed by using the average prices by State and store type; average prices by urban areas and store type; average prices by rural areas and store type; national average prices by store type, or overall national average prices (see table 3 for summary).

Table 2
USDA What We Eat in America (WWEAI) 4-digit food categories

Number	Description	Number	Description
1002	Milk, whole	2804	Nuts and seeds
1004	Milk, reduced fat	2806	Processed soy products
1006	Milk, lowfat	3002	Meat mixed dishes
1008	Milk, nonfat	3004	Poultry mixed dishes
1202	Flavored milk, whole	3006	Seafood mixed dishes
1204	Flavored milk, reduced fat	3202	Rice mixed dishes
1206	Flavored milk, lowfat	3204	Pasta mixed dishes, excludes macaroni and cheese
1208	Flavored milk, nonfat	3206	Macaroni and cheese
1402	Milk shakes and other dairy drinks	3208	Turnovers and other grain-based items
1404	Milk substitutes	3402	Fried rice and lo/chow mein
1602	Cheese	3404	Stir-fry and soy-based sauce mixtures
1604	Cottage/ricotta cheese	3406	Egg rolls, dumplings, sushi
1802	Yogurt, whole and reduced fat	3502	Burritos and tacos
1804	Yogurt, lowfat and nonfat	3504	Nachos
2002	Beef, excludes ground	3506	Other Mexican mixed dishes
2004	Ground beef	3602	Pizza
2006	Pork	3702	Burgers (single code)
2008	Lamb, goat, game	3703	Frankfurter sandwiches (single code)
2010	Liver and organ meats	3704	Chicken/turkey sandwiches (single code)
2202	Chicken, whole pieces	3706	Egg/breakfast sandwiches (single code)
2204	Chicken patties, nuggets and tenders	3708	Other sandwiches (single code)
2206	Turkey, duck, other poultry	3802	Soups
2402	Fish	4002	Rice
2404	Shellfish	4004	Pasta, noodles, cooked grains
2502	Eggs and omelets	4202	Yeast breads
2602	Cold cuts and cured meats	4204	Rolls and buns
2604	Bacon	4206	Bagels and English muffins
2606	Frankfurters	4208	Tortillas
2608	Sausages	4402	Biscuits, muffins, quick breads

—continued

Table 2

USDA What We Eat in America (WWEAI) 4-digit food categories—continued

Number	Description	Number	Description
2802	Beans, peas, legumes	4404	Pancakes, waffles, French toast
4404	Pancakes, waffles, French toast	6414	Onions
4602	Ready-to-eat cereal, higher sugar (>21.2g/100g)	6416	Corn
4604	Ready-to-eat cereal, lower sugar (=<21.2g/100g)	6418	Other starchy vegetables
4802	Oatmeal	6420	Other vegetables and combinations
4804	Grits and other cooked cereals	6422	Vegetable mixed dishes
5002	Potato chips	6802	White potatoes, baked or boiled
5004	Tortilla, corn, other chips	6804	French fries and other fried white potatoes
5006	Popcorn	6806	Mashed potatoes and white potato mixtures
5008	Pretzels/snack mix	7002	Citrus juice
5202	Crackers, excludes saltines	7004	Apple juice
5204	Saltine crackers	7006	Other fruit juice
5402	Cereal bars	7008	Vegetable juice
5404	Nutrition bars	7102	Diet soft drinks
5502	Cakes and pies	7104	Diet sport and energy drinks
5504	Cookies and brownies	7106	Other diet drinks
5506	Doughnuts, sweet rolls, pastries	7202	Soft drinks
5702	Candy containing chocolate	7204	Fruit drinks
5704	Candy not containing chocolate	7206	Sport and energy drinks
5802	Ice cream and frozen dairy desserts	7208	Nutritional beverages
5804	Pudding	7302	Coffee
5806	Gelatins, ices, sorbets	7304	Tea
6002	Apples	7502	Beer
6004	Bananas	7504	Wine
6006	Grapes	7506	Liquor and cocktails
6008	Peaches and nectarines	7702	Tap water
6010	Berries	7704	Bottled water
6012	Citrus fruits	7802	Flavored or carbonated water
6014	Melons	7804	Enhanced or fortified water
6016	Dried fruits	8002	Butter and animal fats
6018	Other fruits and fruit salads	8004	Margarine
6402	Tomatoes	8006	Cream cheese, sour cream, whipped cream
6404	Carrots	8008	Cream and cream substitutes
6406	Other red and orange vegetables	8010	Mayonnaise
6408	Dark green vegetables, excludes lettuce	8012	Salad dressings and vegetable oils

—continued

Table 2
USDA What We Eat in America (WWEAI) 4-digit food categories—continued

Number	Description	Number	Description
6410	Lettuce and lettuce salads	8402	Tomato-based condiments
6412	String beans	8404	Soy-based condiments
8406	Mustard and other condiments	9010	Baby food: yogurt
8408	Olives, pickles, pickled vegetables	9012	Baby food: snacks and sweets
8410	Pasta sauces, tomato-based	9202	Baby juice
8412	Dips, gravies, other sauces	9204	Baby water
8802	Sugars and honey	9402	Formula, ready-to-feed
8804	Sugar substitutes	9404	Formula, prepared from powder
8806	Jams, syrups, toppings	9406	Formula, prepared from concentrate
9002	Baby food: cereals	9602	Human milk
9004	Baby food: fruit	9802	Protein and nutritional powders
9006	Baby food: vegetable	9999	Not included in a food category
9008	Baby food: meat and dinners		

Source: USDA, Agricultural Research Service.

Table 3
Description and count of each imputation method

	Number of items	Share
Total number of items in FoodAPS	259,124	
Items with nonmissing food codes	254,587	98.2% (of all items)
Items with food codes and reported quantity information	218,119	87.5% (of items with food codes)
Items with food codes and ERS imputed quantities	36,388	14.3% (of items with food codes)
Imputations using average prices over:	Number of items	Share of imputed items
FNDDS/SR food code, psu, rural & store type	16,946	46.6%
FNDDS/SR food code, State, rural & store type	3,767	10.4%
4-digit WWEIA code, State, rural & store type	9,858	27.1%
FNDDS/SR food code, State, & store type	261	0.7%
4-digit WWEIA code, State, & store type	466	1.3%
FNDDS/SR food code, rural & store type	977	2.7%
4-digit WWEIA code, rural & store type	636	1.7%
FNDDS/SR food code & store type	27	0.1%
4-digit WWEIA code & store type	42	0.1%
FNDDS/SR food code	67	0.2%
4-digit WWEIA code	14	0.0%
Total number of imputations using price information	33,061	90.9%

—continued

Table 3
Description and count of each imputation method—continued

	Number of items	Share
Imputations using median acquisition by:		
FNDDS/SR food code and exact number of adult equivalents	1,648	4.5%
FNDDS/SR food code and adult equivalents rounded to nearest whole number	1,190	3.3%
4-digit WWEIA code and exact number of adult equivalents	349	1.0%
4-digit WWEIA code and adult equivalents rounded to nearest whole number	114	0.3%
2-digit WWEIA code and exact number of adult equivalents	16	0.0%
2-digit WWEIA code and adult equivalents rounded to nearest whole number	2	0.0%
1-digit WWEIA code and exact number of adult equivalents	3	0.0%
1-digit WWEIA code and adult equivalents rounded to nearest whole number	2	0.0%
FNDDS/SR food code	3	0.0%
Total number of imputations using median acquisitions	3,183	8.8%

FoodAPS = National Household Food Acquisition and Purchase Survey; WWEIA = What We Eat In America.
Source: USDA, Economic Research Service estimates using FoodAPS data.

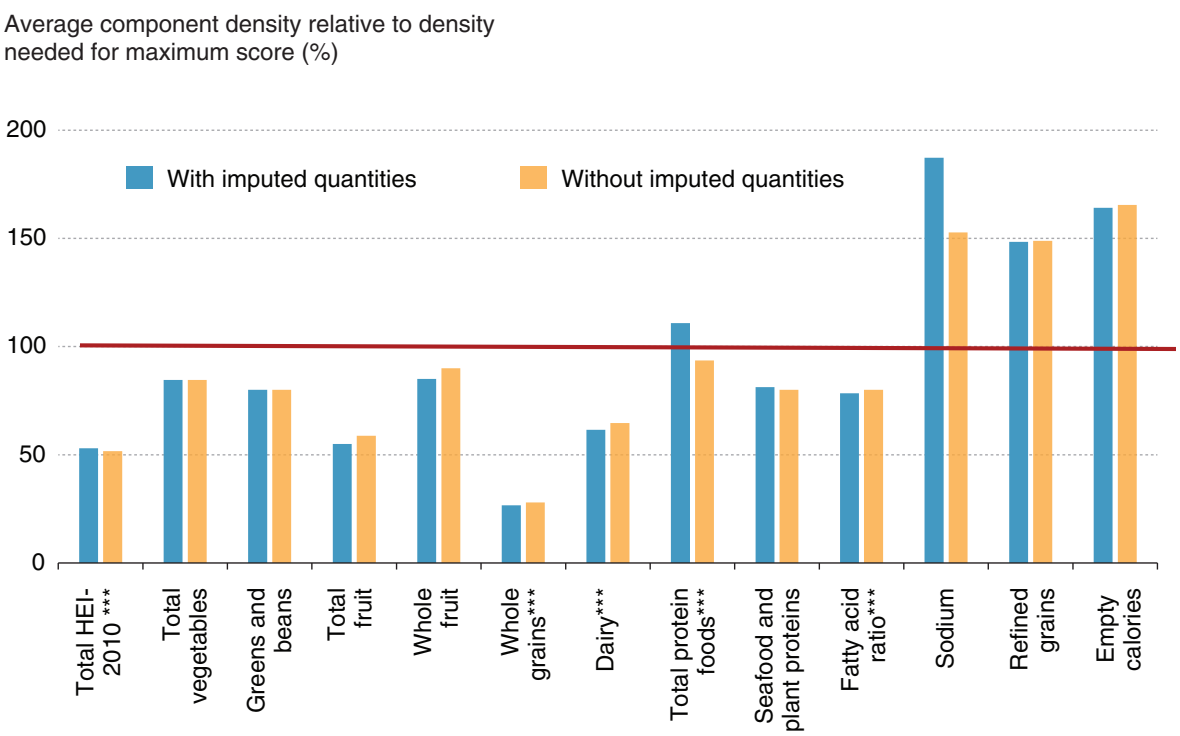
The remaining 3,327 items (9.1 percent) lacked information on quantities and expenditures. For those items, quantity imputations were based on the median quantity acquired for that FNDDS/SR food code or WWEIA 4-digit code by a household with the same composition, meaning the total number of household members and the age and gender of each member. Using this information and the calories needed per day from the 2010 *Dietary Guidelines for Americans* (Appendix 6, p. 78), we were able to estimate the energy requirement of each family member[3] and then calculate the total number of household members that require 2,000 calories per day (which we refer to as an adult equivalent). For example, a 30-year-old man, a 30-year-old woman, a 10-year-old boy, and a 7-year-old girl would have estimated daily calorie needs of 2,400, 1,800, 1,600, and 1,200, respectively. As a household, they would need 7,000 calories per day and make up 3.5 adult equivalents. However, there were 144 items from the FAFH file that had missing quantities that came from outlets such as restaurants. In these cases, we did not match on adult equivalents, but on the number of people at that event. Like the imputations based on average prices, imputations were first matched on the FNDDS/SR food code and the number of adult equivalents. If there was no information for that food code and type of household, quantities were imputed from the median quantity for that food code and the number of adult equivalents, rounded to the nearest whole number. If matches were not available for the FNDDS/SR food code, matches were based on the full 4-digit, 2-digit, or 1-digit WWEIA codes. Table 3 summarizes the breakdown of additional imputations using each method.

[3] Because FoodAPS did not ask information about physical activity, sedentary calorie requirements are assumed. Based on estimated energy requirement for pregnant and lactating women from the Institute of Medicine (2005), pregnant women are assumed to need an additional 250 calories per day and lactating women are assumed to need 365 more calories per day.

Checking the Quality of Imputations

Using our data with imputed quantities, we adapt the HEI-2010 SAS programs developed by the National Cancer Institute (NCI, 2017) into Stata and estimate average HEI-2010 scores and component densities, relative to targets from the 2010 Dietary Guidelines, with and without the imputed quantities (fig. 2). Specifically, we use the simple algorithm method to calculate each household's total HEI score and component densities based on its weekly acquisitions and then calculate averages with and without imputations. The data with imputed quantities and the Stata programs used to calculate HEI-2010 scores for FoodAPS are available on the FoodAPS web page. Although not shown, the average calories per-person, per-week without imputed quantities are approximately 17,600. The imputed calories add roughly 4,200 more, with the average household acquiring 21,800 calories per person, per week. Overall, HEI-2010 scores rise by a little over one point when imputed quantities are incorporated. Although this is a small difference, it is statistically significant. Other significant changes in the component densities from using imputed quantities are a decrease in dairy and whole-grain, a drop in the fatty acid ratio (ratio of unsaturated fats to saturated fats), and an increase in total protein foods. These changes likely reflect the fact that a large share of missing items were from the meat category, and imputing their quantities raises total calories and lowers the fatty acid ratio.

Figure 2
Comparing HEI-2010 scores and component densities with and without imputed quantities



*** Means with and without imputed quantities are estimated to differ with p<0.01. All estimates use sample weights and control for survey design.
HEI-2010 = Healthy Eating Index 2010; FoodAPS = National Household Food Acquisition and Purchase Survey.
Source: USDA, Economic Research Service estimates using FoodAPS data.

Using these imputed quantities, we compare average HEI-2010 scores and component densities from FoodAPS to the 2011-12 National Health and Nutrition Examination Survey (NHANES)[4] (fig. 3). Again, we use the simple mean approach, estimating average scores over households in FoodAPS and respondents in NHANES, using each survey's sample weights and method to account for sampling design to calculate the standard errors. It should be noted that some differences between the two surveys may affect how the HEI-2010 scores and component densities compare.[5] In FoodAPS, HEI-2010 scores and component densities are based on 7 days of household-level acquisitions. In NHANES, HEI-2010 scores and component densities are based on individual-level consumption from 2 days of 24-hour dietary recalls. FoodAPS nutrients are based on the form in which foods were acquired, while NHANES nutrients are based on the form in which they were consumed. While acquisitions over time should approximate consumption over time, both NHANES and FoodAPS were conducted over finite, relatively short periods. This means that using 1 or 2 days of dietary recall data likely underestimates the diversity in our overall diets. Further, 1 week's food purchase does not account for the possibility that people have existing stocks of food, do not consume all the edible parts of the food they acquire or simply throw out some of this food. Despite these differences, the patterns are similar. Both datasets show that Americans acquire (or eat, in the case of NHANES), on a proportional basis, too few fruits, vegetables, whole grains, and dairy compared to recommended levels. They also acquire (consume) too much sodium and too many refined grains and empty calories. We see that total HEI-2010 scores do not differ significantly across the surveys. In NHANES, the score is estimated to be 54 (out of 100) while the average FoodAPS HEI-2010 is 53. There are also no significant differences in many of the component densities—total protein foods; seafood and plant-based protein foods; the fatty acid ratio; and the amount of sodium and refined grains. While the average density for sodium appears much higher in FoodAPS, the variance in FoodAPS is also quite large. This likely reflects the fact that sodium is acquired in many forms—in a food-away-from-home meal and in processed, packaged foods, but also in larger amounts for cooking and seasoning foods at home. These purchases of salt for home seasoning tend to be infrequent, while consumption is spread out more evenly over time, reducing variation in our daily consumption.

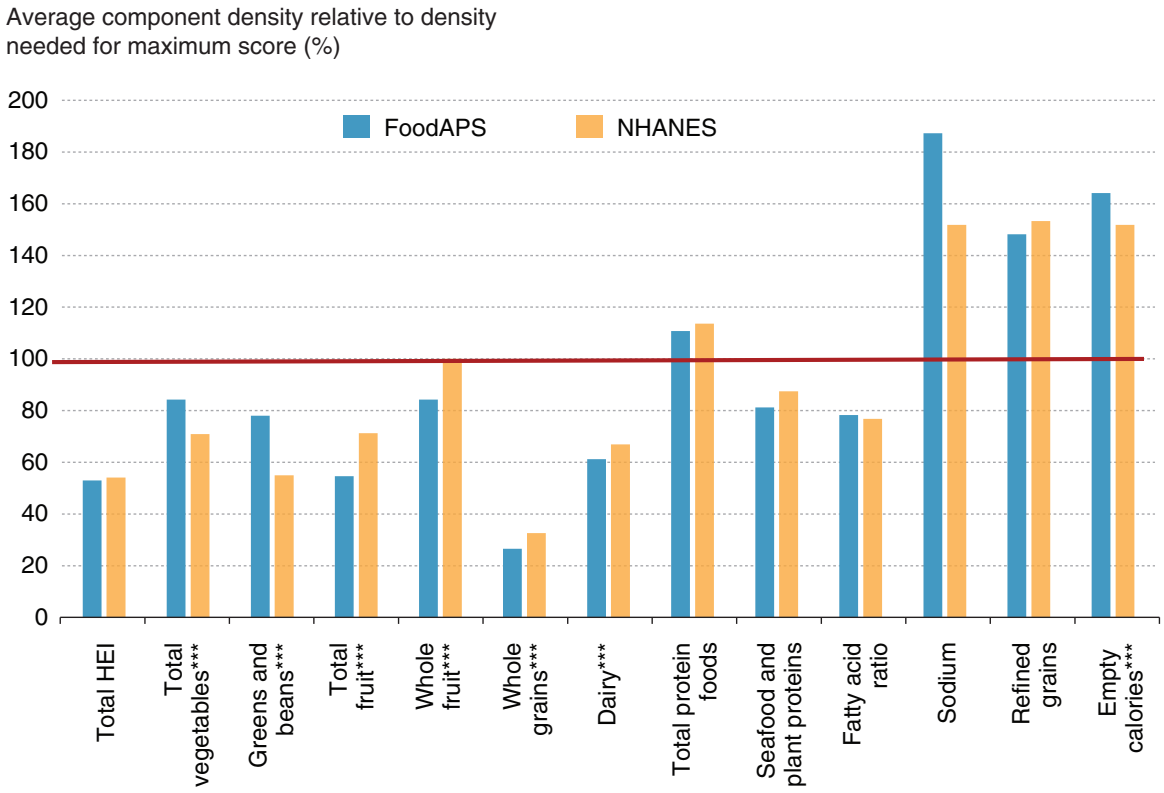
There are differences in component densities between the two surveys that should be noted. For both total vegetables and greens and beans, FoodAPS estimates are significantly higher than NHANES estimates. This may reflect some amount of waste between acquisition and consumption and how that translates to cup equivalents of purchased versus consumed amounts. FoodAPS also seems to underestimate whole-grain densities relative to NHANES. We suspect this may reflect a lack of suitable food codes for certain whole-grain items, like uncooked brown rice or an egg sandwich on

[4] We compare FoodAPS to the 2011-12 National Health and Nutrition Examination Survey (NHANES) because FoodAPS was collected between April 2012 and January 2013. NHANES data are collected in nonoverlapping, 2-year cycles. Thus, the time coverage of 2013-14 NHANES data would not coincide as well with FoodAPS as the 2011-12 data.

[5] It should also be noted that other published numbers on the HEI-2010 using NHANES 2011-12, such as those posted on the CNPP website in describing the HEI-2010, use the population-ratio approach. This approach sums the amount of each HEI component over the entire population and then divides that sum by total calories consumed by the entire population. This is done to better approximate usual intake over 1 or 2 days of intake data. In contrast, the simple algorithm sums each individual's HEI-2010 components and then divides those by his or her total calories. While this may be less representative of usual intake over an entire population, it yields an HEI-2010 score for each unit of observation. This method is thus commonly used when HEI-2010 scores, component densities, or component scores are used as dependent variables, which is their expected use for FoodAPS research.

a whole-grain English muffin. Finally, the differences in empty calories between the two surveys may again reflect differences in how we acquire items versus how we consume them. For example, a bottle of wine, a case of beer, and a 24-pack of soda would all contribute to empty calories. While we might buy all of them on a single shopping trip, we are unlikely to consume them all in 1 day.

Figure 3
Comparing HEI-2010 scores and component densities with and without imputed quantities



*** Means with and without imputed quantities are estimated to differ with p<0.01. All estimates use sample weights and control for survey design
HEI-2010 = Healthy Eating Index 2010; FoodAPS = National Household Food Acquisition and Purchase Survey.
Source: USDA, Economic Research Service estimates using FoodAPS data.

Classifying Foods Into Food Groups

While the HEI-2010 evaluates the quality of foods in relation to USDA recommendations, it does not summarize information about the form in which a food item was purchased. The WWEIA food groups classify foods according to their main ingredient, which does result in mixed dishes being grouped with primary ingredients. This classification system directly links to FNDDS food codes and is useful for some research objectives. Researchers may also be interested in the form in which foods are purchased, however, and how that relates to diet quality. This information may give insight into intermediate steps that consumers take and how they balance time constraints, dietary restrictions, tastes, preferences, and cooking or preparation skills when making food choices.

We developed a classification system for food items to aid researchers interested in studying consumer demand for foods in relation to their quality, as well as likely price premiums for convenience and processing, building on the classification approach for the Quarterly Food-at-Home Price Database (QFAHPD; see Todd et al., 2010). For example, the category of dark green vegetables can be broken down into fresh, frozen with no added ingredients, or canned dark green vegetables. Frozen dark green vegetables with an added sauce (butter or cheese) are separated and grouped instead with prepared dishes.

Classification Methodology

We first separated the food items into eight main categories (Tier 1): grains; vegetables; fruit; dairy products; meat and beans; prepared meals, sides, and salads; other foods; and items that could not be identified (see table 4). Tier 2 further divides each main category into subcategories (e.g., grains are divided into whole and refined grains). The Tier 3 division further refines the subcategories into specific types of foods (e.g., whole-grain breads; whole-grain rice and pasta, etc.), for a total of 82 separate ERS food groups.

Table 4
ERS food groups

Tier 1	Tier 2	Tier 3
1 - Grains		
	101- Whole grain breads, cereal, rice, pasta, and flours	
	Breads (bread, rolls, pita, bagels, tortillas)	10101
	Rice and pasta	10102
	Breakfast cereal	10103
	Flour/bread mixes/frozen dough	10104
	102 - Non-whole-grain breads, cereal, rice, pasta, and flours	
	Breads (bread, rolls, pita, bagels)	10201
	Rice and pasta	10202
	Breakfast cereal	10203
	Flour/bread mixes/frozen dough	10204
2 - Vegetables		
	201 - Starchy vegetables	
	Fresh	20101

—continued

Table 4
ERS food groups—continued

Tier 1	Tier 2	Tier 3
	Frozen	20102
	Canned	20103
	202 - Tomatoes	
	Fresh	20201
	Frozen	20202
	Canned	20203
	203 - Dark green vegetables	
	Fresh	20301
	Frozen	20302
	Canned	20303
	204 - Other red and orange vegetables	
	Fresh	20401
	Frozen	20402
	Canned	20403
	205 - Beans, lentils, and peas or legumes	
	Fresh/Dried	20501
	Frozen	20502
	Canned	20503
	206 - Other/mixed vegetables	
	Fresh	20601
	Frozen	20602
	Canned	20603
3 - Fruit		
	301 - Whole fruit	
	Fresh	30101
	Frozen	30102
	Canned	30103
	Dried	30104
	302 - 100% Fruit and vegetable juices	30201
4 - Milk products		
	401 - Whole milk, yogurt, and cream	
	Milk	40101
	Cream	40102
	Yogurt	40103
	402 - Low-fat and skim milk and low-fat yogurt	
	Milk	40201

—continued

Table 4
ERS food groups—continued

Tier 1	Tier 2	Tier 3
	Cream	40202
	Yogurt	40203
	403 - All cheese, including cheese soups and sauces	
	Cheese	40301
	Processed	40302
5 - Meat other proteins		
	501 - Beef, pork, veal, lamb, and game	
	Fresh	50101
	Frozen	50102
	Canned	50103
	502 - Chicken, turkey, and game birds	
	Fresh	50201
	Frozen	50202
	Canned	50203
	503 - Fish and seafood	
	Fresh	50301
	Frozen	50302
	Canned	50303
	504 - Nuts, nut butters, and seeds	
	Nuts and Seeds	50401
	Nut and Seed Butters and Spreads	50402
	505 - Bacon, sausage, and lunch meats including spreads	50501
	506 - Egg and egg substitutes	50601
	507 - Tofu and meat substitutes	50701
6 - Prepared meals, sides, and salads		
	Ready to eat	60101
	Frozen	60201
	Canned	60301
	Packaged	60401
7 - Other foods		
	701 - Table fats, oils, and salad dressings	
	Fats and oils	70101
	Salad dressing	70102
	702 - Gravies, sauces, condiments, and spices	
	Condiments, gravies, and sauces	70201
	Dry spices	70202

—continued

Table 4
ERS food groups—continued

Tier 1	Tier 2	Tier 3
	703 - Beverages	
	Sweetened coffee and tea	70301
	Unsweetened coffee and tea	70302
	Low-calorie beverages	70303
	All other caloric beverages	70304
	Alcohol	70305
	Water	70306
	704 - Desserts, sweets, and candies	
	Sweeteners	70401
	Jellies/jams	70402
	Candy	70403
	Baked goods (including packaged)	70404
	Cake mixes	70405
	Milk drinks and milk desserts	70406
	All other desserts	70407
	705 - Salty snacks	
	Whole grain snacks	70501
	All other snacks	70502
	706 - Vitamins and meal supplements	70601
	707 - Baby food	70701
	708 - Infant formula	70801
9 - Not coded		
	999-Not coded	99999

Source: USDA, Economic Research Service (ERS) estimates using ERS FoodAPS data.

In general, the assigned FNDDS/SR food codes were used as the first level of information to sort the item. The item description given by the respondent and IRI database information about the item were used as secondary sources to increase accuracy. In the case of the FAFH items, the source of the item (foodstore or other source) was also taken into account in assigning the item to 1 of the 82 ERS food groups. In addition, items identified as belonging to a bundle (such as mashed potatoes and gravy listed as two separate items, or taco, lettuce, cheese, and sour cream listed as four items in a bundle) were placed in the ready-to-eat food group, when they were clearly part of a mixed dish. Beverages (including milk and juice), desserts, salty snacks, and other "a la carte" items that may have been purchased in a bundle, but are clearly separable items, were sorted individually into their respective food groups.

The 82 ERS food groups may not meet the needs of all research objectives. Researchers are encouraged to use item descriptions, nutrient and FPED values, or other item characteristics available in the data to revise or refine the item-sorting to meet project needs.

Summary Statistics

To illustrate the overall nutritional qualities of the ERS food groups, table 5 provides a summary of the mean value of FoodAPS items of each HEI-2010 component or food-plan group per 100 grams in 14 aggregated food categories, which are similar to the broad 4-digit WWEIA categories: whole grains; refined grains; dark-green, red, and orange vegetables and legumes; all other vegetables; whole fruit; 100-percent juice; dairy; meat; other proteins; prepared foods; beverages; desserts; salty snacks; and all other items. Whole grains are clearly differentiated by their whole-grain and refined-grain content per 100 grams. The main vegetable content of foods placed in the dark-green, red, and orange vegetable and legumes (DGROL) group are these vegetables and plant proteins, while all other vegetables include mainly starchy vegetables but also some dark-green and red/orange vegetables, most likely due to the presence of broccoli and carrots in many mixed-vegetable frozen products.

A similar differentiation can be seen between whole fruit and 100-percent juices (juice is the remainder of the total fruit HEI-2010 component, but it is not a component on its own). The main HEI-2010 component in the whole-fruit group is whole fruit, while the main component is juice in the 100-percent juice category. Likewise, the largest HEI-2010 components in the other-protein foods category are seafood and plant proteins, while meat is the largest component in the meat category.

Table 5

Mean HEI-2010 component or Food Pattern group values per 100 grams of food, by aggregated ERS food groups

	14 -category aggregation of ERS Food Groups (indicates ERS Food groups/tiers included)				
HEI component or food pattern group	Whole grains (101)	Refined grains (102)	Dark green, red, orange veg, and legumes (203,204,205)	All other veg (Tier 1= 2, excl. 203,204)	Whole fruit (301)
Whole grains, oz. eq.	2.62	0.12	nc	nc	nc
	(0.018)	(0.003)			
Refined grains, oz. eq.	0.29	3.47	nc	0.01	nc
	(0.008)	(0.008)		(0.001)	
Dark green veg, cup eq.	nc	nc	0.28	0.10	nc
			(0.007)	(0.004)	
Red/orange veg, cup eq.	nc	nc	0.39	0.02	nc
			(0.005)	(0.001)	
Starchy veg, cup eq.	nc	nc	0.01	0.15	nc
			(0.001)	(0.002)	
Whole fruit, cup eq.	0.02	nc	nc	nc	0.66
	(0.001)				(0.002)
Fruit juice, cup eq.	nc	nc	nc	nc	nc
Dairy, cup eq.	nc	0.01	nc	nc	nc
		(0.000)			
Meat, oz. eq.	nc	nc	nc	nc	nc
Seafood & plant protein, oz. eq.	0.08	0.01	0.49	0.00	nc
	(0.005)	(0.001)	(0.018)	(0.001)	
Sodium (mg)	374.87	473.67	65.14	109.46	2.59
	(3.821)	(2.626)	(1.077)	(2.202)	(0.251)
Empty calories (% of total)	16.33	9.66	0.36	0.44	4.55
	(0.246)	(0.102)	(0.049)	(0.042)	(0.137)
Fatty acid ratio (poly+mono)/sat	3.42	3.00	3.32	3.21	4.52
	(0.025)	(0.009)	(0.019)	(0.029)	(0.039)
Number of items	2,827	12,305	5,872	12,957	13,030

—continued

Table 5

Mean HEI-2010 component or Food Pattern group values per 100 grams of food, by aggregated ERS food groups—continued

	14 category aggregation of ERS Food Groups (indicates ERS Food groups/tiers included)				
HEI component or food pattern group	100% juice (302)	Dairy (Tier 1 = 4)	Meat (501,502,505)	Other proteins (503, 504,506,507)	Prepared foods (6)
Whole grains, oz. eq.	nc	nc	nc	nc	0.03
					(0.001)
Refined grains, oz. eq.	nc	nc	0.01	0.04	0.82
			(0.001)	(0.003)	(0.003)
Dark green veg, cup eq.	nc	nc	nc	nc	0.03
					(0.000)
Red/orange veg, cup eq.	0.02	nc	nc	nc	0.04
	(0.001)				(0.000)
Starchy veg, cup eq.	nc	nc	nc	nc	0.09
					(0.001)
Whole fruit, cup eq.	0.15	0.01	nc	nc	0.01
	(0.003)	(0.000)			(0.000)
Fruit juice, cup eq.	0.41	nc	nc	nc	nc
	(0.003)				
Dairy, cup eq.	nc	0.92	0.00	nc	0.14
		(0.008)	(0.001)		(0.001)
Meat, oz. eq.	nc	nc	2.80	nc	0.57
			(0.005)		(0.003)
Seafood & plant protein, oz. eq.	nc	0.00	nc	3.60	0.18
		(0.001)		(0.026)	(0.002)
Sodium (mg)	10.84	253.67	573.72	282.09	546.00
	(0.609)	(2.971)	(4.625)	(3.647)	(3.657)
Empty calories (% of total)	0.00	40.57	28.97	10.11	19.38
	(0.004)	(0.175)	(0.213)	(0.169)	(0.059)
Fatty acid ratio (poly+mono)/sat	2.86	0.59	1.75	3.24	2.61
	(0.017)	(0.005)	(0.004)	(0.032)	(0.005)
Number of items	3,767	16,127	13,361	5,466	79,155

—continued

Table 5

Mean HEI-2010 component or Food Pattern group values per 100 grams of food, by aggregated ERS food groups—continued

HEI component or food pattern group	14 category aggregation of ERS Food Groups (indicates ERS Food groups/tiers included)			
	Beverages (703)	Desserts (704)	Salty snacks (705)	All other items
Whole grains, oz. eq.	nc	0.06	0.49	0.00
		(0.002)	(0.012)	(0.001)
Refined grains, oz. eq.	nc	0.87	2.31	0.03
		(0.007)	(0.022)	(0.002)
Dark green veg, cup eq.	nc	nc	nc	nc
Red/orange veg, cup eq.	nc	nc	0.00	0.03
			(0.001)	(0.001)
Starchy veg, cup eq.	nc	nc	0.56	nc
			(0.008)	
Whole fruit, cup eq.	0.01	0.05	0.01	0.25
	(0.001)	(0.002)	(0.000)	(0.002)
Fruit juice, cup eq.	0.03	nc	nc	nc
	(0.001)			
Dairy, cup eq.	0.01	0.13	0.03	0.01
	(0.000)	(0.001)	(0.001)	(0.000)
Meat, oz. eq.	nc	nc	nc	1.06
				(0.007)
Seafood & plant protein, oz. eq.	nc	0.16	0.12	0.02
		(0.003)	(0.006)	(0.001)
Sodium (mg)	17.48	237.89	625.00	380.30
	(0.329)	(1.395)	(2.822)	(3.355)
Empty calories (% of total)	46.29	58.15	10.04	18.68
	(0.333)	(0.129)	(0.164)	(0.139)
Fatty acid ratio (poly+mono)/sat	3.67	1.90	5.30	3.35
	(0.022)	(0.010)	(0.020)	(0.017)
Number of items	15,237	24,800	10,131	35,356

Notes: USDA, Economic Research Service (ERS) food groups included in each category provided in parentheses under category description. See table 4 for complete list of ERS food groups. Standard error (SE) of mean content in parentheses under mean. nc= no content, mean= 0.00 and se <=0.000. HEI-2010 = Healthy Eating Index 2010. FoodAPS = National Household Food Acquisition and Purchase Survey.

Source: ERS estimates of FoodAPS data.

As expected, the ERS prepared-foods category is a mix of HEI-2010 components, with the relatively large mean values of refined grains, starchy vegetables, dairy, meat, and sodium per 100 grams. The dessert category includes mainly grain-based sweetened foods, and at 58 percent of total calories, this category has the highest share of calories classified as empty calories. The beverages and dairy categories also have high mean percentages of empty calories per 100 grams, at 46 and 41 percent, respectively. Salty snacks contain mainly grain and starchy-vegetable foods and have the highest sodium content of all the categories (625 mg per 100 grams).

Although the ERS food groups result in a classification of foods that is somewhat similar to the WWEIA food categories, the food-constituent contributions of the different classification systems differ in ways that are worth noting (table 6). The WWEIA categories do not differentiate whole grains from refined grains, so there is only one aggregate grain category, and the grain content foods in WWEIA grain categories reflect the fact that most grain-based foods do not contain whole grains. The DGROL aggregation of WWEIA categories has less red and orange vegetable content, and a higher mean of sodium and empty calories as a percent of total calories, compared to the same aggregation of the ERS food groups. We see less dark-green vegetable content in the other-vegetables aggregation, and a higher percent of empty calories, compared to the ERS food group aggregation. This suggests that some of the items classified into the WWEIA vegetable group have added ingredients. The whole-fruit and 100-percent juice aggregations are very similar, with the WWEIA whole fruit aggregation having slightly higher sodium and empty calories, a lower fatty-acid ratio, and a juice aggregation with slightly less sodium but higher empty calories.

The WWEIA and ERS food-group protein-food aggregations are only slightly different from each other. However, the amount of seafood and plant protein is lower and the sodium and empty calorie content higher in the WWEIA other-protein aggregation compared to the ERS food-group aggregation. The prepared-foods WWEIA aggregation has more refined grain, starchy vegetable, dairy, and sodium, and less meat and seafood and plant-protein content than the same aggregation of ERS food groups. Similarly, the WWEIA desserts have more refined grains, more sodium, and more empty calories than the dessert aggregation of ERS food groups, while the salty snacks WWEIA group has more empty calories.

Table 6

Mean HEI-2010 component or Food Pattern group values per 100 grams of food, by aggregated USDA 4-digit food categories

HEI component or food pattern group	4-digit WWEIA category aggregation (indicates 4-digit categories included)				
	Grains (4002-4804)	Dark green, red, orange veg, & legumes (2802; 6402-10)	All other vegetables (6412-20)	Whole fruit (6002-18)	100% juice (7002-8)
Whole grains, oz. eq.	0.51	nc	nc	nc	nc
	(0.007)				
Refined grains, oz. eq.	2.61	nc	0.04	nc	nc
	(0.010)		(0.003)		
Dark green veg, cup eq.	nc	0.28	0.01	nc	nc
		(0.004)	(0.001)		
Red/orange veg, cup eq.	nc	0.26	0.03	nc	0.02
		(0.003)	(0.001)		(0.001)
Starchy veg, cup eq.	nc	nc	0.16	nc	nc
		(0.001)	(0.002)		
Whole fruit, cup eq.	0.01	nc	nc	0.66	0.15
	(0.000)			(0.002)	(0.003)
Fruit juice, cup eq.	nc	nc	nc	nc	0.38
					(0.003)
Dairy, cup eq.	0.02	nc	nc	nc	nc
	(0.000)				
Meat, oz. eq.	nc	0.01	nc	nc	nc
		(0.000)			
Seafood & plant protein, oz. eq.	0.04	0.47	nc	nc	nc
	(0.001)	(0.010)			
Sodium (mg)	464.54	112.03	109.35	2.91	10.27
	(1.831)	(1.261)	(1.293)	(0.229)	(0.546)
Empty calories (% of total)	12.74	4.81	2.58	5.28	2.09
	(0.089)	(0.094)	(0.053)	(0.132)	(0.141)
Fatty acid ratio (poly+mono)/sat	3.21	3.30	2.73	4.42	2.92
	(0.008)	(0.013)	(0.013)	(0.036)	(0.020)
Number of items	21,227	11,877	11,934	14,590	4,467

—continued

Table 6

Mean HEI-2010 component or Food Pattern group values per 100 grams of food, by aggregated USDA 4-digit food categories—continued

HEI component or food pattern group	4-digit WWEIA category aggregation (indicates 4-digit categories included)			
	Dairy (1002-1804; 8006-8)	Meat (2002-2206;2602-8)	Other proteins (2402-2502; 2804-2806)	Prepared foods (3002-3802; 6422-6806)
Whole grains, oz. eq.	nc	nc	0.00	0.02
			(0.001)	(0.000)
Refined grains, oz. eq.	0.01	0.18	0.08	0.95
	(0.001)	(0.003)	(0.003)	(0.004)
Dark green veg, cup eq.	nc	nc	nc	0.01
				(0.000)
Red/orange veg, cup eq.	nc	nc	nc	0.04
				(0.000)
Starchy veg, cup eq.	nc	nc	nc	0.14
				(0.001)
Whole fruit, cup eq.	nc	nc	0.01	nc
			(0.001)	
Fruit juice, cup eq.	nc	nc	nc	nc
Dairy, cup eq.	0.82	nc	0.02	0.18
	(0.007)		(0.001)	(0.001)
Meat, oz. eq.	nc	2.73	0.01	0.48
		(0.004)	(0.001)	(0.003)
Seafood & plant protein, oz. eq.	nc	0.01	3.24	0.10
		(0.000)	(0.020)	(0.001)
Sodium (mg)	234.66	593.26	325.53	596.18
	(2.427)	(3.302)	(2.845)	(5.646)
Empty calories (% of total)	44.11	26.67	12.50	19.43
	(0.159)	(0.163)	(0.163)	(0.062)
Fatty acid ratio (poly+mono)/sat	0.65	2.07	3.25	2.29
	(0.005)	(0.006)	(0.024)	(0.007)
Number of items	21,312	21,220	8,030	50,130

—continued

Table 6

Mean HEI-2010 component or Food Pattern group values per 100 grams of food, by aggregated USDA 4-digit food categories—continued

HEI component or food pattern group	4-digit WWEIA category aggregation (indicates 4-digit categories included)			
	Beverages (7102-7804)	Desserts (5402;5502-5806)	Salty snacks (5002-5204)	All other items
Whole grains, oz. eq.	nc	0.02	0.52	0.01
		(0.001)	(0.011)	(0.001)
Refined grains, oz. eq.	nc	1.03	2.30	0.05
		(0.008)	(0.021)	(0.002)
Dark green veg, cup eq.	nc	nc	nc	0.04
				(0.002)
Red/orange veg, cup eq.	nc	nc	0.00	0.04
			(0.001)	(0.001)
Starchy veg, cup eq.	nc	nc	0.53	nc
			(0.008)	
Whole fruit, cup eq.	0.01	0.06	nc	0.33
	(0.000)	(0.002)		(0.002)
Fruit juice, cup eq.	0.02	nc	nc	nc
	(0.000)			
Dairy, cup eq.	0.02	0.10	0.04	0.02
	(0.001)	(0.001)	(0.001)	(0.001)
Meat, oz. eq.	nc	nc	nc	nc
Seafood & plant protein, oz. eq.	nc	0.16	0.08	0.04
		(0.003)	(0.003)	(0.002)
Sodium (mg)	19.54	264.87	609.74	303.60
	(0.369)	(1.666)	(2.543)	(3.619)
Empty calories (% of total)	47.61	61.06	11.17	13.43
	(0.345)	(0.134)	(0.161)	(0.146)
Fatty acid ratio (poly+mono)/sat	3.67	2.04	5.14	4.36
	(0.023)	(0.011)	(0.021)	(0.023)
Number of items	14,367	19,318	10,839	29,279

Notes: WWEIA food categories included in each column provided in parenthesis under category description. See table 2 for complete list of the WWEIA food categories. Standard error (SE) of mean content in parentheses under mean.nc= no content, mean= 0.00 and se <=0.000.

WWEIA = What We Eat in America; FoodAPS = National Household Food Acquisition and Purchase Survey. HEI = Healthy Eating Index.

Source: USDA, Economic Research Service estimates of FoodAPS data.

Conclusions

The FoodAPS data include a number of item-level measures and identifications that can be used to study the nature of household food acquisitions. These include item descriptions, expenditures, assigned food codes, nutrient values, and FPED values. This report describes ERS's extensions to the item-level data to include imputed quantities, a new food group classification, and the HEI-2010 scores that can be constructed using FPED values and reported and imputed quantities. The data required to impute quantities are not available on the public-use file; among users accessing data through NORC, accessing all the required files entails multiple Third Party Agreements (TPAs). Posting the imputed quantities to the public use files allows all users to calculate HEI-2010 scores in a manner consistent with calculations in ERS publications and facilitates comparison of research results across multiple papers. Having a second defined hierarchy of food groups simplifies the process of aggregating individual food items into broader groups and allows for more consistency among multiple research projects, again facilitating comparison of results across publications.

References

Bowman, S.A., J.C. Clemens, J.E. Friday, R.C. Thoerig, and. J.A Moshfegh. 2014. Food Patterns Equivalents Database 2011-12: Methodology and User Guide. Food Surveys Research Group, Beltsville Human Nutrition Research Center, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, Maryland.

Guenther, Patricia, Sharon I. Kirkpatrick, Jill Reedy, Susan M. Krebs-Smith, Dennis W. Buckman, Kevin W. Dodd, Kellie O. Casavale, and Raymond J. Carroll. 2014. "The Healthy Eating Index-2010 Is a Valid and Reliable Measure of Diet Quality According to the 2010 Dietary Guidelines for Americans," *The Journal of Nutrition,* doi: 10.3945/jn.113.183079.

Guenther Patricia M., K.O. Casavale, J. Reedy, S.I. Kirkpatrick, , H.A. Hiza, K. J. Kuczynski, L.L. Kahle, and S.M. Krebs-Smith. 2013. "Update of the Healthy Eating Index: HEI-2010," *Journal of the Academy of Nutrition and Dietetics* 113(4):569-80

Institute of Medicine (IOM). 2005. "Dietary reference intakes for energy, carbohydrate, fiber, fat, fatty acids, cholesterol, protein, and amino acids." Panel on Macronutrients, Panel on the Definition of Dietary Fiber, Subcommittee on Upper Reference Levels of Nutrients, Subcommittee on Interpretation and Uses of Dietary Reference Intakes, and the Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board.

National Cancer Institute (NCI). 2017. HEI Tools for Researchers.

Schap, T., K. Kuczynski, and H. Hiza. 2017. "Healthy Eating Index—Beyond the Score," *Journal of the Academy of Nutrition and Dietetics* 117(4)519-21.

Todd, Jessica E., Lisa Mancino, Ephraim Leibtag, and Christina Tripodo. 2010, *Methodology Behind the Quarterly Food-at-Home Price Database*, Technical Bulletin No. 1926, Economic Research Service, U.S. Dept. of Agriculture.

U.S. Department of Agriculture, Agricultural Research Service (ARS). 2016. *USDA Food and Nutrient Database for Dietary Studies 2013-2014.* Food Surveys Research Group Home Page.

U.S. Department of Agriculture (USDA) and U.S. Department of Health and Human Services (HHS). 2010. *Dietary Guidelines for Americans*, 2010, Seventh Edition. U.S. Government Printing Office, Washington, DC. December

U.S. Department of Health and Human Services (HHS) and U.S. Department of Agriculture (USDA). 2015. *Dietary Guidelines for Americans*, 2015-2020, Eighth Edition. U.S. Government Printing Office, Washington, DC. December.

Appendix 1. National Household Food Acquisition and Purchase Survey (FoodAPS) Codebook: ERS Imputed Quantities (IQ) for Food-at-Home (FAH) Items– Public Use File faps_eiq_fahitem_puf

HHNUM

Variable: HHNUM	Definition: 6-digit unique identifier for household	Type: Numeric
	143,050 responses with 4,367 unique values. Individual responses not shown.	

EVENTID

Variable: EVENTID	Definition: Unique identifier for each event and can be used to link the event to the items in the faps_fahitem_puf file.	Type: Numeric
	Note: EVENTID is unique across FAH & FAFH files.	
	143,050 responses with 4,367 unique values. Individual responses not shown.	

ITEMNUM

Variable: ITEMNUM	Definition: Unique identifier for each event and can be used to link the event to the items in the faps_fahitem_puf file.	Type: Numeric
	Note: To uniquely identify an item entry, EVENTID and ITEMNUM. ITEMNUM does not uniquely identify any particular food item, such as “12 oz box of Cheerios.”	
Range	1-202	
Missing observations (.):	0 (out of 143,050)	

HEI_GRAMS

Variable: HEI_GRAMS	Definition: Total edible gram weight of the food item			Type: Numeric	
	Note: HEI_GRAMS are imputed one when TOTGRAMSEDIBLE and TOTGRAMSEDIBLEIMP are missing.				
	N	Min	Max	Mean	#Missing (.)
	143,050	.4993	640307.2	898.294	0

IMP_GRAM_FLAG

Variable: IMP_GRAM_FLAG	Definition: FLAG-edible gram weight is imputed as described in Technical Bulletin…...			Type: Numeric
	Value	Count	Percent	Value description
	.	102,887	71.92	Gram weight not imputed by ERS
	1	40,163	28.08	Gram weight imputed by ERS

Appendix 1b. National Household Food Acquisition and Purchase Survey (FoodAPS) Codebook: ERS Imputed Quantities (IQ) for Food-at-Home (FAFH) Items– Public Use File faps_eiq_fafhitem_puf

HHNUM

Variable: HHNUM	Definition: 6-digit unique identifier for household	Type: Numeric
	116,074 responses with 4,305 unique values. Individual responses not shown.	

EVENTID

Variable: EVENTID	Definition: Unique identifier for each event and can be used to link the event to the items in the faps_fahitem_puf file.	Type: Numeric
	Note: EVENTID is unique across FAH & FAFH files.	
	116,074 responses with 37,407 unique values. Individual responses not shown.	

ITEMNUM

Variable: ITEMNUM	Definition: Unique identifier for each event and can be used to link the event to the items in the faps_fahitem_puf file.	Type: Numeric
	Note: To uniquely identify an item entry, EVENTID and ITEMNUM. ITEMNUM does not uniquely identify any particular food item, such as “12 oz box of Cheerios.”	
Range	1-61	
Missing observations (.):	0 (out of 116,074)	

HEI_GRAMS

Variable: HEI_GRAMS	Definition: Total edible gram weight of the food item			Type: Numeric
	Note: HEI_GRAMS are imputed when TOTGRAMS are missing and foodcode is non-missing.			
	N	Min	Max	Mean	#Missing (.)
	115,7888	1	113664	331.0923	286

IMP_GRAM_FLAG

Variable: IMP_GRAM_FLAG	Definition: FLAG-edible gram weight is imputed as described in Technical Bulletin…..			Type: Numeric
	Value	Count	Percent	Value description
	.	115,866	99.84	Gram weight not imputed by ERS
	1	188	0.16	Gram weight imputed by ERS

Appendix 2. Overview of programs to calculate Healthy Eating Index-2010 scores

Appendixes 3-5 provide example do files that can be used with faps_eiq_fahitem_puf, faps_fahnutrients_puf and faps_fafhnutrients_puf to construct weekly HEI-2010 scores at the household level.

These programs aggregate all of the items for the whole week for each household. One can modify this program to aggregate by place type, day of the week, or any other unit of analysis desired. Background on the HEI-2010 can be found here: http://epi.grants.cancer.gov/hei/. We detail the steps to use these programs below.

- Step one: Get item-level data into pyramid-equivalent quantities by using example program in Appendix 3.
- Step two: Save example program in Appendix 4a as "$hei\hei-2010 scores.do" and example program in Appendix 4b as "$hei\legumes.do," making sure that "$hei" points to the folder on your computer where you have saved these macros.
- Step three: Calculate weekly HEI-2010 scores, component densities, and component scores for each household using example program in Appendix 5.

Appendix 3. Example STATA program to get item-level data into pyramid-equivalent quantities

```
capture log close

/*STEP 1--APPLY QUANTITIES TO FAH NUTRIENT DATA*/
global hei *"hei" Should refer to folder where quantity data and HEI programs are
stored
global foodaps *"foodaps" Should refer to folder where main FoodAPS data are
stored--they could be in the same directory as above

use "$foodaps\faps_fahnutrients.dta"

drop foodcode-refusesourcempr

merge 1:1 hhnum eventid itemnum using "$hei\faps_eiq_fahitem_puf.dta"

drop _m

/*getting pyramid equivalents per amount of food acquired*/
foreach var in d_total f_total f_juice f_citmlb f_other ///
g_total g_refined g_whole ///
pf_total pf_mps_total pf_seafd_hi pf_seafd_low pf_soy pf_nutsds pf_legumes pf_eggs ///
v_legumes v_total v_drkgr v_redor_total ///
energy carb satfat totfat monofat polyfat alcohol add_sugars a_drinks solid_fats
sodium {
gen grm_`var'=`var'*(hei_grams/100)
}

save "$hei\nutrients_wpyrvar_qs_fah.dta", replace


/*STEP 2--APPLY QUANTITIES TO FAFH NUTRIENT DATA*/
/*Data for FAFH data*/
use "$foodaps\faps_fafhnutrients.dta", clear

merge 1:1 hhnum eventid itemnum using "$hei\faps_eiq_fafhitem_puf.dta"
drop _m

/*getting pyramid equivalents per amount of food acquired*/
gen hei_grams=gramstotal if hei_grams==.

foreach var in d_total f_total f_juice f_citmlb f_other ///
g_total g_refined g_whole ///
pf_total pf_mps_total pf_seafd_hi pf_seafd_low pf_soy pf_nutsds pf_legumes pf_eggs ///
v_legumes v_total v_drkgr v_redor_total ///
energy carb satfat totfat monofat polyfat alcohol add_sugars a_drinks solid_fats
sodium {
gen grm_`var'=`var'*(hei_grams/100)
}

save "$hei\nutrients_wpyrvar_qs_fafh.dta", replace
```

Appendix 4a. Example STATA macro program—hei-2010 score.do

```
/*This do file creates HEI-2010 component densities and scores*/

gen monopoly=monofat+polyfat

gen addsugc=16*add_sugars
gen solfatc=9*solid_fats

gen maxalcgr=13*(energy/1000)
gen ethcal=7*alcohol
gen exalccal=7*(alcohol-maxalcgr)
replace exalccal=0 if alcohol<=maxalcgr

gen emptycal10=addsugc+solfatc+exalccal


gen vegden=legume_added_v_total/(energy/1000)
gen heix1_totalveg=5*(vegden/1.1)
replace heix1_totalveg=5 if heix1_totalveg>5
replace heix1_totalveg=0 if heix1_totalveg<0

gen grbnden=legume_added_v_drkgr/(energy/1000)
gen heix2_greens_and_bean=5*(grbnden/.2)
replace heix2_greens_and_bean=5 if heix2_greens_and_bean>5
replace heix2_greens_and_bean=0 if heix2_greens_and_bean<0

gen frtden=f_total/(energy/1000)
gen heix3_totalfruit=5*(frtden/.8)
replace heix3_totalfruit=5 if heix3_totalfruit>5
replace heix3_totalfruit=0 if heix3_totalfruit<0

gen wholefrt=f_total-f_juice
gen whfrden=wholefrt/(energy/1000)
gen heix4_wholefruit=5*(whfrden/.4)
replace heix4_wholefruit=5 if heix4_wholefruit>5
replace heix4_wholefruit=0 if heix4_wholefruit<0

gen wgrnden=g_whole/(energy/1000)
gen heix5_wholegrain=10*(wgrnden/1.5)
replace heix5_wholegrain=10 if heix5_wholegrain>10
replace heix5_wholegrain=0 if heix5_wholegrain<0

gen dairyden=d_total/(energy/1000)
gen heix6_totaldairy=10*(dairyden/1.3)
replace heix6_totaldairy=10 if heix6_totaldairy>10
replace heix6_totaldairy=0 if heix6_totaldairy<0

gen meatden=legume_added_allmeat/(energy/1000)
gen heix7_totprot=5*(meatden/2.5)
replace heix7_totprot=5 if heix7_totprot>5
replace heix7_totprot=0 if heix7_totprot<0

gen seaplden=legume_added_seaplant/(energy/1000)
gen heix8_seaplant_prot=5*(seaplden/.8)
replace heix8_seaplant_prot=5 if heix8_seaplant_prot>5
replace heix8_seaplant_prot=0 if heix8_seaplant_prot<0

gen faratio=monopoly/satfat if satfat>0
```

```
gen farmin=1.2
gen farmax=2.5
gen heix9_fattyacid=0 if satfat==0 & monopoly==0
replace heix9_fattyacid=10 if satfat==0 & monopoly>0
replace heix9_fattyacid=10 if faratio>=farmax & faratio !=.
replace heix9_fattyacid=0 if faratio<=farmin & faratio !=.
replace heix9_fattyacid=10*((faratio-farmin)/(farmax-farmin)) if faratio !=.

gen sodden=sodium/energy
gen sodmin=1.1
gen sodmax=2
gen heix10_sodium=10
replace heix10_sodium=0 if sodden>=sodmax
replace heix10_sodium=10-(10*(sodden-sodmin)/(sodmax-sodmin))

gen rgden=g_refined/(energy/1000)
gen rgmin=1.8
gen rgmax=4.3
gen heix11_refinedgrain=10
replace heix11_refinedgrain=0 if rgden>=rgmax
replace heix11_refinedgrain=10-(10*(rgden-rgmin)/(rgmax-rgmin))

gen sofa_perc=100*emptycal10/energy
gen sofamin=19
gen sofamax=50
gen heix12_sofaas=0 if sofa_perc>=sofamax
replace heix12_sofaas=20 if sofa_perc<=sofamin
replace heix12_sofaas=20-(20*(sofa_perc-sofamin)/(sofamax-sofamin))

foreach var in vegden grbnden frtden whfrden wgrnden dairyden meatden seaplden
faratio sodden rgden sofa_perc {
replace `var'=0 if `var'==.
}

foreach var in 1_totalveg 2_greens_and_bean 3_totalfruit 4_wholefruit 5_wholegrain
6_totaldairy 7_totprot 8_seaplant 9_fattyacid 10_sodium 11_refinedgrain 12_sofaas {
replace heix`var'=0 if energy==0
}

foreach var in 1_totalveg 2_greens_and_bean 3_totalfruit 4_wholefruit 5_wholegrain
6_totaldairy 7_totprot 8_seaplant 9_fattyacid 10_sodium 11_refinedgrain 12_sofaas {
replace heix`var'=0 if heix`var'<0 & heix`var'!=.
}

foreach var in 9_fattyacid 10_sodium 11_refinedgrain {
replace heix`var'=10 if heix`var'>10 & heix`var'!=.
}

replace heix12_sofaas=20 if heix12_sofaas>20 & heix12_sofaas!=.

gen hei2010_total_score=heix1_totalveg+heix2_greens_and_bean+heix3_totalfruit+ ///
heix4_wholefruit+heix5_wholegrain+heix6_totaldairy+heix7_totprot+heix8_seaplant ///
+heix9_fattyacid+heix10_sodium+heix11_refinedgrain+heix12_sofaas

label var hei2010_total_score "total hei-2010 score"
label var  heix1_totalveg "hei-2010 component 1 total vegetables"
label var  heix2_greens_and_bean "hei-2010 component 2 greens and beans"
label var heix3_totalfruit "hei-2010 component 3 total fruit"
label var heix4_wholefruit "hei-2010 component 4 whole fruit"
label var heix5_wholegrain "hei-2010 component 5 whole grains"
label var heix6_totaldairy "hei-2010 component 6 dairy"
label var heix7_totprot "hei-2010 component 7 total protein foods"
```

```
label var heix8_seaplant_prot "hei-2010 component 8 seafood and plant protein"
label var heix9_fattyacid "hei-2010 component 9 fatty acid ratio"
label var heix10_sodium "hei-2010 component 10 sodium"
label var heix11_refinedgrain "hei-2010 component 11 refined grains"
label var heix12_sofaas "hei-2010 component 12 sofaas calories"
label var vegden "density of mped total vegetables per 1000 kcal"
label var grbnden "density of mped of dark green veg and beans per 1000 kcal"
label var frtden "density of mped total fruit per 1000 kcal"
label var whfrden "density of mped whole fruit per 1000 kcal"
label var wgrnden "density of mped of whole grain per 1000 kcal"
label var dairyden "density of mped of dairy per 1000 kcal"
label var meatden "density of mped total meat/protein per 1000 kcal"
label var seaplden "denstiy of mped of seafood and plant protein per 1000 kcal"
label var faratio "fatty acid ratio"
label var sodden "density of sodium per 1000 kcal"
label var rgden "density of mped of refined grains per 1000 kcal"
label var sofa_perc "percent of calories from added sugar, solid fat, and alcohol"
```

Appendix 4b. Example STATA macro program—legumes.do

```
/* This program calculates legumes that get counted as meat and those that get
counted as veggies*/
/** This macro gets called into the program that calculates HEI 2010 scores**/
gen allmeat=pf_mps_total+pf_eggs+pf_nutsds+pf_soy
gen seaplant=pf_seafd_hi+pf_seafd_low+pf_nutsds + pf_soy

gen mbmax=2.5*(energy/1000)

gen needmeat=mbmax-allmeat if allmeat<mbmax
gen meatleg=4*v_legumes

/*Needs more meat, and all beans go to meat*/
gen all2meat=1 if meatleg<=needmeat /*folks who don't meet meat max and the amount
of legumes they consume is less than the amount they  need to reach mbmax*/
foreach var in allmeat seaplant {
gen legume_added_`var'=`var'+meatleg if all2meat==1
}
foreach var in v_total v_drkgr {
gen legume_added_`var'=`var' if all2meat==1
}

/*Needs more meat, and some beans go to meat, some go to veggies*/
gen meatveg=1 if meatleg>needmeat
gen extrmeat=meatleg-needmeat
gen extrleg=extrmeat/4

foreach var in allmeat seaplant {
replace legume_added_`var'=`var'+needmeat if meatveg==1 /*folks who don't meet
meat max and the amount of legumes they consume is more than the amount they  need
to reach mbmax--rest go to veggies*/
}
foreach var in v_total v_drkgr {
replace legume_added_`var'=`var'+extrleg if meatveg==1
}

gen all2veg=1 if allmeat>=mbmax /*Folks who meet the meat requirement so all
legumes count as veggies*/
foreach var in allmeat seaplant {
replace legume_added_`var'=`var' if all2veg==1
}
foreach var in v_total v_drkgr {
replace legume_added_`var'=`var'+v_legumes if all2veg==1
}
```

Appendix 5. Example STATA programs to calculate household level Healthy Eating Index-2010 scores, component densities, and component scores

```
/*THIS PROGRAM COMBINES FAH AND FAFH DATA AND CALCULATES WEEKLY HEI-2010 SCORE,
COMPONENT SCORES, AND COMPONENT DENSITIES FOR EACH HOUSEHOLD*/
capture log close

/*STEP 1--APPLY QUANTITIES TO FAH NUTRIENT DATA*/
global hei  *"hei" Should refer to folder where quantity data and HEI programs are
stored

use "$hei\nutrients_wpyrvar_qs_fah.dta", clear
append using "$hei\nutrients_wpyrvar_qs_fafh.dta"

/*Calculating total FPED by week */
collapse (sum) grm_* , by (hhnum)

foreach var in d_total f_total f_juice f_citmlb f_other ///
g_total g_refined g_whole ///
pf_total pf_mps_total pf_seafd_hi pf_seafd_low pf_soy pf_nutsds pf_legumes pf_eggs ///
v_legumes v_total v_drkgr v_redor_total ///
carb satfat totfat monofat polyfat alcohol add_sugars a_drinks solid_fats sodium
energy {
rename grm_`var' `var'
}

do "$hei\legumes.do"
       /*Pulls in legumes macro program to help figure out whether legume
       acquisitions should be counted as meat or veggies*/
do "$hei\hei-2010 score.do"/*Calling in macro to calculates HEI-2010 scores*/

order hhnum
sort hhnum

save "$hei\hei_scores_by_week.dta", replace
```