Westat®

Improving lives through research™

Review of the FoodAPS 2012 Imputation Approaches for Income and Price Data

Authors

Jianzhu Li
Wendy Van de Kerckhove
Tom Krenzke

December 30, 2016

Prepared for:
Economic Research Service
U.S. Department of Agriculture
355 E Street, SW
Washington, DC 20024-3221

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

Preferred citation:

Li, J., Van de Kerckhove, W., and Krenzke, T. (2016). *Review of the FoodAPS 2012 Imputation Approaches for Income and Price Data.* Prepared for the Economic Research Service, U.S. Department of Agriculture. Washington, D.C.

This report is part of a series of five reports. The citations for the other reports are as follows:

Krenzke, T., and Kali, J. (2016). *Review of the FoodAPS 2012 Sample Design.* Prepared for the Economic Research Service, U.S. Department of Agriculture. Washington, D.C.

Maitland, A., and Li, L. (2016). *Review of the Completeness and Accuracy of FoodAPS 2012 Data.* Prepared for the Economic Research Service, U.S. Department of Agriculture. Washington, D.C.

Petraglia, E., Van de Kerckhove, W., and Krenzke, T. (2016). *Review of the Potential for Nonresponse Bias in FoodAPS 2012.* Prepared for the Economic Research Service, U.S. Department of Agriculture. Washington, D.C.

Yan, T., and Maitland, A. (2016). *Review of the FoodAPS 2012 Instrument Design, Response Burden, Use of Incentives, and Response Rates.* Prepared for the Economic Research Service, U.S. Department of Agriculture. Washington, D.C.

Table of Contents

Tables **Page**

Figures **Page**

Executive Summary

The 2012 National Household Food Acquisition and Purchase Survey (FoodAPS) (hereafter referred to as "FoodAPS-1") is a household survey fielded primarily in 2012 and designed to capture detailed information on the food acquisitions of U.S. households. FoodAPS-1 was sponsored by the U.S. Department of Agriculture and managed by its Economic Research Service (ERS). In 2015, ERS contracted with Westat to conduct an independent assessment of the quality of the FoodAPS-1 sample design, instrumentation, data collection procedures, and resulting data. This report is part of a series of five reports that constitute that assessment.

This report presents an evaluation of the imputations that were performed for several variables, including income and prices for food items in the FoodAPS-1 data. A bottom-up approach was used to impute household income, which started with imputing each income component at the person level, and then obtaining the household income by aggregating person-level income across household members. The missing rates for the income components range from 1.5 percent to 4 percent, with monthly earnings having the highest missing rate. The purpose of imputing income components at the person level instead of imputing household income directly was to take advantage of the large pool of person-level characteristics, which served as candidate covariates in the imputation models. The imputed household-level income from aggregating imputed person-level income components was set to missing again if it was significantly less than household expenses. (This step of the imputation process was intended to maintain the consistency between total household income and expenses.) The missing value was re-imputed through a household-level model, and later re-distributed to six person-level income component variables.

For food-at-home (FAH) items, imputation was performed for missing item prices for both free and purchased items. Prices were missing for 7.6 percent of FAH items. For free items, respondents were not required to report price, so it was missing at a rate of 95.0 percent. For purchased items, the missing value rate was 5.5 percent, and missing values occurred when the receipt was not provided or was unreadable. To impute for missing price, items that had a Universal Product Code (UPC) appearing multiple times in the sample were assigned the mean price for that UPC at a similar location. A hot deck procedure was implemented otherwise. Some item prices were not imputed because of missing values among the imputation model variables.

Imputation was also used to fill in missing prices for purchased food-away-from-home (FAFH) items. Less than 1 percent of school items and approximately 8.8 percent of purchased non-school

items were missing prices after edits. School and non-school acquisitions were treated separately, with most school-item prices resolved through edits or deterministic methods. For non-school items, a two-step approach was used. First, the missing item prices were imputed as the median price of the non-zero sample prices within a cell defined by menu group, place type, food category, bundle indicator, and relative size. Then, the imputed prices were ratio adjusted so that the reported and imputed prices (plus tip where reported) summed to the total payment for the acquisition less 10 percent (for tax).

Missing data, if not accounted for properly, may lead to serious bias in the estimates derived from the survey data. Item nonresponse bias analyses (NRBA) were conducted on the items that have been imputed to understand the potential for bias and identify the variables most useful for imputation process. The analyses serve as the basis for understanding the missing data and provide insights on the use of an appropriate imputation strategy. The variables that were used in the NRBA of income (or item prices) include the variables that were used in the imputation models and also those that were not used in the imputation models but may be related to income (or item prices) or response status of income (or item prices). We computed the overall response rates, response rates by subgroups, and performed multivariate analyses such as logistic regressions and classification trees. The evaluations also focused on a few main concerns about the imputation of household income and food prices:

- General structure of the imputation model for income;
- Large variation in imputed values of food prices in the same stores;
- Remaining missing values in food prices after imputation;
- Imputation of food prices based on within-sample data only;
- Impact of outlying income values on imputation and analysis; and
- Using reported net earnings to estimate gross earnings.

We also compared the means of household income and food prices and their correlations with other auxiliary variables before and after imputation.

As a summary, the imputation of household income for FoodAPS-1 was done reasonably well. IVEware, the software that was used to perform single or multiple imputations of missing values, adopts the sequential regression imputation method and incorporates a large number of important predictors in the imputation model. Multiple imputation also allows the estimation of imputation

variance. It is appropriate to treat income as a mixed variable and impute it in two steps: first using a logistic regression to impute zero versus non-zero status, and then using a normal linear regression to impute non-zero values.

The original imputation strategy for income, however, could be improved from the following aspects:

- Account for household-level characteristics when imputing income poverty-level groups.
- Consider imputing before-tax and after-tax earnings simultaneously in the imputation process.
- Consider using total expenditure as a predictor in imputation models rather than using it as a lower bound to identify and re-impute unreasonable income values.
- Consider the inclusion of more household characteristics in the imputation model (e.g., sample target group and household size).
- Consider imputing income at the household level directly rather than aggregating from the person level.
- Consider other non-parametric approaches for imputing income. Income has a skewed distribution. Imputation models may easily be misspecified without doing appropriate diagnostic checks.

For FAH item prices, the predictor variables were related to price and the missingness of price, suggesting that the imputation process should have reduced bias in price estimates. We did not find any issues with the current imputation if it will be used in analyses that involve aggregate prices over different types of items, such as estimates of the total FAH expenditures for a household. The distribution of item prices was similar for reported and imputed values, and correlations with our analysis variables were preserved. However, if an analyst is interested in a particular type of item, such as cereal, then we found that the imputed values might not always be reasonable. Another drawback of the current imputation approach is that 3.2 percent of items still had a missing price after the imputation process since no imputation was done for items that had missing values for one of the predictor variables. We recommend the following for improving the FAH imputation process:

- As a first step, for items with a UPC appearing multiple times in the dataset, use the deterministic approach taken in FoodAPS-1, but use donors with the same value of QUANTITY when possible.

- Otherwise, if it is possible to match the item to an external database, such as household food purchase data from the market research company IRI (if feasible and accessible), impute using the price from the external database rather than just relying on within-sample data.

- For the remaining items, a hot deck approach can be used similar to that in FoodAPS-1, but first imputing for any missing values of the predictor variables and then using the results to impute for price.

For FAFH item prices, the analysis supported the choice of predictor variables, as they were again found to be related to price and whether price was missing. Our analysis of FAFH prices was limited since we did not have access to all the variables used in imputation. We also did not have information to distinguish between the result of the mean cell imputation and the ratio adjustment. We did not find any serious issues with the imputation, although we are unclear whether the low percentage of zero prices among the imputed values is a concern. For FAFH imputation, we recommend the following:

- Further review the discrepancy in the percentage of zero prices among reported and imputed items.

- Consider treating FAFH item price as a mixed variable and imputing in two steps: first, imputing for zero versus non-zero status, and then imputing for non-zero prices.

Overview 1

The 2012 National Household Food Acquisition and Purchase Survey (FoodAPS) (hereafter referred to as "FoodAPS-1") is a household survey fielded primarily in 2012 and designed to capture detailed information on the food acquisitions of U.S. households. FoodAPS-1 was sponsored by the U.S. Department of Agriculture (USDA) and managed by its Economic Research Service (ERS). In 2015, ERS contracted with Westat to conduct an independent assessment of the quality of the FoodAPS-1 sample design, instrumentation, data collection procedures, and resulting data. This report is part of a series of five reports that constitute that assessment.

As part of its processing of collected data, Mathematica Policy Research (Mathematica), referred to as the survey contractor thereafter, imputed prices for a number of food items with missing price information. It also imputed values for income components that respondents were unable to or unwilling to provide. Westat conducted an independent assessment of the procedures used by the survey contractor to impute values for missing data elements. This report provides the results of this component of the assessment.

The survey contractor performed imputations for several variables, including income and prices for food items in the FoodAPS-1 data. Westat evaluated the imputation approaches implemented in FoodAPS-1 for income and price data, and investigated the estimation of gross and net earnings. This technical report summarizes the steps Westat has taken for assessing the imputation and estimation approaches. Chapter 1 gives an overview of the imputation of missing income and food items, respectively. Chapter 2 highlights the main concerns on the imputation and estimation techniques. Chapter 3 presents the results from an item nonresponse bias analyses. Chapter 4 discusses the evaluation of imputation models.

It should be noted that Westat performed this analysis prior to the availability of revised final weights and, therefore, used the original adjusted household weights and revised variance estimation codes. It is not necessary to update the analyses with the new weights because the evaluation's focus is on the quality of imputation models and methodologies, and this assessment is not affected by the change in weights.

The contractor for FoodAPS-1 used a bottom-up approach to impute household income, which started with imputing each income component at the person level, and then obtaining the household income by aggregating person-level income across household members. The six income components are earnings, welfare and child support, retirement and disability income, unemployment insurance income, investment, as well as other income. The purpose of imputing income components at the person level, instead of imputing household income directly, was to take advantage of the large pool of person-level characteristics, which served as candidate covariates in the imputation models. The missing rates for the income components range from 1.5 percent to 4 percent, with monthly earnings having the highest missing rate.[1]

The imputation procedure for missing data on FoodAPS-1 household income took four steps:

- Step 1 – Impute household-level covariates using a single imputation through IVEware (http://www.isr.umich.edu/src/smp/ive/) and impute household income poverty-level group using simple random imputation. This missing values in income group was imputed based on empirical distribution of the observed values.
- Step 2 – Impute person-level covariates and six income component variables using multiple imputation through IVEware.
- Step 3 – Aggregate income components to person-level income and then to household-level income, and set the household income to missing if it is significantly less than household expenditures.
- Step 4 – Impute household income for any case that was set to missing in Step 3 through a household-level model, and re-distribute it to six person-level income component variables. Details about the imputation approaches and steps can be found in Appendix A.

For food-at-home (FAH) items, imputation was performed for missing item prices for both free and purchased items. Prices were missing for 7.6 percent of FAH items. For free items, respondents were not required to report price, so it was missing at a rate of 95.0 percent. For purchased items, the missing value rate was 5.5 percent, and missing values occurred when the receipt was not provided or was unreadable. To impute for missing price, items that had a UPC appearing multiple times in the sample were assigned the mean price for that UPC at a similar location. A hot deck procedure was implemented otherwise. Some item prices were not imputed because of missing

[1] The missing rates account for "don't know" and refusals. The imputation rates may be higher than those since some zeroes were viewed as passive nonresponse and imputed.

values among the variables used in the imputation process. More details of the imputation can be found in Appendix B.

Imputation also was used to fill in missing prices for purchased food-away-from-home (FAFH) items. Less than 1 percent of school items and approximately 8.8 percent of purchased non-school items were missing prices after edits. School and non-school acquisitions were treated separately, with most school-item prices resolved through edits or deterministic methods. For non-school items, a two-step approach was used. First, the missing item prices were imputed as the median price of the non-zero sample prices within a cell defined by menu group, place type, food category, bundle indicator, and relative size. Then the imputed prices were ratio adjusted so that the reported and imputed prices (plus tip where reported) summed to the total payment for the acquisition less 10 percent (for tax). Further information can be found in Appendix C.

2 Main Concerns

Missing data, if not accounted for properly, may lead to serious bias in empirical results derived from the survey data. Unit nonresponse was taken into account in the weighting process. Imputation was, therefore, used to address the item nonresponse. The main topics investigated herein include:

- Bias due to item nonresponse;
- General structure of the imputation model for income;
- Large variation in imputed values of food prices in the same stores;
- Remaining missing values in food prices after imputation;
- Imputation of food prices based on within-sample data only;
- Impact of outlying income values on imputation and analysis; and
- Using reported net earnings to estimate gross earnings.

The evaluations were conducted using the following data files:

- Household file
- Household imputation file
- Individual file
- Individual imputation file
- FAH items, FAFH items, FAH event, FAH item IRI,[2] and FAFH event files

The evaluation results are described in the following chapters.

[2] IRI is a market research company from which ERS obtained data on household food purchases and retail food sales.

Item Nonresponse Bias Analysis 3

The survey contractor imputed the values of income components and item prices that respondents were unable to or unwilling to provide. Westat conducted an independent item nonresponse bias analyses (NRBA) on the items that were imputed by the survey contractor to understand the potential for bias and identify the variables most useful for the imputation process. The analyses serve as the basis for understanding the missing data and provide insights on the use of an appropriate imputation strategy. WesNRBA, Westat proprietary software, that is written in the form of a SAS macro, was used to conduct all NRBA listed above. This chapter provides the results of this assessment.

3.1 Identification of Variables for Analyses

The variables that were used in the NRBA of income and item prices include the following types:

- Variables that were used in the original imputation models. For income, these are at the household level and person level. For item price, these are at the item level and event level.
- Variables that were not used in the imputation models but may be related to income or item price.
- Variables that were not used in the imputation models but may be related to response status of income or item price.

3.1.1 Income

For income, the variables used in the NRBA were collected from the screener, the Initial and Final Interviews, the sampling frame, or the sample design process. Table 3-1 provides a list of categorical covariates that were used in Westat's income NRBA analyses. The last two columns in the table indicate whether a variable was used as a predictor in the survey contractor's person-level imputation models and whether the variable was defined at the household level or the person level. These variables were missing for less than 5 percent of all records. The majority of them have zero or very few missing values. The variables in Table 3-1 were imputed by the survey contractor, if necessary, to have no or few missing values during Step 1 and Step 2 of their income imputation process.

Table 3-1. FoodAPS variables used in income NRBA

Variable name	Description	Values	Used as predictor in person-level imputation models	Household or person level
HHpd_earn	Derived indicator variable for whether the household indicated earnings as a source of income at either the screening or at the Final Interview, propagated to the person level	1 = Yes 0 = No	Yes	Household
HHpd_unem	Derived indicator variable for whether the household indicated unemployment compensation as a source of income at either the screening or at the Final Interview, propagated to the person level	1 = Yes 0 = No	Yes	Household
HHpd_reti	Derived indicator variable for whether the household indicated retirement/disability as a source of income at either the screening or at the Final Interview, propagated to the person level	1 = Yes 0 = No	Yes	Household
HHpd_inve	Derived indicator variable for whether the household indicated investment as a source of income at either the screening or at the Final Interview, propagated to the person level	1 = Yes 0 = No	Yes	Household
HHpd_welf	Derived indicator variable for whether the household indicated welfare as a source of income at either the screening or at the Final Interview, propagated to the person level	1 = Yes 0 = No	Yes	Household
HHpd_oth	Derived indicator variable for whether the household indicated other sources of income at either the screening or at the Final Interview, propagated to the person level	1 = Yes 0 = No	Yes	Household
Relationr	Whether or not the person is respondent/spouse or partner	1 = Primary respondent or spouse or partner 0 = Other	Yes	Person
age_grp	Age group	0 if age ≥ 0 and age < 16 1 if age ≥ 16 and age ≤ 24 2 if age ≥ 25 and age ≤ 44 3 if age ≥ 45 and age ≤ 64 4 if age ≥ 65	Yes	Person

Table 3-1. FoodAPS variables used in Income NRBA (continued)

Variable name	Description	Values	Used as predictor in person-level imputation models	Household or person level
Sexr	sex	1 = Male 2 = Female	Yes	Person
Racer1	Race = White	1 = Yes 0 = No	Yes	Person
Racer2	Race = Black	1 = Yes 0 = No	Yes	Person
Racer3	Race = Amer Indian/Alaskan Native	1 = Yes 0 = No	Yes	Person
Racer4	Race = Asian	1 = Yes 0 = No	Yes	Person
Racer5	Race = Hawaiian/Pacific Islander	1 = Yes 0 = No	Yes	Person
Racer6	Race = Other	1 = Yes 0 = No	Yes	Person
Educr	Education level	1 = High school or less 2 = High school graduate or equivalent 3 = Some college 4 = College graduate 5 = More than college	Yes	Person
Hispr	Hispanic origin	1 = No 2 = Yes	Yes	Person
Maritalr	Marital status	1 = Married 2 = Divorced, widowed, separated 3 = Never married	Yes	Person
Healthr	Health status	1 = Excellent 2 = Very good 3 = Good 4 = Fair 5 = Poor	Yes	Person
RBMIr	BMI weight category	1 = Not overweight 2 = Overweight 3 = Obese	Yes	Person
Smoker	Smoke/chew tobacco	1 = Yes 0 = No	Yes	Person
Workr	Work status	1 = Working at a job or business 2 = With a job or business but not at work 3 = Looking for work 4 = Not working at a job or business	Yes	Person

Table 3-1. FoodAPS variables used in Income NRBA (continued)

Variable name	Description	Values	Used as predictor in person-level imputation models	Household or person level
Uatype	Urban area type	1 = Urbanized area 2 = Urban cluster 3 = Neither	Yes	Household
Metromicro	Metro- or micro-area	1 = Metro 2 = Micro 3 = Neither	Yes	Household
Anyfinprobs	Indicator variable for whether the household has any financial problems	1 = Yes 0 = No	Yes	Household
Workfarm	Migrant or seasonal farm worker	1 = Yes 0 = No	Yes	Household
Workselfemploy	Self-employment status	1 = Yes 0 = No	Yes	Household
Foodpantry_imp	Past 30 days–Food pantry or bank	1 = Yes 0 = No	Yes	Household
Rsnapnow	Receiving SNAP at time of survey	0 = No 1 = Yes 2 = Match confirms SNAP nonparticipation	Yes	Household
Q11	Currently receive SNAP?	1 = Yes 0 = No, missing value (./.B/.D/.R)	Yes	Household
Finances	HH financial condition	1 = Very comfortable and secure 2 = Able to make ends meet without much difficulty 3 = Occasionally have some difficulty making ends meet 4 = Tough to make ends meet but keeping your head above water 5 = In over your head	Yes	Household
Liqassets2000_imp	Income-F8b-$2,000 or more liquid assets	1 = Yes 0 = No	Yes	Household
Liqassets3000_imp	Income-F8b-$3,000 or more liquid assets	1 = Yes 0 = No	Yes	Household
Rfoodsecscore	Food Security Score	From 0 to 10, categorical	Yes	Household
Auto_imp	Own/lease car or truck	0 = No 1 = Yes, own 2 = Yes, Lease 3 = Own and lease	Yes	Household

Table 3-1. FoodAPS variables used in Income NRBA (continued)

Variable name	Description	Values	Used as predictor in person-level imputation models	Household or person level
Housingtype	Rent or own home	1 = Rent 2 = Own 3 = Other, do not pay for housing	Yes	Household
Healthycost_imp	Costs too much to eat healthy foods	1 = Agree 2 = Disagree	Yes	Household
Healthytime_imp	Too busy to prepare healthy foods	1 = Agree 2 = Disagree	Yes	Household
Healthytaster	Respondent thinks healthy foods don't taste good	1 = Agree 2 = Disagree	Yes	Household
Billsreview	Reviews bills	1 = Never 2 = Rarely 3 = Sometimes 4 = Usually 5 = Always 6 = Not applicable	Yes	Household
Billspay	Pays bills on time	1 = Never 2 = Rarely 3 = Sometimes 4 = Usually 5 = Always 6 = Not applicable	Yes	Household
Billspayabovemin	Pays more than minimum on credit card	1 = Never 2 = Rarely 3 = Sometimes 4 = Usually 5 = Always 6 = Not applicable	Yes	Household
Anyjobchange	HH member changed job in past 3 months	1 = Yes 0 = No	Yes	Household
Anyillness	Illness/disability in past 3 months	1 = Yes 0 = No	Yes	Household
Initiallang	Initial Interview language	1 = English 2 = Spanish 3 = Korean	Yes	Household
Finallang	Final Interview language	1 = English 2 = Spanish 3 = Korean	Yes	Household

Table 3-1. FoodAPS variables used in Income NRBA (continued)

Variable name	Description	Values	Used as predictor in person-level imputation models	Household or person level
Targetgroup[1]	Sampling target group used for weight construction (post stratification and trimming)	1 = NonSNAP household, with income <100% of the Federal Poverty Guideline 2 = NonSNAP household, with income >=100% and <185% of the Federal Poverty Guideline 3 = NonSNAP household, with income >=185% of the Federal Poverty Guideline 4 = SNAP household	No	Household
Region	Census region	1 = Midwest 2 = Northeast 3 = South 4 = West	No	Household
FNS_Region	Food and Nutrition Service Region	1 = Mid-Atlantic 2 = Midwest 3 = Mountains/Plains 4 = Northeast 5 = Southeast 6 = Southwest 7 = West	No	Household
Hhsizer	Household size	1 = 1 person 2 = 2 persons 3 = 3 persons 4 = 4 persons 5 = 5 persons 6 = 6 persons 7 = 7 or more persons	No	Household
Wichhr	Is anyone in household receiving benefits from WIC?	1 = yes 0 = No	No	Household

[1] The analysis in this report was performed prior to the revision of the sampling domain (TARGETGROUP) values in March 2016. That is, TARGETGROUP is derived using reported values of income from the Initial Interview, and using the initial imputed income values conducted as part of the weighting process. Since the evaluation, a new target group variable was derived using the multiply imputed values. An "r" at the end of a variable name means that the original variable was recoded for the imputation analysis.

3.1.2 FAH Price

The NRBA for FAH item price looked at the relationship between whether price is missing prior to imputation and characteristics of the item, event, and household. The characteristics included variables that were used in the survey contractor's imputation process and additional variables that were not used in imputation.

The NRBA variables used in the survey contractor's imputation process were:

- Place type (PLCTYP_R, derived);
- Package size unit (PKGSZUNT, derived); and
- IRI department (IRI_DEPT).

The NRBA variables not used in imputation were:

- Quantity of the item (collapsed QUANTITY);
- Source of barcode (BARCODESOURCE);
- Total number of items associated with an event (collapsed ITEMSTOT);
- Whether a loyalty card was used (LOYALTYCARD);
- Whether respondent paid with cash (CASH);
- Whether respondent paid with Supplemental Nutrition Assistance Program (SNAP) electronic benefit transfer (EBT) (EBT_SNAP);
- Household size (collapsed HHSIZE);
- Sampling domain (TARGETGROUP);[3]
- WIC household indicator (WICHH);
- Financial condition (FINCONDITION);
- Food and Nutrition Service (FNS) region (REGION, derived from ST, which identifies state of residence);

[3] The analysis in this report was performed prior to the revision of the sampling domain (TARGETGROUP) values in March 2016. That is, TARGETGROUP is derived using reported values of income from the initial interview, and using the initial imputed income values conducted as part of the weighting process. Since the evaluation, a new target group variable was derived using the multiply imputed values.

- County-level percentage that have low access to store (quartiles); and
- County-level percentage that are low income and low access (quartiles).

The exact place type and package size unit variables used in imputation were not available, so they were derived based on the descriptions in a technical memorandum.[4] The county-level access variables were obtained from the USDA Food Environment Atlas data at http://www.ers.usda.gov/data-products/food-environment-atlas/data-access-and-documentation-downloads.aspx.

The primary sampling unit (PSU), IRI aisle, and package size or weight were also used as sorting variables in the imputation, but they are not included in the NRBA because they have too many categories to use as auxiliary variables for this analysis. The other variables were selected because it was hypothesized that they might be related to price or the missingness of price. Food prices are external to the household (set by the market), so household variables will likely be of only limited value in imputing. However, it was hypothesized that household income could be related to price, in that higher income households may be willing to pay more for the same item (e.g., shop at the closest store even if it is a bit more expensive). In addition, household size could be relevant, in that larger households can buy in bulk, although this might already be captured by the item size/weight variable. Access variables were also considered, as this implies a shortage of supply.

Only auxiliary variables with less than 5 percent of missing values were considered for this analysis. There were two exceptions: package size unit, since it was used in the imputation, and whether a loyalty card was used, since it seemed an important indicator of price and did not greatly exceed 5 percent missing. Ideally, we would impute for missing values of the analysis variables, but this was not feasible given the scope of this task. Missing values of an auxiliary variable were handled differently depending on the analysis. For response rates by subgroup, missing values were a separate subgroup. The chi-square tests of independence between response status and the auxiliary variable excluded any observations with a missing value for that variable. Otherwise, the test could just indicate whether a missing value in the auxiliary variable is related to a missing value of price, which is not of interest. For the classification tree analysis, the R procedure (rpart) treated missing values differently depending on whether the auxiliary variable was continuous or categorical. For categorical variables, it treated it as a separate category. For continuous variables, it used a surrogate variable to make the split for observations where the split variable was missing.

[4] This internal technical memorandum, titled "The National Household Food Acquisition And Purchase Survey – Food-At-Home Items Documentation," was prepared by Cole and Baxter from Mathematica in 2014.

3.1.3 FAFH Price

The NRBA for FAFH item price looked at the relationship between whether price is missing and characteristics of the item, event, and household. The characteristics included variables that were used in the survey contractor's imputation process and additional variables that were not used in imputation.

NRBA variables used in the survey contractor's imputation process were:

- Menu group (MENUGRP);
- Type of place (PLACEGROUP, derived);
- Indicator of whether the item was bundled (BUNDLED, 1 if 1<=BUNDLETYPE<6; 0 otherwise); and
- Relative beverage size (BEVSIZE, derived).

NRBA variables not used in imputation were:

- Quantity (collapsed QUANTITY);
- Type of food book that contained the acquisition (BOOKTYPE);
- Number of household members who ate the meal(s) (collapsed NUMHHPEOPLE);
- Whether the meal was for breakfast (BREAKFAST);
- Whether the meal was for lunch (LUNCH);
- Whether the meal was for dinner (DINNER_SUPPER);
- Whether the meal was a snack or drink (SNACK_DRINK);
- Whether respondent paid with cash (CASH);
- Household size (collapsed HHSIZE);
- Sampling domain (TARGETGROUP[5]);

[5] The analysis in this report was performed prior to the revision of the sampling domain (TARGETGROUP) values in March 2016. That is, TARGETGROUP is derived using reported values of income from the initial interview, and using the initial imputed income values conducted as part of the weighting process. Since the evaluation, a new target group variable was derived using the multiply imputed values.

- Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) household indicator (WICHH);
- Financial condition (FINCONDITION); and
- FNS region (REGION, derived from ST, which identifies state of residence).

Type of place and relative beverage size were derived based on the description in a technical memorandum.[6] In imputation, relative size was used for beverages and chicken; however, the food code used to identify chicken was not on the dataset we received, so we could use relative size only for beverages in our analysis. Food category (the 4-digit food group from What We Eat in America [WWEIA]) was also used to form the imputation cells, but it was not available on the dataset, so it was not used in this analysis. It would also likely have too many categories for the purpose of the item NRBA.

Only auxiliary variables with less than 5 percent missing values were considered for this analysis. The one exception was beverage size, but this was included because it was used in imputation.[7] The treatment of missing values followed that for the FAH item NRBA, as was described in Section 3.1.2.

3.2 NRBA Results for Income

At the individual level, the survey contractor used IVEware to conduct multiple imputation and generate five sets of imputed values for each income component. In this section, the analysis results were computed using the first out of the five multiply-imputed values. Nonrespondents were defined as those who had imputation flags as 1 (e.g., RINCEARN = 1 identifies missing/imputed monthly income from earnings). This definition of missingness accounts for passive refusals and missing data due to other reasons; for example, inconsistency between zero household income and non-zero total household expenses, and inconsistent reported retirement income at household level and individual level.

[6] This internal technical memorandum, titled "The National Household Food Acquisition And Purchase Survey – Food-Away-From-Home Items Documentation," was prepared by Cole et al. from Mathematica in 2015.

[7] For imputation, if beverage size was missing, it was assumed to be size medium.

3.2.1 Overall Item Response Rates

We computed item response rates as the ratio of the number of respondents for whom an in-scope response was obtained, to the number of respondents who were asked to answer that item. Weighted response rates (or imputation rates) for each income component are shown in Table 3-2. The missing rates are about 5 to 6 percent for most of the income components. Retirement income has the lowest response rate at 91.4 percent.

Table 3-2. Weighted item response rates for six income components based on the first version of imputed values

Income component	Response rate (%)
Earnings	93.4
Unemployment	94.5
Retirement	91.4
Welfare	94.6
Investment	93.6
Other	94.3

3.2.2 Item Response Rates by Subgroups

Response rates were also computed by subgroups. The subgroups were defined by items from the survey questionnaire that have high response rates or other key variables available for the unit respondents. Rao-Scott Chi-square tests[8] were used to detect a significant relationship between the item response indicator and the analysis variable of interest. The categorical variables that were identified in Table 3-1 were used in this analysis. Table 3-3 summarizes the significant relationship between the response indicator for each income component and subgroup variable. Detailed response rates for each income component and by subgroups can be found in Appendix D. For example, in Table D-1, for the subgroup defined as "Retirement/disability as a source of household income = No", the weighted response rate for person earnings is 92.4 percent.

[8] Rao, J.N.K., and Scott, A.J. (1981). The analysis of categorical data from complex sample surveys: chi-squared tests for goodness-of-fit and independence in two-way tables. *Journal of the American Statistical Association*, 76, 221–230.

Table 3-3. Subgroups with significantly different item response rates

Subgroup	Earnings	Unemployment	Retirement	Welfare	Investment	Other
Earnings as a source of household income	*	✓		✓	✓	✓
Unemployment compensation as a source of household income	*	*				*
Retirement/disability as a source of household income	✓	✓	✓	✓	✓	✓
Investment as a source of household income						
Welfare as a source of household income					*	
Other source of household income			✓			✓
Relation	✓		✓			
Age group						
Sex	✓	✓	✓	✓	✓	✓
White						
Black						
American Indian/Alaskan Native					✓	
Asian	*					
Hawaiian/Pacific Islander		*	*	*	✓	*
Other race						
Education level					✓	*
Hispanic origin						
Marital status						
Health status						
BMI weight category				*	✓	*
Smoke/chew tobacco	*	✓	*		✓	✓
Work status		✓	✓	✓	✓	*
Urban area type	*	✓		✓		✓
Metro- or micro-area	*					
Household has any financial problems	✓	✓		✓	✓	✓
Migrant or seasonal farm worker						
Self-employment status						

Table 3-3. Subgroups with significantly different item response rates (continued)

Subgroup	Earnings	Unemployment	Retirement	Welfare	Investment	Others
Past 30 days–Food pantry or bank	✓					
Receiving SNAP at time of survey	*	*	✓	*		
Currently receive SNAP?			✓			
Household financial condition					✓	✓
$2,000 or more liquid assets			✓			
$3,000 or more liquid assets			*			
Food Security Score					✓	
Own/lease car or truck		✓	✓	*	*	
Rent or own home						
Costs too much to eat healthy foods						
Too busy to prepare healthy foods						
Respondent thinks healthy foods don't taste good						
Reviews bills						
Pays bills on time		*		*	✓	
Pays more than minimum on credit card					✓	*
Household member changed job in past 3 months						
Illness/disability in past 3 months						
Initial Interview language		*		✓		*
Final Interview language						
Sampling target group	✓	✓	✓	✓	✓	✓
Census region					*	
Food and Nutrition Service Region						
Household size						
Is anyone in household receiving benefits from WIC?						

Note:

✓ Denotes significant at 5 percent confidence level.

* Denotes marginally significant at 10 percent confidence level.

The response rates are significantly different among subgroups defined by whether the household indicated retirement/disability as a source of income, sex, and sample target group for all the income components. The sample target group was created based on the household income reported in the Final Interview. When reported income was zero, income was imputed by the survey contractor as the mean in the secondary sampling unit (SSU) by household size (or entire SSU if no donor existed). This variable itself was not used in the imputation process. However, a related variable, household income poverty group, was used. The poverty group variable was reported at screener and its missing values were imputed using a simple random imputation approach. This variable was used to restrict the lower and upper boundaries for imputed income. The issues with poverty groups are discussed in Section 4.1. Other covariates that are significantly related to differential subgroup response rates include whether the household indicated monthly earnings as a source of income, if a smoker or not, work status, if the household has any financial problems, etc. (see Table 3-3). The results also show that the household-level characteristics impact the response status more than the person-level characteristics, which indicates that doing imputation at the household level may have been a feasible and reasonable alternative to the approach used.

3.2.3 Multivariate Analysis

The response rates by subgroup in the previous section were useful in evaluating the relationship between income response status and each auxiliary variable individually. To account for potential relationships among the auxiliary variables, multivariate analyses of item nonresponse were conducted using logistic regressions. The dependent variable is the item response status for each income component and the candidate independent variables are the key characteristics that are available for both respondents and nonrespondents (i.e., the variables in Table 3-1). A stepwise selection procedure is used to choose the predictors to be included in the final regression model. The significance criterion for entering the model based on a Chi-square score statistic is 0.1, and the significance criterion for staying in the model based on a Wald Chi-square is 0.05. The significant predictors in logistic regressions are presented in Table 3-4. Variables that are highly correlated with the response status should be considered as candidates in the imputation process in order to reduce nonresponse bias. Moreover, variables that highly related to the income variables should also be considered for the study. This is explored further in later sections of this report. The target sampling group and FNS region are significant predictors for several income components but were not used as predictors in the imputation models by the survey contractor.

Table 3-4. Variables included in logistic regression models

Variables	Earnings	Unemployment	Retirement	Welfare	Investment	Others
Earnings as a source of household income	✓	✓	*	✓	✓	✓
Unemployment compensation as a source of household income	✓	✓		*		✓
Retirement/disability as a source of household income	✓	✓	✓	✓	✓	✓
Investment as a source of household income		*	*	*		
Welfare as a source of household income	*					
Other source of household income			✓			✓
Relation	✓	✓	*	✓	*	✓
Age group	✓	✓		✓	✓	✓
Sex	✓	✓	*	✓		✓
White						
Black						
American Indian/Alaskan Native						
Asian				*		
Hawaiian/Pacific Islander						
Other race						
Education level	*	*	✓	*	✓	*
Hispanic origin						
Marital status						
Health status	*					
BMI weight category						
Smoke/chew tobacco			*			*
Work status	*	✓	✓	✓	✓	✓
Urban area type	✓	*		✓	✓	✓
Metro- or micro-area	✓	*				
Household has any financial problems	✓	✓	*	✓	✓	✓
Migrant or seasonal farm worker						
Self-employment status		*	*	*	*	*
Past 30 days–Food pantry or bank	✓		✓			
Receiving SNAP at time of survey						

Table 3-4. Variables included in logistic regression models (continued)

Variables	Earnings	Unemployment	Retirement	Welfare	Investment	Others
Currently receive SNAP?						
Household financial condition			*	✓	✓	*
$2,000 or more liquid assets	*					*
$3,000 or more liquid assets			✓			
Food Security Score	✓	✓	✓			
Own/lease car or truck		✓	✓	✓		✓
Rent or own home						
Costs too much to eat healthy foods						
Too busy to prepare healthy foods						
Respondent thinks healthy foods don't taste good						
Reviews bills	*	✓				
Pays bills on time			*			✓
Pays more than minimum on credit card	✓	*	✓	✓	✓	✓
Household member changed job in past 3 months						
Illness/disability in past 3 months		*	*	*	*	*
Initial Interview language		✓		✓	✓	
Final Interview language			✓			
Sampling target group	✓	✓	✓	✓	✓	✓
Census region			✓			
Food and Nutrition Service Region	*	*	✓		*	*
Household size	*					
Is anyone in household receiving benefits from WIC?						*

Note:

✓Denotes predictors that are highly significant (the p-values for the Wald Chi-square test is less than 0.05).

* Denotes predictors that were selected into models but are not highly significant.

3.3 NRBA Results for FAH Prices

The item NRBA for FAH price focused on purchased items.[9] Free items (FREE = 1) and items imputed using the multiple-UPC deterministic method (IMPUTEMETHOD = 1, 2, or 3) were excluded from the analysis.[10] Respondents were considered items with a non-missing price (non-missing TOTITEMEXPNOCOUPONS). Nonrespondents were the items with missing price values that were imputed using hot deck or not imputed.

3.3.1 Response Rates Overall and By Subgroup

A response rate analysis was performed to evaluate nonresponse bias in FAH item price. The analysis was similar to the one described in Section 3.2. The results are presented in Table E-1 in Appendix E. A total of 138,855 items were included in this analysis. The overall weighted response rate (i.e., weighted percentage of items with non-missing price) was 95.4 percent. The three variables used in imputation were significantly related to response status, with more missing prices for items: that were not obtained in a superstore or supermarket; that were measured in pounds, dry ounces, or grams; or that were from the liquor department. This supports the use of these variables for imputation, to the extent that they are related to price (see Section 4.2).

Several other item and event characteristics are associated with missing prices, with a higher percent missing when the quantity of the item was greater than one, the barcode was scanned from a non-UPC barcode on the item or a Food Book barcode, fewer items were purchased at the event, a loyalty card was not used, the purchase was paid for with cash, and the purchase was not paid for with SNAP EBT.

Household-level characteristics were not used in the imputation of item price. This analysis indicates that they are not as strongly associated with missing price as item-level or event-level characteristics, although there are moderately higher rates of missing prices among WIC households and non-SNAP households below poverty.

---

[9] Since the reason for a missing price value differs for free and purchased items, the two types of items should be analyzed separately. Only purchased items were considered here, given the scope of this task. However, it could also be informative to analyze whether the characteristics of free items differ from those of purchased items. If the auxiliary variables related to a purchased items having a missing price differ from those related to an item being free, it could indicate a benefit in using different predictor variables for the two sets of cases.

[10] The item NRBA helps inform the best variables to use in the hot deck imputation model. Therefore, items where the price could be imputed deterministically based on the UPC code were not included here.

3.3.2 Multivariate Analysis

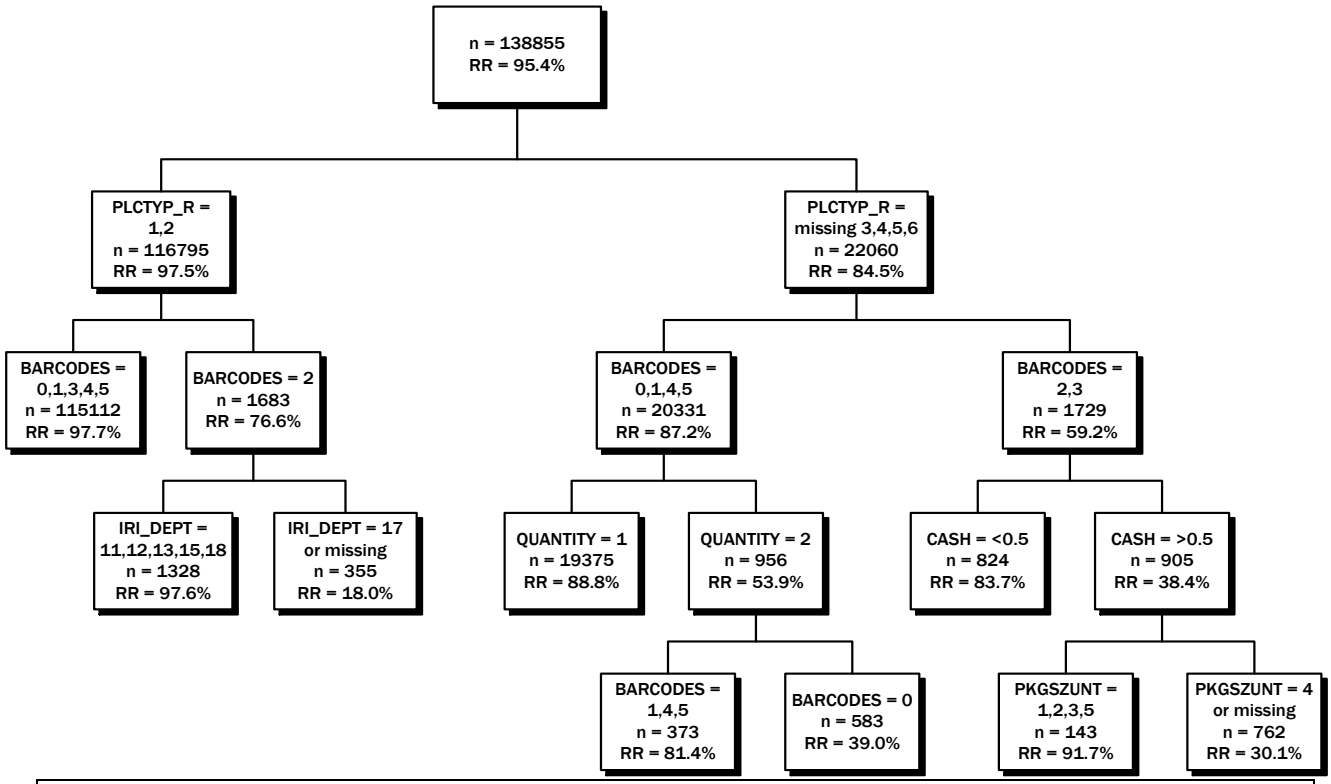
The analysis in Section 3.3.1 is useful in evaluating the relationship between price response status and each auxiliary variable individually but does not account for potential relationships among the auxiliary variables. To incorporate the interactions, a classification tree analysis was run using the rpart package in the R software. The classification tree identifies the domains with the most differential response rates, as defined by combinations of the auxiliary variables. The weights for this analysis were scaled to sum to the sample size so that significance would not be overstated, given that the procedure does not account for the complex sample design. The minimum cell size was set to 30, and the threshold complexity parameter was specified as 0.01.

Figure 3-1 shows the resulting classification tree from the multivariate analysis of the relationship between price response status and auxiliary characteristics. Although the focus of this analysis was to identify the subgroups with the lowest response rates, it is standard practice to report error rates associated with the trees. For 10 cross validations, the cross-validated error rate was 4.1 percent. Place type is the primary indicator of missing FAH item price and defines the first split in the tree, with place types of 1 or 2 (superstores or supermarkets) having higher response rates than other place types. Within place type, the barcode source is the next most significant predictor.

Overall, the highest response rate was found for items purchased in superstores or supermarkets that either did not have a barcode or had a UPC or Food Book barcode (as opposed to some other barcode). Approximately 97.7 percent of such items had a non-missing price. At the other end, only 38.4 percent of prices were non-missing for items that were not purchased in a superstore, supermarket, or place type was unknown; had a non-UPC barcode on the item or a Food Book barcode; and were paid for in cash.[11] As with the analysis in Section 3.3.1, these results support the use of place type in imputation. The source of the barcode and whether the payment was made in cash were not used in imputation but are shown to be significantly related to price response status within place type. Including these two variables may improve the imputation, depending on their relationship to price. This is explored further in Section 4.2.

[11]There were two cells with lower response rates. However, for one, the split was primarily based on whether the IRI descriptor IRI_DEPT was missing or not. Within the subgroup, if the IRI department is missing, the price was more likely to be missing. The split based on PKGSZUNT had a similar interpretation.

Figure 3-1. Classification tree for FAH item price



Variable	Values
PLCTYP_R	1=superstore, 2=supermarket, 3=convenience store or small grocery, 4=other food stores, 5=eating places, 6=all other
BARCODES	0=no barcode, 1=scanned UPC on item, 2=scanned other barcode on item, 3=scanned Food Book barcode, 4=assigned Food Book barcode, 5=extracted UPC from item description
IRI_DEPT	11=beverages, 12=frozen, 13=general food, 15=health, 17=liquor,18=refrigerated
QUANTITY	1=one, 2=more than one
CASH	0=no, 1=yes
PKGSZUNT	1=oz, liter, 2=lbs, dryoz, gram, 3=count, piece, 4=other, 5=not applicable

Note: In each cell, "n" is the unweighted sample size and "RR" is the weighted percentage of items with a non-missing price. Variable names are defined in Section 3.1.2; BARCODESOURCE is abbreviated as BARCODES in this figure.

3.4 NRBA Results for FAFH Prices

The item NRBA for FAFH focused on purchased, non-school items. Free items (FREE = 1) and school items (MENUID = 3) were excluded from the analysis. No imputation of price was done for free items. Only 52 school items required imputation, and of these, 36 were imputed using a different method than for other items (using in-sample median paid school meal price). Respondents were considered to be purchased non-school items that have a non-missing price and did not require

imputation (IMPCOSTMETHOD = .v). Nonrespondents were the purchased non-school items with missing price values or zero values that were imputed as non-zero.[12]

3.4.1 Response Rates Overall and By Subgroup

The results of the response rate analysis for FAFH item price are shown in Table F-1 in Appendix F. For the 59,893 items included in this analysis, the overall weighted response rate (i.e., weighted percentage of items with non-missing price) was 92.0 percent. The four variables used in imputation had the strongest relationship to response status (p values < .0001). There is a higher percentage of missing prices for food items (i.e., not beverages) that were not obtained from a top national chain (as defined by MENUGRP) and for items that were not part of a bundle.

Of the variables not used in imputation, there was a significant relationship between missing FAFH price to the number of household members that shared the meal, the type of meal, and region. A higher percentage of prices were missing when five or more household members shared the meal, the meal was dinner, the meal was not a snack or beverage, or the item was obtained in the Southwest, Northeast, or Southeast.

3.4.2 Multivariate Analysis

As was done with FAH price, a classification tree analysis was run to analyze the multivariate relationship between FAFH price response status and auxiliary characteristics. The minimum cell size was set to 30, and the threshold complexity parameter (cp) was specified as 0.00044, based on a review of a plot of the cp by the cross-validation error. The tree is provided in Figure 3-2. The cross-validated error rate was 7.8 percent with 10 cross validations. The tree divides the items into subgroups with response rates ranging from 96.4 percent to less than 10 percent, although the cells with less than a 10 percent response rate contain fewer than 50 items each. The first split in the tree is based on whether the item was bundled. Within non-bundled items, the next split is based on

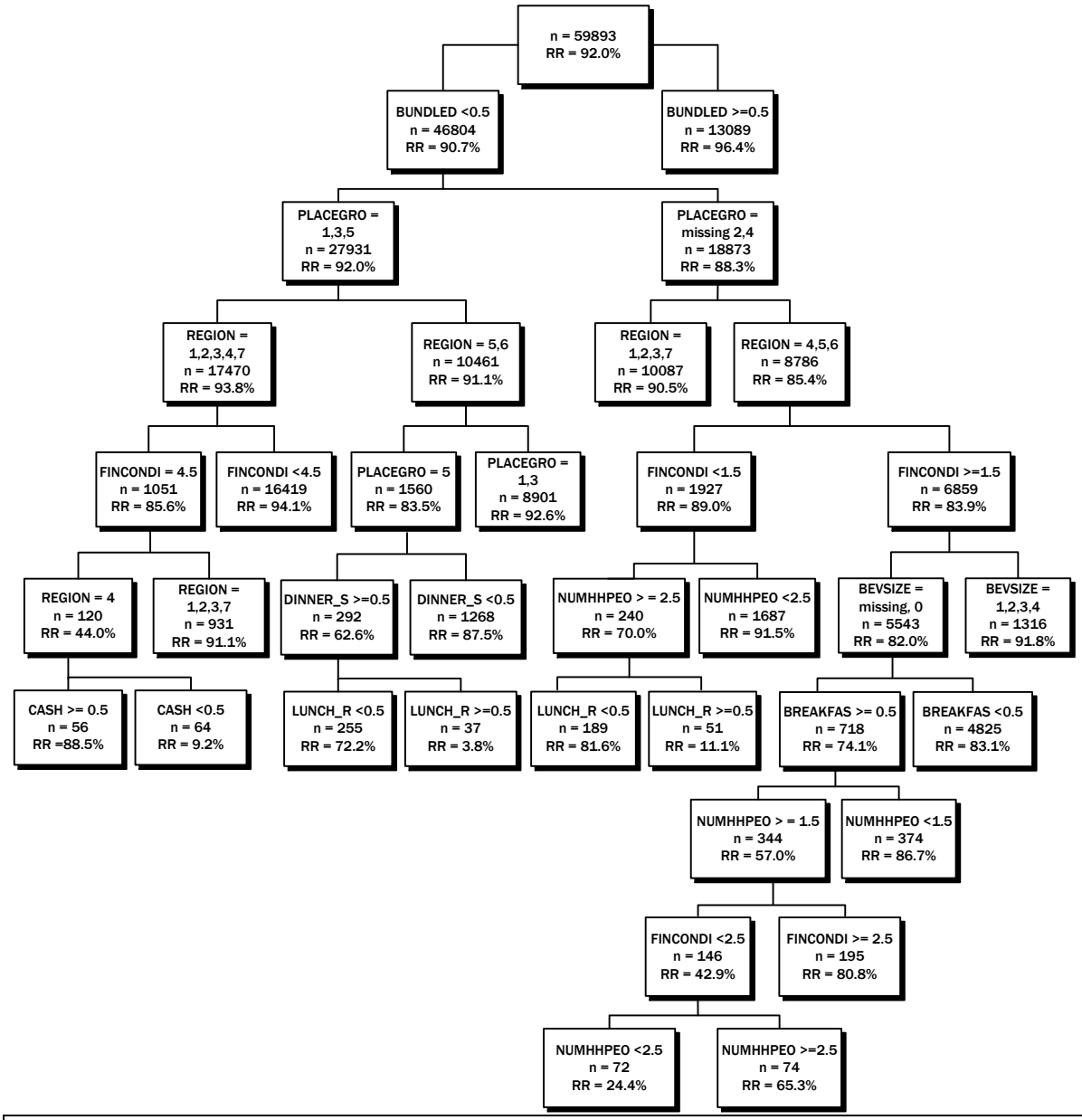
[12]There are 211 purchased (FREE = 0), non-school (MENUID ≠ 3) items that have IMPCOSTMETHOD = .v (valid skip, not imputed) but are missing item price. It is unclear why the missing price is considered a valid skip, but the 211 items are treated as respondents for the purpose of this analysis. There are also 211 items marked as free (FREE = 1) but that have a value other than .v for IMPCOSTMETHOD. This occurred when a non-household member purchased the item and then gave it to the household for free. These items are treated as free in this analysis and excluded.

place type, with items from restaurants that are not a top fast food restaurant (or missing place type) having a higher percentage of missing prices. This analysis supports the survey contractor's choice to use the bundle indicator and place type in the imputation process, given their relationship to price response status.

3.5 Summary

The results of the NRBA showed that the survey contractor made good choices of covariates when imputing income and item prices for the purpose of reducing nonresponse bias. But a few variables that were not used in the imputation process should have been considered. For example, including the following variables in the process could have helped reduce the bias due to item nonresponse: sampling domain and FNS region when imputing household income, including the source of the barcode and whether the payment was made in cash when imputing FAH item prices, and including the number of household members that shared the meal, the type of meal, and region when imputing FAFH item prices. These variables were found to be related to the presence of missing values. If these variables are related to the outcome (income or price) and are not highly correlated with the other covariates, including them in the imputation process could help reduce nonresponse bias.

Figure 3-2. Classification tree for FAFH item price



Variable	Values
BUNDLED	0=not part of a bundle, 1=part of a bundle
PLACEGRO	1=top fast food restaurants, 2=top non-fast food restaurants, 3=food store, 4=other, restaurant, 5=other, non-restaurant
REGION	1=Mid-Atlantic, 2=Midwest, 3=Mountains/Plains, 4=Northeast, 5=Southeast, 6=Southwest, 7=West
FINCONDI	1=very comfortable and secure, 2=able to make ends meet without much difficulty, 3=occasionally have some difficulty making ends meet, 4=tough to make ends meet but keeping head above water, 5=in over your head
DINNER_S	0=no, 1=yes
NUMHHPEO	1=one, 2=two, 3=three or four, 4=five or more
BEVSIZE	0=not a beverage, 1=x-small or small, 2=medium, 3=large, 4=x-large
CASH	0=no, 1=yes
LUNCH_R	0=no, 1=yes
BREAKFAS	0=no, 1=yes

Note: In each cell, "n" is the unweighted sample size and "RR" is the weighted percentage of items with a non-missing price. Variable names are defined in Section 3.1.3 and are abbreviated to 8 characters in this figure.

Evaluation of Imputation Models and Imputation Results 4

4.1 Income

Several evaluations were conducted to assess the imputation of income at the individual and household level. The evaluation results are reported below.

4.1.1 Imputation of Household Income Poverty-Level Groups

Household income poverty-level groups[13] (A: income less than 100% of poverty guideline, B: income greater than or equal to 100% and less than 185% of poverty guideline, or C: income greater than or equal to 185% of poverty guideline) were collected during screening. The income poverty-level group indicates the range of income for a household depending on its household size. For example, if a household has four members and income below $23,000, the household is in group A. This household will be placed in group C if its income is above $43,000. When imputing person-level income, the imputed values were restricted by the upper bound of household income poverty-level group. Moreover, the person-level imputation was done in two batches, with the first batch imputing the missing income values in groups A and B, and the second batch imputing the missing values in group C (without upper bound). The income poverty-level group was subject to a 5.3 percent missing rate. Its missing values were imputed through simple random imputation before being used in the person-level imputation.

Westat believes this imputation strategy was inappropriate because it ignored the correlation between income groups and household size, as well as other household-level characteristics. Income group should have been imputed along with other household-level variables in Step 1. In this evaluation we did not include imputed income groups as a covariate for any analysis since we are skeptical of the quality of the imputed values. The improper imputation of income group may not have an overwhelmingly negative impact on the imputation of individual income components,

[13]The cutoffs to determine the income poverty-level groups can be found in screener questions 10a and 10b.

however, since income group has a relatively low missing rate of 5.3 percent and was used only as the upper boundary in the imputation process.

4.1.2 Zero Income Components

A large proportion of zero incomes was reported for all the income components. During the imputation process, some zero income values were redefined as missing and, therefore, imputed. Meanwhile, some originally missing values were imputed as zeros. Table 4-1 compares the percent of zero values for each income component at the person level before and after imputation. After imputation, the percent of positive values increased for earnings and retirement and disability income by two to three percentage points, and it increased slightly for the other income components. Two reasons explain this. First, for retirement income, zeroes were treated as passive refusal if it was reported that there was such type income in the household screener. Second, household income was imputed and redistributed to each income component for each household member if total income was zero but total expense was positive (Imputation Step 4). The imputed income was mostly allocated to earnings and retirement if a household did not indicate a single income source at the screener.

Table 4-1. Percent of zero values for each income component before and after imputation

Income components	Percent of zero values (%)					
	Impute1	Impute2	Impute3	Impute4	Impute5	Reported
Earnings	48.2	48.4	48.3	48.2	48.4	50.8
Investments	96.7	96.7	96.7	96.7	96.6	96.8
Retirement	76.3	76.3	76.3	76.4	76.5	79.7
Welfare	95.2	95.2	95.2	95.2	95.2	95.4
Unemployment	97.6	97.6	97.5	97.5	97.5	97.7
Other	96.4	96.4	96.4	96.4	96.5	96.9

4.1.3 Distribution of Imputed Income Values

Besides imputation, the survey contractor also conducted data editing and used flags to identify edited and imputed values. Figure D-2 in Appendix D shows six plots of distributions of reported versus edited or imputed values for each income component. The edited values are plotted separately from the imputed values. The magnitude of edited values varies across five imputations, and the edited values are much larger than the imputed values in general. The codebook does not

give a detailed explanation about how the edited values were generated. They may come from the process of redistributing household income to person level when household income was "significantly" less than household expense and, therefore, re-imputed.

In Step 4, the survey contractor identified income values as "unreasonable" if total household expenditure was at least 10 percent higher than income and re-imputed household income with the total expenditure being used as the lower bound in this case. This step impacted 218 (about 4.5%) households. The purpose of re-imputation was to capture possible under-reported income using the assumption that the amount of total expense excluding sporadic one-time expenses should not be significantly more than the monthly income. To be consistent with the definition of unreasonable income values, 90.9 percent [=1/(1+10%)] of total expenditure, instead of total expenditure, should have been used as the lower bound. This allows the imputed income values to be higher than total expenditure by 10 percent or less. Using total expenditure as the lower bound generated larger re-imputed values and, therefore, introduced positive bias to income estimates. On the other hand, the total expense accounts for many items such as rent, mortgage, home insurance, property tax, child care, utility bills, and these reported amounts may be subject to large measurement error. Also, it is possible that some households were using their savings to pay for the monthly bills for a certain period. Instead of using total expenditure to force the imputed income to be strictly higher than expenditure, an alternative option would have been to use total expenditure as a predictor in the imputation model to account for the correlation between income and expenditure.

Figure D-3 in Appendix D shows the imputed values versus reported values for household income. A majority of the imputed values are smaller than $20,000. The outlying imputed household income values may correspond to the outlying edited values from the income components at the person level. Moreover, the imputed household income values have large variation across five imputations at the tail, which also corresponds to the large variation in the edited income components at the person level.

IVEware does two-step imputation to mixed variables such as income. It first uses a logistic regression to impute zero versus non-zero status, then a linear regression is used to impute the non-zero status. This is a reasonable procedure. Doing imputation five times is a common practice. Multiple imputation theory suggests that a small number of imputations may yield excellent results. Other researchers found that a larger number of imputations may be needed to improve the efficiency under specific scenarios. Obviously, doing more imputations would require more computational effort and would produce a larger imputation dataset.

High variation in imputed values exists at the individual level. ERS provided some examples for which the five imputed values for a single record differed dramatically. Under the two-step imputation algorithm, it is possible that some of the five imputed values are zero, while others are fairly large positive values, especially if the logistic and linear regression models do not have very strong predictors. However, this large variation at the individual level does not necessarily mean it will introduce large imputation variance to survey estimates. Table A6a from a technical memorandum prepared by the survey contractor[14] shows that the variance between multiples is very small compared with the variance within multiples for mean household income.

4.1.4 Net Versus Gross Earnings

When imputing earnings, the survey contractor multiplied reported net earnings (before tax) by a factor of 1.4 to obtain gross earnings (after tax) before imputing missing values. This ensures that earnings were defined comparably in the imputation model. Separately, ERS evaluated a regression model for earnings and concluded that different adjustment factors should be applied to different age groups when estimating gross earnings. This approach also yielded an overall factor close to 1.4. Westat proposes that, in the future, net and gross earnings be imputed simultaneously using IVEware, which helps maintain the correlation at the individual level and requires no additional work beyond the imputation process.

4.1.5 Imputation Models

The survey contractor for FoodAPS-1 considered two options for imputing missing income: One was to impute only household-level income through household-level modeling; the other was to impute missing person-level income data and obtain household income by aggregating person-level income across household members. For the first option, one can impute the total household income directly or impute each household income component separately and then aggregate. The second option was chosen by the survey contractor to address the concern that household-level imputation cannot incorporate covariates that strongly correlate with person income contributing to household income. However, imputing individual income components may have introduced too much noise to

[14]This internal technical memorandum, titled "The National Household Food Acquisition And Purchase Survey – Multiple Imputation of Missing Income Data," was prepared by Zhou and Sukasih from Mathematica in 2015.

the aggregated household income. Moreover, if the person-level characteristics are not highly correlated with the income components, aggregating the imputed income components may not work better than imputing the household income directly. To answer this question, we fitted models at both the household and person levels.

The reported income amounts are very skewed. Therefore, in the imputation of income, a log transformation was done. In our evaluation of the household linear regression model, we used log (reported income),[15] excluding the zero income values, as the dependent variable and a large pool of household characteristics (see Table 3-1) plus log (total expenditure) as the predictors. The regression model accounted for the design features such as stratification and clustering, as well as sample weights. The adjusted R-square was 0.77. This indicates the existence of a strong correlation between log (income) and the household-level predictors. We ran the same regression for log (post-imputation income) by taking the first set of imputed values from multiple imputation. The adjusted R-square was 0.72. The correlation was well retained after imputation. As discussed above, during the FoodAPS-1 imputation process, some household income values were re-imputed and bounded by total household expense. This process helps reinforce the correlation between income and expenditure. Expenditure is a highly significant predictor in the regression model.

In our evaluation at the person level, we used log (reported income components),[16] excluding the zero values, as the dependent variable, and a large pool of person and household characteristics (see Table 3-1) as the predictors. Again, the sample design features were incorporated into the model. The adjusted R-square values in Table 4-2 were, in general, lower than that from the household-level model, especially for investment income. This indicates that the associations between some income components and predictors at the person level are low. In the linear regressions we included five additional household-level covariates that were not used for imputing household income (see Table 3-1, Section B). The adjusted R-square would be even lower without the use of these five variables (e.g., the adjusted R-square is only 0.39 in the model for earnings without imputed values). The imputed values from the imputation models may contain much noise. Doing the imputation at the household level directly may be a better choice, which is of higher quality and requires less effort, if publishing person-level income components is not a concern.

[15]We subset to the reported household income values with imputation flag being zero.

[16]We subset to the reported individual income values with imputation flag being zero and two. Edited values were treated as reported in the regression models.

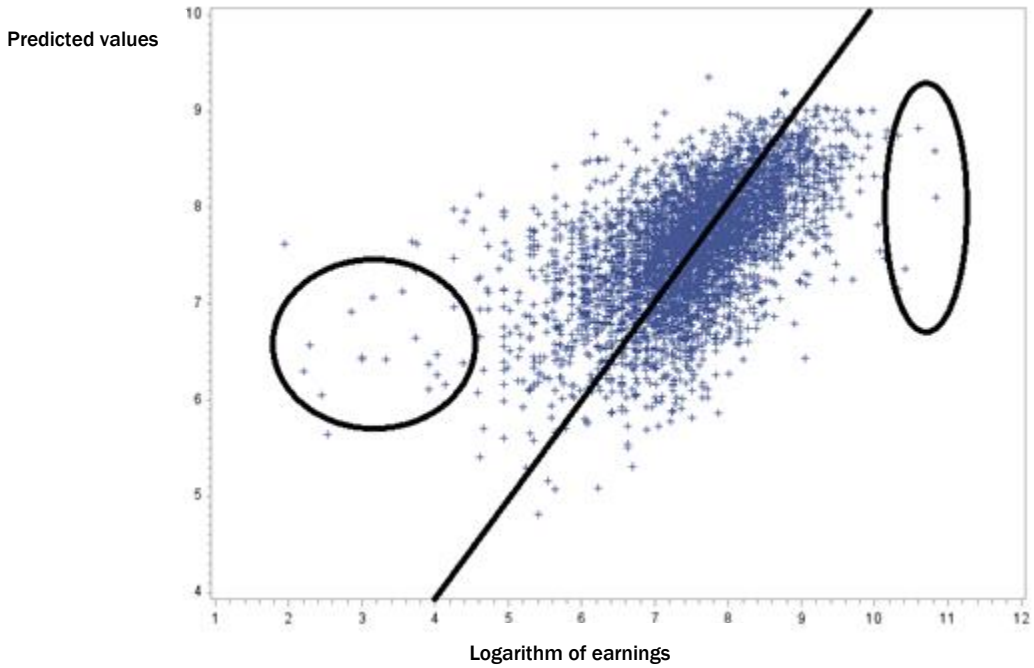
Table 4-2. Adjusted R-squares for linear regression models with and without imputed values

Income components	Adjusted R-square with reported income values (with imputed values)
Earnings	0.43 (0.42)
Investments	0.28 (0.28)
Retirement	0.39 (0.24)
Welfare	0.41 (0.41)
Unemployment	0.54 (0.59)
Other	0.38 (0.36)

The adjusted R-square dropped a lot after imputed values were used in the model for retirement. This may be because it has a large proportion of imputed values (almost 15%). Meanwhile, Figure D-2 shows that the imputed values of retirement in the first imputation are generally smaller compared with the other four imputations and compared with the reported values. The correlation may be somewhat distorted for this reason.

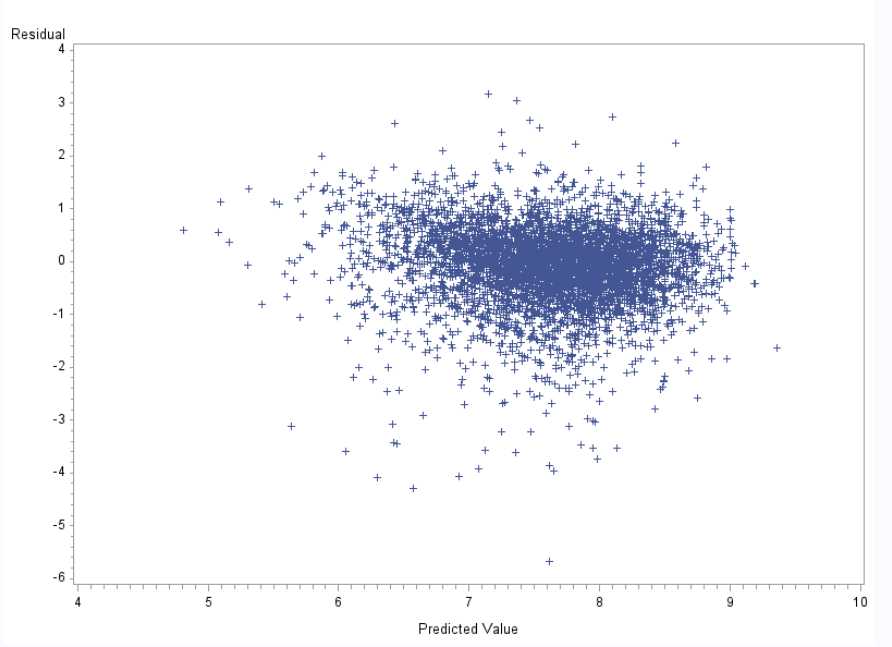
Figure 4-1 shows the predicted values versus the Y (log of earnings) values for the earnings model, with a 45-degree reference line. The predicted values tend to be smaller than the observed Y values for large outlying Ys and larger than the observed Ys for very small Ys. The imputed values from the imputation models, as opposed to edited values, are, therefore, less likely to be outlying.

Figure 4-1. Predicted values versus Y (logarithm of earnings)



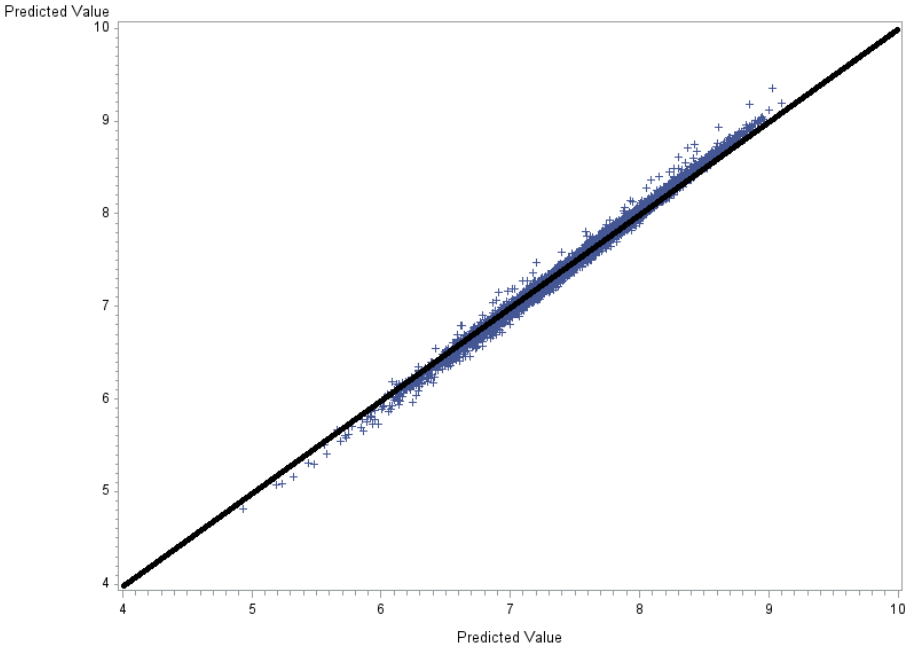
The plot of residual versus predicted values in Figure 4-2 looks reasonable except that there are some large negative residuals.

Figure 4-2. Residual versus predicted values



There are some outlying earnings values (e.g., >20,000 monthly) in the data. Figure 4-3 shows the impact of outliers on the predicted values. The vertical axis is the predicted values from the model including the outliers (defined as >15,000 or <100), whereas the horizontal axis is the predicted values from the model excluding the outliers. The predicted values changed slightly. ERS also found that the removal of outlying values does not change the regression estimation much in the individual gross earning model.

Figure 4-3. Impact of outliers on predicted values



4.1.6 Mean Income Before and After Imputation

Table 4-3 gives the mean household income by subgroups before and after the imputations done by the survey contractor. This analysis indicates if imputation works to reduce the nonresponse bias. For example, if a variable should be included in the imputation model (related to both response propensity and the outcome), then after imputation there is some expectation to observe a change in mean income using this variable as subgroup. The subgroups are defined by the categorical variables at the household level in Table 3-1. The right-most column "relative difference" shows the difference between the means after and before imputation relative to the standard error of the mean before imputation. We use a relative difference greater than 2 as a guideline to identify important differences because it indicates that the means after imputation are two standard errors away from the means before imputation. Relative differences greater than 2 are highlighted in red.

Positive relative differences in Table 4.3 indicate that the mean income increases after imputation. The mean income before imputation was computed using RHHINCOME, the reported household income that simply sums up the reported income components within a household. However, some reported zero values were recognized as passive refusals or inconsistent with household expenses in Step 2 of the imputation process. Such zero values were recoded to missing and later imputed. Also, in Step 4, household income was re-imputed if it was significantly lower than household expenses.

Both imputation steps contributed to the increase in mean household income. The low-income households have larger increase in the mean income, for example, than households that were receiving SNAP at the time of survey, households that did not have $2,000 or more liquid assets, households that did not own a car or truck, households that were in poverty, etc. The variables TARGETGROUP, HHSIZER, and FNS_REGION have some categories with large relative differences in mean incomes. For example, the relative difference for non-SNAP households with income <100 percent of the federal poverty guideline is as large as 32.27.

Table 4-3. Mean household income by subgroups before and after imputation

Subgroup variable	Category	Sample size	Mean before imputation	S.E.	Mean after imputation	S.E.	Relative difference
Earnings as a source of household income	No	1,297	2,139	157	2,431	176	1.86
	Yes	3,525	5,797	277	6,182	297	1.39
Unemployment compensation as a source of household income	No	4,512	4,972	254	5,333	274	1.42
	Yes	310	3,884	325	4,210	325	1.00
Retirement/disability as a source of household income	No	2,924	5,495	324	5,892	347	1.22
	Yes	1,898	4,014	208	4,316	233	1.45
Investment as a source of household income	No	4,449	4,609	216	4,959	237	1.62
	Yes	373	7,010	625	7,439	688	0.69
Welfare as a source of household income	No	4,223	5,017	259	5,387	279	1.43
	Yes	599	3,634	264	3,861	299	0.86
Other source of household income	No	4,317	4,998	255	5,358	276	1.41
	Yes	505	4,152	360	4,509	377	0.99
Urban area type	Urbanized area	2,677	5,092	348	5,496	366	1.16
	Urban cluster	231	3,973	367	4,179	363	0.56
	Neither	1,918	5,176	378	5,499	402	0.85
Metro- or micro-area	Metro	3,705	5,403	349	5,805	361	1.15
	Micro	439	4,088	257	4,333	262	0.95
	Neither	682	4,396	387	4,651	397	0.66
Household has any financial problems	No	3,453	5,494	312	5,897	336	1.29
	Yes	1,373	2,986	193	3,117	200	0.68
Migrant or seasonal farm worker	No	4,802	5,084	272	5,444	291	1.32
	Yes	24	4,071	701	4,130	731	#
Self-employment status	No	4,259	4,888	286	5,226	303	1.18
	Yes	567	6,387	467	6,883	457	1.06
Past 30 days–Food pantry or bank	No	4,504	5,202	277	5,568	296	1.32
	Yes	322	1,733	188	1,894	184	0.86
Receiving SNAP at time of survey	No	3,245	5,550	294	5,920	312	1.26
	Yes	1,581	2,105	117	2,397	174	2.49
Currently receive SNAP?	No	3453	5,499	285	5,866	304	1.29
	Yes	1,373	1,904	111	2,192	185	2.61

Table 4-3. Mean household income by subgroups before and after imputation (continued)

Subgroup variable	Category	Sample size	Mean before imputation	S.E.	Mean after imputation	S.E.	Relative difference
Household financial condition	Very comfortable and secure	658	7,346	596	7,861	613	0.86
	Able to make ends meet without much difficulty	1,363	5,529	508	5,905	522	0.74
	Occasionally have some difficulty making ends meet	1,425	4,065	230	4,362	271	1.29
	Tough to make ends meet but keeping your head above water	1,090	2,710	196	2,915	205	1.05
	In over your head	290	1,801	248	1,939	240	0.56
$2,000 or more liquid assets	No	3,180	3,091	109	3,441	152	3.20
	Yes	1,646	6,761	419	7,127	439	0.87
$3,000 or more liquid assets	No	3,453	3,564	335	3,898	343	1.00
	Yes	1,373	6,762	400	7,148	413	0.96
Food Security Score	0	2,522	6,003	343	6,410	364	1.19
	1	541	4,046	260	4,420	306	1.44
	2	419	3,213	245	3,406	252	0.79
	3	473	2,812	249	3,086	277	1.10
	4	166	2,506	286	2,661	313	0.54
	5	146	1,983	257	2,238	226	0.99
	6	190	2,545	310	2,712	355	0.54
	7	148	1,849	206	1,910	220	0.29
	8	109	1,576	216	1,710	250	0.62
	9	50	1,490	159	1,557	148	0.42
	10	62	1,235	165	1,366	233	0.79

Table 4-3. Mean household income by subgroups before and after imputation (continued)

Subgroup variable	Category	Sample size	Mean before imputation	S.E.	Mean after imputation	S.E.	Relative difference
Own/lease car or truck	No	768	1,704	161	2,041	223	2.09
	Yes, own	3,913	5,496	289	5,843	308	1.20
	Yes, Lease	79	4,533	614	5,743	973	1.97
	Own and lease	66	5,631	969	6,206	962	0.59
Rent or own home	Rent	2,370	3,923	515	4,212	524	0.56
	Own	2,300	5,832	334	6,236	355	1.21
	Other, do not pay for housing	156	3,239	554	3,450	558	0.38
Costs too much to eat healthy foods	Disagree	2,785	5,399	318	5,778	335	1.19
	Agree	2,041	4,418	530	4,732	522	0.59
Too busy to prepare healthy foods	Disagree	3,848	5,177	322	5,542	344	1.14
	Agree	978	4,733	217	5,066	230	1.53
Respondent thinks healthy foods don't taste good	Disagree	4,200	5,253	293	5,608	312	1.21
	Agree	626	3,642	227	4,024	269	1.68
Reviews bills	Never	497	4,206	376	4,396	374	0.51
	Rarely	322	4,331	337	4,649	332	0.94
	Sometimes	777	5,630	1,120	6,008	1,117	0.34
	Usually	783	5,666	482	6,025	472	0.75
	Always	2,415	4,983	271	5,369	295	1.43
	Not applicable	32	1,845	609	2,002	635	0.26
Pays bills on time	Never	81	3,033	691	3,052	686	0.03
	Rarely	136	2,404	369	2,491	366	0.24
	Sometimes	621	2,635	219	2,906	261	1.23
	Usually	1,209	4,924	354	5,163	353	0.68
	Always	2,763	5,486	321	5,902	343	1.30
	Not applicable	16	2,607	1,332	2,746	1,267	#
Pays more than minimum on credit card	Never	408	3,772	318	4,202	324	1.35
	Rarely	183	4,275	623	4,450	627	0.28
	Sometimes	481	4,324	255	4,760	305	1.71
	Usually	428	6,068	473	6,302	465	0.49
	Always	1,355	6,560	440	6,995	466	0.99
	Not applicable	1,971	3,285	287	3,564	306	0.97

Table 4-3. Mean household income by subgroups before and after imputation (continued)

Subgroup variable	Category	Sample size	Mean before imputation	S.E.	Mean after imputation	S.E.	Relative difference
Household member changed job in past 3 months	No	4,320	5,091	286	5,431	301	1.19
	Yes	506	4,968	402	5,533	408	1.41
Illness/disability in past 3 months	No	4,455	5,180	285	5,539	306	1.26
	Yes	371	3,692	346	4,036	400	0.99
Initial Interview language	English	4,440	5,182	285	5,530	303	1.22
	Spanish	376	3,147	360	3,647	349	1.39
	Korean	10	1,148	0	4,196	2250	#
Final Interview language	English	4,414	5,190	287	5,532	304	1.19
	Spanish	383	3,234	335	3,754	336	1.55
	Korean	29	1,767	459	4,276	1,553	#
Sampling target group	NonSNAP household, with income <100% of the Federal Poverty Guideline	434	835	35	1,967	371	32.27
	NonSNAP household, with income >=100% and <185% of the Federal Poverty Guideline	878	1,887	49	2,170	94	5.80
	NonSNAP household, with income >=185% of the Federal Poverty Guideline	1,933	6,743	348	7,045	361	0.87
	SNAP household	1,581	2,105	117	2,397	174	2.49
Census region	Midwest	1,170	4,622	213	4,894	224	1.28
	Northeast	816	5,726	848	6,272	838	0.64
	South	1,784	4,901	462	5,209	488	0.67
	West	1,056	5,733	801	6,191	870	0.57

Table 4-3. Mean household income by subgroups before and after imputation (continued)

Subgroup variable	Category	Sample size	Mean before imputation	S.E.	Mean after imputation	S.E.	Relative difference
Food and Nutrition Service Region	Mid-Atlantic	471	5,732	1,399	6,027	1,442	0.21
	Midwest	787	4,507	257	4,749	267	0.94
	Mountains/Plains	383	5,083	108	5,481	90	3.68
	Northeast	415	5,027	115	5,717	258	5.98
	Southeast	1,133	5,375	602	5,704	632	0.55
	Southwest	662	4,165	587	4,481	606	0.54
	West	975	5,839	814	6,306	885	0.57
Household size	1 person	1,024	3,242	519	3,557	528	0.61
	2 persons	1,337	5,506	199	5,801	209	1.48
	3 persons	874	7,023	672	7,375	670	0.52
	4 persons	770	6,188	388	6,552	441	0.94
	5 persons	444	5,625	441	6,577	705	2.16
	6 persons	202	5,480	574	5,990	642	0.89
	7 or more persons	175	5,280	513	5,725	548	0.87
Is anyone in household receiving benefits from WIC?	No	4,365	5,154	280	5,516	299	1.29
	Yes	461	3,386	207	3,661	225	1.33

Note: The relative difference is the difference between the means divided by the standard error of the mean before imputation. The relative difference is suppressed and shown as "#" if the sample size is smaller than 30. We use a relative difference greater than 2 as a guideline to identify important differences because it indicates that the means after imputation are two standard errors away from the means before imputation. Relative differences greater than 2 are highlighted in red.

4.1.7 Correlations Before and After Imputation

If imputation were done appropriately, the correlations between variables should be well preserved. The correlations between household income and other variables are shown in Table 4-4. The correlations are similar before and after imputation except that for TARGETGROUP and TOTALEXP_R the correlations changed slightly. The correlation between household income and target group dropped from 0.42 to 0.39 possibly because this variable was not used in the imputation process. The correlation between household income and total expenditure increased from 0.39 and 0.43. Step 4 of the imputation process strengthens this correlation by re-imputing the under-reported income using total expenditure as the lower bound. Although not a major concern, this issue could have been resolved by using TARGETGROUP and TOTALEXP_R appropriately in the imputation process, as suggested in Section 3.2.3 and Section 4.1.3.

Table 4-4. Correlation between household income and other variables before and after imputation

Variables	Correlation before imputation	Correlation after imputation
Earnings as a source of household income	0.33	0.33
Unemployment compensation as a source of household income	0.05	0.05
Retirement/disability as a source of household income	0.15	0.16
Investment as a source of household income	0.17	0.17
Welfare as a source of household income	0.08	0.08
Other source of household income	0.05	0.05
Urban area type	0.04	0.05
Metro- or micro-area	0.09	0.10
Household has any financial problems	0.16	0.18
Migrant or seasonal farm worker	0.01	0.01
Self-employment status	0.09	0.10
Past 30 days–Food pantry or bank	0.11	0.12
Receiving SNAP at time of survey	0.21	0.21
Currently receive SNAP?	0.20	0.20
Household financial condition	0.29	0.30
$2,000 or more liquid assets	0.32	0.32
$3,000 or more liquid assets	0.28	0.28
Food Security Score	0.25	0.26
Own/lease car or truck	0.21	0.20
Rent or own home	0.17	0.18
Costs too much to eat healthy foods	0.08	0.08
Too busy to prepare healthy foods	0.03	0.03
Respondent thinks healthy foods don't taste good	0.09	0.08
Reviews bills	0.09	0.10
Pays bills on time	0.15	0.16
Pays more than minimum on credit card	0.26	0.26

Table 4-4. Correlation between household income and other variables before and after imputation (continued)

Variables	Correlation before imputation	Correlation after imputation
Household member changed job in past 3 months	0.01	0.01
Illness/disability in past 3 months	0.07	0.06
Initial Interview language	0.08	0.07
Final Interview language	0.08	0.07
Sampling target group	0.42	0.39
Census region	0.08	0.10
Food and Nutrition Service Region	0.10	0.11
Household size	0.25	0.25
Is anyone in household receiving benefits from WIC?	0.06	0.06
Total expenditure	0.39	0.43

4.2 FAH Prices

To evaluate the hot deck for FAH item prices, we compared analysis results with and without imputation. The analysis results included item price means and correlations with the item, event, and household characteristics listed in Section 3.1.2. As in the item NRBA, the analysis was limited to purchased items and excludes items with price imputed using the multiple-UPC deterministic method. The multiple-UPC method will be discussed in section 4.2.2 below on "Imputation based on within-sample data."

Table E-2 in Appendix E shows the weighted mean FAH item price by subgroup before and after imputation. The mean price changed by less than two standard errors for all subgroups in the analysis. Overall, the mean item price was $2.92 for reported items, and $2.91 after hot deck imputation. The item NRBA showed higher rates of missing prices for certain subgroups of items, so we would expect the mean price to change after imputation if the imputation corrected for these potential sources of bias. However, some subgroups with lower response rates had higher mean prices (such as place types other than superstores and supermarkets) while others had lower mean prices (such as paying with cash), and less than 3 percent of items had a price imputed, so it seems reasonable that there was little change in the mean price.

Correlations before and after imputation are shown in Table 4-5. The results provide no evidence of attenuation of the correlations as a result of the imputation. They also indicate that IRI department, place type, and barcode source have the strongest correlations with FAH item price. For each, the

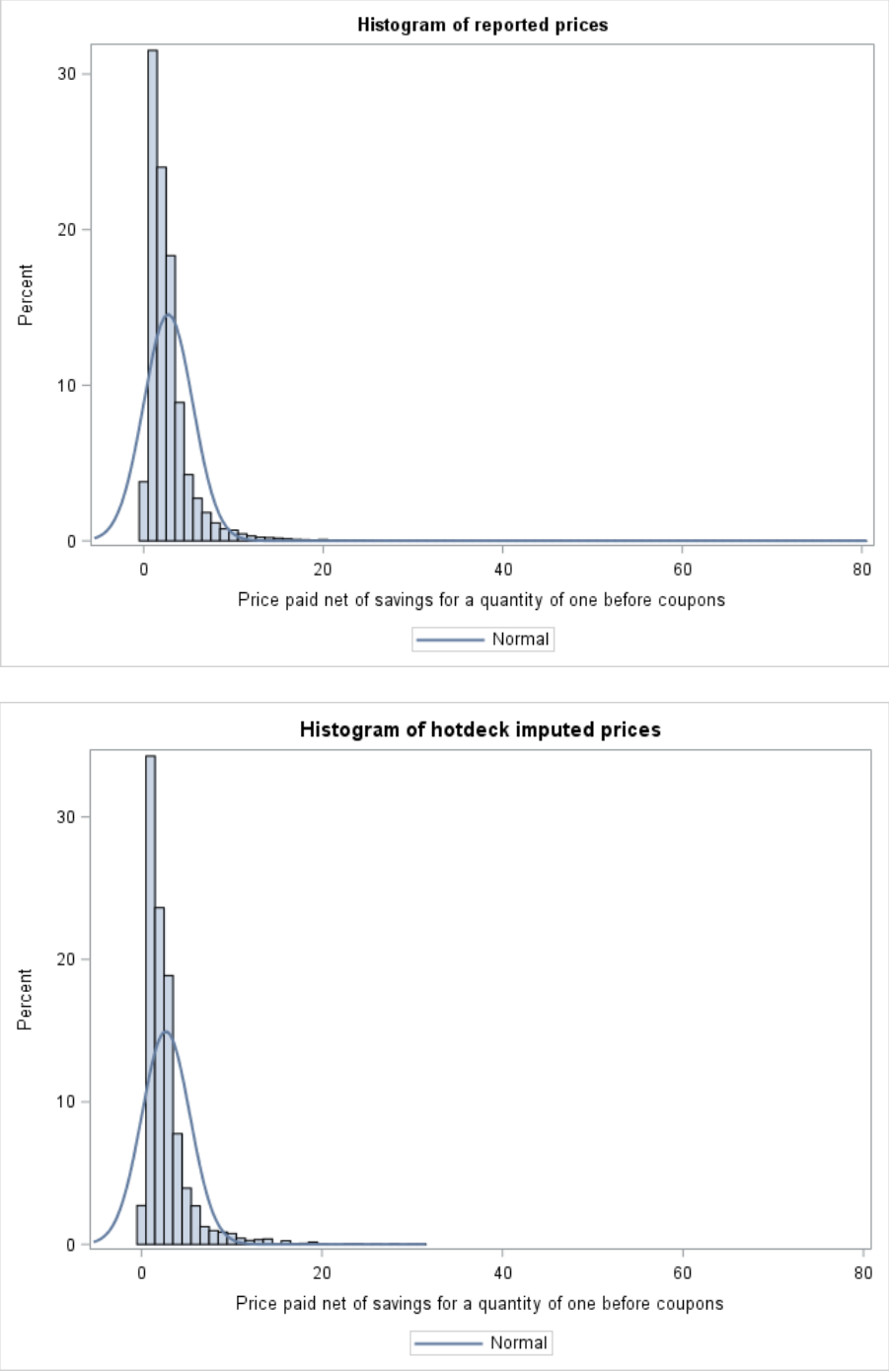
correlation with FAH item price is over 0.2. IRI department and place type were used in imputation, and barcode type was not used.

Table 4-5. Correlations with FAH item price before and after hot deck imputation

Auxiliary variable	Correlation with FAH item price	
	Before imputation	After imputation
Place type	0.203	0.192
Package size unit	0.048	0.051
IRI department	0.317	0.311
Quantity	0.047	0.041
Barcode source	0.214	0.210
Number of items	0.079	0.077
Loyalty card	0.033	0.033
Paid with cash	0.059	0.058
Paid with SNAP EBT	0.007	0.007
Household size	0.001	0.001
Target group	0.070	0.069
WIC household	0.028	0.027
Financial condition	0.062	0.062
FNS region	0.070	0.069
Low access	0.009	0.011
Low income–low access	0.044	0.045

As an additional review of the imputation, histograms of reported prices and hot deck-imputed prices were produced and are shown in Figure 4-4. These show similar distributions of reported and imputed prices.

Figure 4-4. Histograms of FAH item price



The next step in the evaluation of the FAH imputation was to run a main effects ANOVA model with item price as the dependent variable and the variables used in imputation (stratification and sorting variables) as predictors. Item price was defined as the price paid net of savings for a quantity of one before coupons, and was equal to TOTITEMEXPNOCOUPONS/QUANTITY. The

predictors were place type (derived as described above), PSU, IRI department (IRI_DEPT), IRI aisle (IRI_AISLE), package size in ounces, package size in pounds, package size as a count, and weight. The package size in ounces was set to the package size (PKGSIZE) expressed in ounces if the package size unit was ounces, or it could be converted to ounces and was set to zero otherwise. The other package size and weight variables were similarly defined. Cases with a package size unit other than oz., lbs., count, liter, dry oz., grams, or piece were excluded from the model since they were excluded from the imputations.

The R-square of the ANOVA model is 0.2913 (adjusted R-square of 0.2906), indicating that the predictor variables explain 29 percent of the variation in item price. The model was then refit with barcode source and number of items as additional predictors. Barcode source had one of the highest correlations with FAH item price and is also significantly related to whether price was missing. The number of items associated with the event was highly related to whether price was missing. Including these variables in the model, however, led to only a small increase in the adjusted R-square to 0.2955. These two variables might be related to others that are already included in the imputation.

4.2.1 Remaining Missing Values After Imputation

For FAH, 3.2 percent of items were still missing a price after the imputation process. The survey contractor did not impute a price if the FAH item was missing package size or weight, missing the IRI department and aisle code, or the package size/weight was not specified in ounces, pounds, or count (or could not be converted to these units). We recommend using sequential imputation in the future to impute for missing values of the predictor variables, and then using the resulting complete data to impute for the price. In addition, it is our understanding that for the purpose of nutrition information, weight needs to be converted to a common unit for all items. If it is feasible given the schedule, this conversion could be done prior to imputation so that weight can be used in the imputation process even if it was not initially provided in one of the three major units.

4.2.2 Imputation Based on Within-Sample Data

ERS expressed concern that only sample data were used to impute for missing prices. For items with UPCs appearing multiple times in the database, using the sample data should give a good indication of price given the granularity of the UPC. Some examples of prices imputed using this method were reviewed as part of the evaluation. In one, the price of two 1.25 oz. bags of the same type of chips

was imputed as $2.00 based on the price of an item with the same UPC bought at the same location (PLACEID). In a second, the price was missing for a gallon of 2 percent milk from a supermarket. There was no other item in the database with the same UPC and PLACEID, so the price was imputed using the price of items with the same barcode that were purchased at a the same supermarket chain in the same PSU. In a third, there were no items with the same UPC in the same store and PSU, so the price was imputed using the mean price of items with the same UPC at the same store (PLACENAME). The imputed values in all three examples seem reasonable. In the last example, one of the donor records had a quantity of 24. The price may have been lower given that the item was purchased in bulk. Therefore, we would recommend also considering quantity in the imputation, using donors that have the same quantity and place, if there are any, and ignoring quantity otherwise.

For the hot deck imputation of purchased FAH items, we performed a spot check of a random 1 percent of imputed values. Some of the imputed prices are unreasonable for the given item. For example, one price was imputed as $11.99 for a 16-oz. box of cereal, and another has an imputed price of $0.69 for a 64-oz. bottle of juice. Rather than imputing based on sample data, it may be preferable to impute the price using the national value from an external database (such as the IRI) when available for the item, even though it will not reflect the local variation in the price.

4.3 FAFH Prices

The analysis described in Section 4.2 was repeated for FAFH item prices. As in the FAFH item NRBA, the analysis was limited to purchased, non-school items. In the following discussion, by "imputation" we mean both the mean cell imputation and ratio adjustment (where applied), as described in section 1. There were no data available to distinguish between the two steps.

The first step in evaluating the imputation was to compare the mean item price by subgroup before and after the imputation, as shown in Table F-2 in Appendix F. Overall, the mean price increased from $2.81 to $2.89. The largest change (from $0.48 to $3.33) was observed for items with a missing value of quantity. However, less than 1 percent of items had a missing value of quantity, and two-thirds of these also had a missing price prior to imputation. In addition, the imputation brings the mean price for items with a missing value of quantity closer to the mean price for items without a missing value of quantity. Large differences were also observed for items that were part of a bundle, where the mean price increased from $1.95 to $2.14, and items purchased by SNAP households,

where the mean price increased from $2.34 to $2.47. An increase in mean item price after imputation seems reasonable, since a higher percentage of missing prices was observed for items that tended to cost more, such as items that were not obtained from a top national chain, not part of a bundle, not a beverage, or were purchased in the Northeast. The imputation seems to correct for this bias. The increase in mean price might also be related to the treatment of zero prices, which is discussed further below.

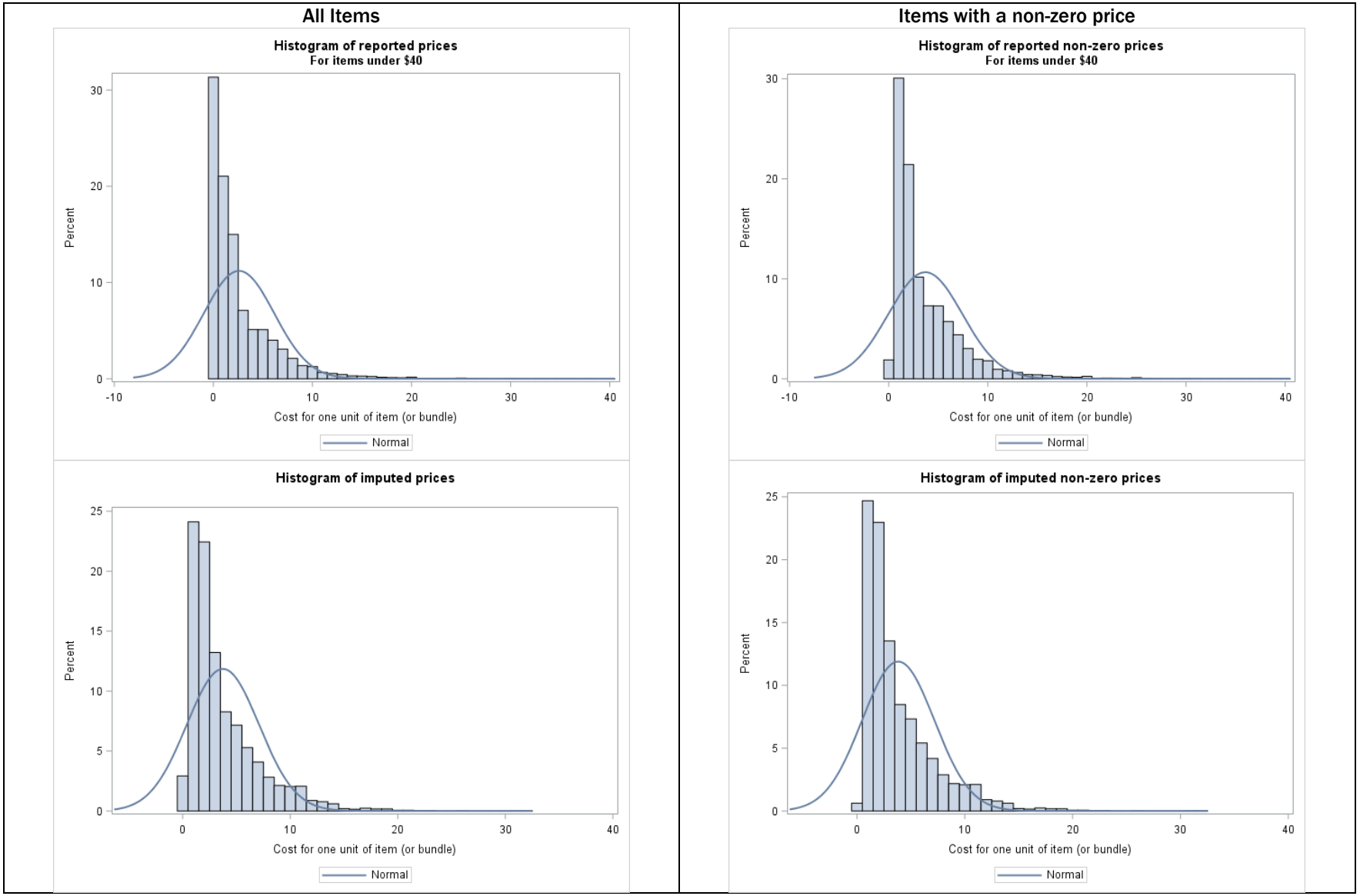
The correlation of item price with each of the auxiliary variables is given in Table 4-6. Results are provided before and after imputation. The analysis provides no evidence of attenuation in the correlations as a result of the imputation process. The predictor variables menu group, place type, and relative beverage size, along with the variable indicating whether the item was purchased for dinner/supper, have the highest correlations with item price.

Table 4-6. Correlations with FAFH item price before and after imputation, for non-school items

Auxiliary variable	Correlation with FAFH item price	
	Before imputation	After imputation
Menu group	0.161	0.168
Place type	0.125	0.136
Bundle indicator	0.088	0.076
Relative beverage size	0.147	0.154
Quantity	0.054	0.054
Book type	0.031	0.032
Number of HH members that shared meal	0.053	0.061
Breakfast	0.066	0.070
Lunch	0.029	0.027
Dinner	0.114	0.117
Snack or drink	0.053	0.060
Paid with cash	0.088	0.092
Target group	0.040	0.038
WIC household	0.022	0.020
Financial condition	0.046	0.049
FNS region	0.055	0.053

Histograms of reported and imputed FAFH item prices are shown in Figure 4-5. The histograms for reported prices are limited to items with a price under $40 to exclude the impact of outliers. Given the large number of items in FAFH that were purchased but have a price of zero, two sets of histograms were produced: the ones on the left include all items, and the ones on the right are limited to items with a non-zero price. As in Figure 4-4, the histograms on the top are for reported prices, and those on the bottom are for imputed prices. The main difference in the distributions of reported and imputed prices is the prevalence of zero prices. Approximately 30 percent of reported prices are zero, compared with 2 percent of imputed prices. Our understanding is that a zero price was only imputed if all prices within the cell were zero or if the adjustment factor in the ratio adjustment was very small (less than 0.5) and more than 70 percent of prices in the cell were zero. A more thorough investigation of the data would be needed to understand if it makes sense to only impute a zero price under these two conditions, but we were not able to look into this further under the scope of this task. The types of items with a higher prevalence of missing prices, such as non-bundled items, might also be less likely to have zero prices. The process could possibly be improved by first imputing for whether the price is zero or not and then imputing for the non-zero prices.

Figure 4-5. Histograms of FAFH item prices, for purchased non-school items



The final step in the evaluation of FAFH price imputation was to run an ANOVA model with FAFH price as the dependent variable, and the imputation cells as the independent variable, excluding one observation that was an extreme outlier. The adjusted R-square from this model is 0.0903. A major limitation of this analysis is that we could not exactly re-create the imputation cells since we were missing the WWEIA food category variable, which could be a significant predictor of price, and we did not have relative size for chicken. The model also does not reflect the effect of the ratio adjustment. We tried refitting the model with an indicator for dinner/supper as an additional independent variable, since this was significantly related to price being missing, and had a higher correlation with item price. This increased the adjusted R-square to 0.1005. This indicates that it could be beneficial to include this additional variable in the creation of the imputation cells. However, if the item was for dinner/supper, the total paid at the event is likely also higher, so this might have been partially accounted for in the ratio adjustment. The item NRBA and evaluation did not suggest any significant item-level variables that were not already being accounted for in the imputation.

4.3.1 Remaining Missing Values After Imputation

Less than 1 percent of FAFH items had a missing price after the imputation process. Prices were not imputed for these items because the residual (difference between the total paid and the sum of the non-missing items prices) was too small. It is unclear to us why the prices for these items were not imputed as zero. There may be a logical reason, but we could not find this explained in the documentation.

4.3.2 Imputation Based on Within-Sample Data

The use of within-sample data may not be as large of a concern for FAFH items as FAH items, since the imputed price is not solely based on prices of other items in the database but also on the total amount paid for that particular event. The price of other items is used in the mean imputation step and the total paid for the event in the ratio adjustment. As a spot check, we reviewed imputed prices for a random 1 percent of FAFH purchased, non-school items. As with FAH, some prices may not be reasonable for a particular item, such as $1.62 for a small ice cream product.

5 Conclusions

The survey contractor's imputation of household income was done reasonably well. IVEware, the software that was used to perform single or multiple imputations of missing values, adopts the sequential regression imputation method and incorporates a large number of important predictors in the imputation model. Multiple imputation also allows the estimation of imputation variance. It also was appropriate to treat income as a mixed variable and impute it in two steps: first using a logistic regression to impute zero versus non-zero status and then using a normal linear regression to impute non-zero values.

However, the original imputation strategy could be improved from the following aspects:

- Account for household-level characteristics when imputing income poverty-level groups. In other words, income group may be imputed in step 1a along with all the other household variables.
- Consider imputing before-tax and after-tax earnings simultaneously in the imputation process.
- Consider using total expenditure as a predictor in imputation models rather than using it as a lower bound to identify and re-impute unreasonable income values.
- Consider the inclusion of more household characteristics in the imputation model (e.g., TARGETGROUP and HHSIZER).
- Consider imputing income at the household level directly rather than aggregating from the person level.
- Consider other semi-parametric approaches (e.g., Judkins et al., 2007) for imputing income. Income has a skewed distribution. Imputation models may easily be mis-specified without doing appropriate diagnostic checks.

For FAH item prices, the predictor variables were related to both price and the missingness of price, suggesting that the imputation process should have reduced bias in price estimates. We did not find any issues with the current imputation if it will be used in analyses that involve aggregate prices over different types of items, such as estimates of the total FAH expenditures for a household. The distribution of item prices was similar for reported and imputed values, and correlations with our analysis variables were preserved. However, if an analyst is interested in a particular type of item,

such as cereal, then we found that the imputed values might not always be reasonable. Another drawback of the current imputation approach is that 3.2 percent of items still had a missing price after the imputation process, since no imputation was done for items that had missing values for one of the predictor variables. We would recommend the following for improving the FAH imputation process:

- As a first step, for items with a UPC appearing multiple times in the dataset, use the deterministic approach taken in FoodAPS-1, but using donors with the same value of QUANTITY when possible.
- Otherwise, if it is possible to match the item to an external database, such as IRI, impute using the price from the external database rather than just relying on within-sample data.
- For the remaining items, a hot deck approach can be used similar to that in FoodAPS-1, but first imputing for any missing values of the predictor variables and then using the results to impute for price.

For FAFH item prices, the analysis supported the choice of predictor variables, as they were again found to be related to price and whether price was missing. Our analysis of FAFH prices was limited since we did not have access to all the variables used in imputation. We also did not have information to distinguish between the results of the mean cell imputation and the ratio adjustment. We did not find any serious issues with the imputation, although we are unclear whether the low percentage of zero prices among the imputed values is a concern. For FAFH imputation, we recommend the following:

- Further review the discrepancy in the percentage of zero prices among reported and imputed items.
- Consider treating FAFH item price as a mixed variable and imputing in two steps: first, imputing for zero versus non-zero status, and then imputing for non-zero prices.

Appendix A

Imputation of Income

Appendix A
Imputation of Income

In the FoodAPS-1 questionnaire, income information was collected for persons age 16 and older at the screener stage and interview stage. At the screener stage, a respondent who is any adult resident in the household answered two questions related to income: (1) whether there was each of a number of income types in the household; and (2) the income group of the household (the actual income amount was not collected). At the Final Interview, the primary respondent (the primary food shopper or meal planner) answered questions about whether each household member had any of the six types of income and the amount of each type of monthly income. To aid in the collection of the income data, an Income Worksheet was left with the primary respondent at the beginning of the week. The worksheet asked for the same information to be collected during the Final Interview, so the primary respondent had a week to gather information about income sources and amounts from all household members age 16 or more.

There are six types of income at the individual level. They are: (1) Earnings; (2) Unemployment compensation; (3) Retirement and disability; (4) Welfare, child support, alimony; (5) Investment; and (6) Other sources. An indicator summarized whether the primary respondent reported completely to income questions about the persons:

- No income and zeroes on all income amounts – complete;
- Have income and positive on some and zeroes on other income amounts – complete;
- Missing income indicator and missing all income amounts;
- Have income but DK or Refusal on all income amounts;
- Have income but reported all zeroes – passive refusal;
- Have income but reported zeroes on some and DK or Refusal on other income amounts; and
- Have income but reported zeroes and positive values for some and DK or Refusal on other income amounts.

The survey contractor used a multiple imputation technique to impute the missing income data and capture the imputation error variance component of the total variance. It was implemented through the sequential regression multivariate imputation (SRMI) method. The SRMI was processed using

SAS-callable IVEware v0.1. IVEware does model selection through a stepwise searching process based on a specified minimum marginal R^2.

The imputation of income took four steps. First, missing values were imputed in the household-level covariates and poverty-level income boundaries. The imputed values were used later at the person-level income imputation. The household-level covariates were imputed through a single imputation using IVEware. The household income poverty-level groups (categorized as A/B/C) were imputed randomly based on the distribution of the observed values.

Next, the six types of income were imputed at the person-level, and the household total income was calculated by aggregating all six income variables for a household member and then aggregating income across all members of the household. The person-level missing incomes were classified into 4 types:

- Missing values, DKs, and refusals;
- Passive refusal: where the indicator of income source was reported as "yes" but the amount of a type of income was reported as 0 for the person across all six types of income;
- Passive refusal: where the household screener indicator of retirement income source at household level was reported as "yes" but the aggregated amount of retirement income across members in the interview was reported as zero; and
- Inconsistency between expense and income: Total household income was reported as zero but total household expenses were non-zero under normal financial conditions.

When imputing incomes at the person-level, all person-, household-, and area-level covariates were included in the modeling. Missing values in the person-level covariates were also imputed in this process. Total household income based on the income group during screening was used as an upper bound to impute each type of income at the person level. Imputation was done in two batches. The first batch imputed missing income for cases in income groups A and B, and the second batch imputed cases in group C, where no upper bound was applied.

Finally, the household income was cross-checked against household expenses. If income was smaller than expenses with no financial difficulties being reported, household income was re-imputed through household-level imputation and then distributed into six types of income and across all household members.

Appendix B

Imputation of Missing Prices for Food at Home (FAH) Items

Appendix B
Imputation of Missing Prices for Food at Home (FAH) Items[17]

FAH items are defined as "foods and drinks that are brought home and used to prepare meals that are consumed at home or elsewhere." The unit price for a FAH item is defined as "the price paid net of savings for a quantity of one before coupons." This value could be missing for a particular item for three reasons: (1) no receipt was provided, (2) the receipt was unreadable, or (3) the item was free. This section summarizes the process the survey contractor used to impute for the missing values. Prices are imputed for both purchased and free items, although the imputed values are recorded in separate fields (IMPUTEDEXP for purchased and IMPUTEDVALUE for free).

Prior to imputation, items with Universal Product Code (UPCs) appearing multiple times in the dataset were reviewed for outliers. This was done to identify and correct data entry errors. Next, items were identified that were missing package size or weight, missing the IRI Department and Aisle code, or the package size/weight was not specified in ounces, pounds, or count.[18] No imputation was done for such items.

For the remaining items, the following process was used:

- For variable weight items (such as meat or produce) with UPCs (219 items), extracted the price from the barcode.
- For other items with a UPC appearing multiple times in the dataset (996 items), used a deterministic method, taking the mean value for sample items with the same UPC
  - In the same location (PLACEID) if available;
  - Otherwise, in the same place (PLACENAME, e.g., Wal-mart or McDonald's) in the PSU if available; and
  - Otherwise, in the same place (PLACENAME) over the whole sample.

[17]Based on an internal technical memorandum titled "The National Household Food Acquisition And Purchase Survey – Food-At-Home Items Documentation," which was prepared by Cole and Baxter from Mathematica in 2014.

[18]Per the technical memorandum by Cole and Baxter (2014), "Units specified as liter, dry oz., grams, or piece were converted to the three primary units; other units were excluded from imputations."

- Otherwise (5,061 items), applied a stratified hotdeck method using a SAS macro provided by Ellis (2007),[19] where:
  - Strata were defined by place type (superstore, supermarket, convenience store or small grocery, all other food stores, eating places, and all other places) and package size unit (OZ, LBS, and COUNT);
  - Within strata, the items were sorted by PSU, IRI Department and Aisle, and package size or weight; and
  - The donor comes from the nearest neighbor.

The hot deck model was evaluated by applying it to the 996 items that were imputed deterministically and comparing the results. This process was used to determine the final number of sorting variables to include in the model. Per the technical memorandum by Cole and Baxter (2014), "The final hot deck specification is the one that yielded lowest variance in the final distribution of item prices. Final hot deck results obtain an exact match for 80.4 percent for the 996 items that could be imputed deterministically."

The imputation was done without taking into account the total price paid for the acquisition, since the total price could include non-food items of unknown price.

Table B-1 summarizes the prevalence of missing item prices before and after imputation. As can be seen from the table, 3.2 percent of FAH items were still missing a price after the imputation process. All are items without a UPC.

Table B-1. Prevalence of missing item prices before and after price imputation

	Total items	Before imputation		After imputation	
		# Missing price	% Missing price	# Missing price	% Missing price
Category of items					
Items with UPC	83,115	4,619	5.6	0	0.0
Items with Food Book barcode	23,225	3,037	13.1	2,727	11.7
Other from Blue Page or receipt	36,721	3,249	8.8	1,902	5.2
Purchased or free					
Purchased	139,608	7,615	5.5	2,846	2.0
Free	3,239	3,078	95.0	1,692	52.2
Not reported	214	212	99.1	91	42.5
Total	143,061	10,905	7.6	4,629	3.2

Source: Internal technical memorandum by Cole and Baxter (2014).

[19]Ellis, Bruce. "A Consolidated Macro for Iterative Hot Deck Imputation." Northeast SAS Users Group, 2007. www.nesug.org/proceedings/nesug07/po/po03.pdf.

Appendix C

Imputation of Missing Prices for Food Away From Home (FAFH) Items

Appendix C
Imputation of Missing Prices for Food Away From Home (FAFH) Items[20]

FAFH items are defined as "foods and drinks that are obtained and consumed away from home, and prepared foods that are brought into the home." Defining item price is more complicated for FAFH items than FAH, as some items can occur as part of a bundle (such as a combo meal at a fast food restaurant or an entrée with sides) and are not individually priced. Such items receive a zero price. Another difference is that imputation was done for free items for FAH but not for FAFH. Finally, the FAFH items have the additional restriction that the item prices should add up to the total for the acquisition (less tip). This is not necessarily true for FAH events where the total price of the acquisition might include non-food as well as food items, for which neither item descriptions nor prices were collected.

Several edits were performed on the FAFH data prior to imputation. After the edits, there were 2,088 acquisitions that had at least one item requiring imputation (out of 37,408 acquisitions). School acquisitions and non-school acquisitions were considered separately for this process.

For the school acquisitions requiring imputation, median paid school meal price was assigned to acquisitions if the total paid was not reported and the acquisition included only items eligible for reimbursable meals. Medians were calculated by meal and age group from acquisitions by children identified from the Initial Interview as receiving full-price meals. The remaining 18 items with missing prices were included in the non-school imputations.

For the 1,971 non-school acquisitions (6.0% of all non-school acquisitions) requiring item price imputation, a two-step process was used. First, the missing item prices were imputed using the median of the non-zero sample prices within the cell, where 563 cells were formed based on menu group (beverage, top 60, generic, school), type of place (top fast food restaurants, top non-fast food restaurants, food store, other restaurant, other non-restaurant), food category (4-digit food group from What We Eat in America [WWEIA]), indicator for whether the item was bundled, and relative size (where appropriate). The technical memorandum by Cole et al. (2015) notes that "deterministic

[20]Based on the internal technical memorandum, titled "The National Household Food Acquisition And Purchase Survey – Food-Away-From-Home Items Documentation," which was prepared by Cole et al. from Mathematica in 2015.

methods were chosen because the adjustment to TotalPaid provides a simple way to account for variations in price due to geographic location, serving size, or quality without modeling those attributes."

Second, the imputed prices were ratio-adjusted so that the reported and imputed prices (plus tip where reported) summed to the total payment for the acquisition, less an assumed 10 percent (for tax). If the adjustment factor exceeded an acceptable upper bound (1.6) and the acquisition included items with zero price, then values were imputed for the zero price items, with the assumption that they should have been non-zero. If the adjustment factor was below an acceptable lower bound (0.6), then the items with the highest percentage of zero prices in the sample had their price set to zero, assuming there were at least 70 percent zero prices for that item in the sample.

After this process, 220 acquisitions (0.7% of non-school acquisitions) were left with some missing price values because the residual (difference between the total paid and the sum of the non-missing items prices) was too small. For example, if a package of meat was missing a price and the residual was $0.15 before imputing its value, the meat price was left as missing because $0.15 is too low to be a valid price for meat.

Appendix D

Tables and Figures for Income Imputation

Appendix D
Tables and Figures for Income Imputation

For each of the six income components captured by FoodAPS-1, the following six tables (D-1 through D-6) provide comparisons of the weighted response rates for income values for different values of selected categorical variables. The Chi-square statistics were not calculated for the subgroups that have at least one category with 0 or 100 percent response rate, labeled by "—".

Table D-1. Bivariate NRBA for individual earnings

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
OVERALL		93.4		
Earnings as a source of household income	No	91.7	2.89	0.0893
	Yes	93.8		
Unemployment compensation as a source of household income	No	93.3	3.20	0.0739
	Yes	95.8		
Retirement/disability as a source of household income	No	92.4	6.30	0.0121
	Yes	95.1		
Investment as a source of household income	No	93.3	0.35	0.5536
	Yes	94.2		
Welfare as a source of household income	No	93.3	2.00	0.1571
	Yes	95.1		
Other source of household income	No	93.4	0.04	0.8347
	Yes	93.2		
Relation	Other	90.1	10.59	0.0011
	Primary respondent or spouse or partner	94.3		
Age group	age ≥ 16 and age ≤ 24	94.0	0.41	0.8881
	age ≥ 25 and age ≤ 44	93.5		
	age ≥ 45 and age ≤ 64	93.1		
	age ≥ 65	93.6		
Sex	Male	94.1	4.30	0.0381
	Female	92.8		
White	No	92.0	2.35	0.1256
	Yes	93.8		
Black	No	93.5	0.01	0.9067
	Yes	93.3		
American Indian/Alaskan Native	No	93.5	0.38	0.5355
	Yes	90.3		

Table D-1. Bivariate NRBA for individual earning (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Asian	No	93.6	3.11	0.0777
	Yes	86.6		
Hawaiian/Pacific Islander	No	93.4	–	–
	Yes	100.0		
Other	No	93.5	0.49	0.4842
	Yes	92.6		
Education level	High school or less	92.6	2.99	0.1488
	High school graduate or equivalent	94.6		
	Some college	94.6		
	College graduate	93.5		
	More than college	88.4		
Hispanic origin	No	93.6	0.81	0.3685
	Yes	92.5		
Marital status	Married	93.5	0.11	0.9295
	Divorced, widowed, separated	93.6		
	Never married	93.2		
Health status	Excellent	93.0	2.45	0.5021
	Very good	93.8		
	Good	92.7		
	Fair	94.8		
	Poor	94.8		
BMI weight category	Not overweight	92.5	3.33	0.1663
	Overweight	94.1		
	Obese	93.9		
Smoke/chew tobacco	No	93.1	3.00	0.0835
	Yes	94.7		
Work status	Working at a job or business	94.2	2.68	0.3378
	With a job or business but not at work	91.4		
	Looking for work	90.3		
	Not working at a job or business	92.8		
Urban area type	Urbanized area	92.8	5.08	0.0779
	Urban cluster	97.4		
	Neither	93.7		

Table D-1. Bivariate NRBA for individual earning (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Metro- or micro-area	Metro	92.7	4.20	0.0827
	Micro	94.2		
	Neither	95.8		
Household has any financial problems	No	93.0	3.86	0.0494
	Yes	95.2		
Migrant or seasonal farm worker	No	93.4	–	–
	Yes	100.0		
Self-employment status	No	93.6	1.07	0.302
	Yes	92.2		
Past 30 days—Food pantry or bank	No	93.3	5.32	0.0211
	Yes	96.1		
Receiving SNAP at time of survey	No	93.9	2.80	0.0943
	Yes	90.9		
Currently receive SNAP?	No	93.7	1.49	0.2225
	Yes	91.6		
Household financial condition	Very comfortable and secure	92.5	3.65	0.2752
	Able to make ends meet without much difficulty	92.7		
	Occasionally have some difficulty making ends meet	94.4		
	Tough to make ends meet but keeping your head above water	94.8		
	In over your head	94.2		
$2,000 or more liquid assets	No	92.8	1.88	0.1706
	Yes	94.0		
$3,000 or more liquid assets	No	93.2	0.52	0.4703
	Yes	93.8		

Table D-1. Bivariate NRBA for individual earning (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Food Security Score	0	92.8	5.19	0.399
	1	95.7		
	2	94.5		
	3	93.5		
	4	94.5		
	5	91.4		
	6	96.2		
	7	95.1		
	8	97.4		
	9	90.8		
	10	96.5		
Own/lease car or truck	No	90.7	4.58	0.1456
	Yes, own	93.7		
	Yes, lease	89.8		
	Own and lease	93.6		
Rent or own home	Rent	93.4	0.08	0.9625
	Own	93.5		
	Other, do not pay for housing	92.7		
Costs too much to eat healthy foods	Disagree	93.4	0.03	0.8738
	Agree	93.6		
Too busy to prepare healthy foods	Disagree	93.5	0.07	0.7921
	Agree	93.1		
Respondent thinks healthy foods don't taste good	Disagree	93.4	0.06	0.8132
	Agree	93.7		
Reviews bills	Never	94.6	4.64	0.3642
	Rarely	94.2		
	Sometimes	94.6		
	Usually	94.1		
	Always	92.6		
	Not applicable	86.8		

Table D-1. Bivariate NRBA for individual earning (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Pays bills on time	Never	98.0	6.01	0.139
	Rarely	94.7		
	Sometimes	92.9		
	Usually	95.4		
	Always	92.7		
	Not applicable	87.8		
Pays more than minimum on credit card	Never	92.3	5.00	0.1846
	Rarely	95.6		
	Sometimes	95.3		
	Usually	94.9		
	Always	92.3		
	Not applicable	93.8		
Household member changed job in past 3 months	No	93.6	0.53	0.4666
	Yes	92.1		
Illness/disability in past 3 months	No	93.5	0.26	0.6122
	Yes	92.4		
Initial Interview language	English	93.6	1.92	0.2429
	Spanish	90.9		
	Korean	84.5		
Final Interview language	English	93.6	1.75	0.356
	Spanish	91.7		
	Korean	85.2		
Sampling target group	NonSNAP household, with income <100% of the Federal Poverty Guideline	84.3	24.85	<.0001
	NonSNAP household, with income >=100% and <185% of the Federal Poverty Guideline	92.3		
	NonSNAP household, with income >=185% of the Federal Poverty Guideline	95.1		
	SNAP household	90.9		
Census region	MidWest	94.4	1.68	0.344
	Northeast	90.5		
	South	94.1		
	West	92.9		

Table D-1. Bivariate NRBA for individual earning (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Food and Nutrition Service Region	Mid-Atlantic	95.5	2.53	0.2515
	Midwest	94.5		
	Mountains/Plains	94.2		
	Northeast	86.6		
	Southeast	93.9		
	Southwest	93.9		
	West	93.1		
Household size	1 person	94.3	5.80	0.3144
	2 persons	94.2		
	3 persons	93.5		
	4 persons	92.7		
	5 persons	90.0		
	6 persons	92.9		
	7 or more persons	93.2		
Is anyone in household receiving benefits from WIC?	No	93.4	0.15	0.7013
	Yes	93.9		

Table D-2. Bivariate NRBA for individual unemployment compensation

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
OVERALL		94.5		
Earnings as a source of household income	No	91.5	7.97	0.0047
	Yes	95.1		
Unemployment compensation as a source of household income	No	94.3	3.69	0.0547
	Yes	96.8		
Retirement/disability as a source of household income	No	93.4	11.34	0.0008
	Yes	96.4		
Investment as a source of household income	No	94.3	0.90	0.3426
	Yes	95.7		
Welfare as a source of household income	No	94.4	1.09	0.2971
	Yes	95.6		
Other source of household income	No	94.5	0.00	0.9538
	Yes	94.5		
Relation	Other	93.3	1.85	0.1733
	Primary respondent or spouse or partner	94.8		
Age group	age ≥ 16 and age ≤ 24	95.1	0.83	0.8056
	age ≥ 25 and age ≤ 44	94.6		
	age ≥ 45 and age ≤ 64	94.1		
	age ≥ 65	94.6		
Sex	Male	95.4	11.21	0.0008
	Female	93.6		
White	No	93.7	0.99	0.3186
	Yes	94.7		
Black	No	94.5	0.00	0.9448
	Yes	94.4		
American Indian/Alaskan Native	No	94.5	1.55	0.2132
	Yes	96.0		
Asian	No	94.6	1.97	0.1604
	Yes	89.3		
Hawaiian/Pacific Islander	No	94.5	3.30	0.0692
	Yes	98.6		
Other	No	94.5	0.04	0.8423
	Yes	94.2		

Table D-2. Bivariate NRBA for individual unemployment compensation (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Education level	High school or less	93.6	2.52	0.1849
	High school graduate or equivalent	95.9		
	Some college	95.1		
	College graduate	94.6		
	More than college	90.4		
Hispanic origin	No	94.8	1.88	0.1707
	Yes	93.1		
Marital status	Married	94.6	0.04	0.97
	Divorced, widowed, separated	94.5		
	Never married	94.3		
Health status	Excellent	93.6	1.88	0.5949
	Very good	94.5		
	Good	94.3		
	Fair	95.7		
	Poor	96.1		
BMI weight category	Not overweight	93.7	2.83	0.2205
	Overweight	95.0		
	Obese	94.9		
Smoke/chew tobacco	No	94.1	8.29	0.004
	Yes	96.1		
Work status	Working at a job or business	96.1	9.44	0.0139
	With a job or business but not at work	91.9		
	Looking for work	89.5		
	Not working at a job or business	92.8		
Urban area type	Urbanized area	93.7	5.43	0.0582
	Urban cluster	97.7		
	Neither	95.1		
Metro- or micro-area	Metro	93.9	2.79	0.17
	Micro	95.2		
	Neither	96.4		

Table D-2. Bivariate NRBA for individual unemployment compensation (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Household has any financial problems	No	93.9	12.94	0.0003
	Yes	97.3		
Migrant or seasonal farm worker	No	94.5	–	–
	Yes	100.0		
Self-employment status	No	94.6	0.77	0.3792
	Yes	93.5		
Past 30 days—Food pantry or bank	No	94.5	0.17	0.6833
	Yes	95.0		
Receiving SNAP at time of survey	No	94.9	3.24	0.072
	Yes	92.2		
Currently receive SNAP?	No	94.7	2.00	0.1577
	Yes	92.8		
Household financial condition	Very comfortable and secure	93.1	4.27	0.2067
	Able to make ends meet without much difficulty	94.1		
	Occasionally have some difficulty making ends meet	95.3		
	Tough to make ends meet but keeping your head above water	96.0		
	In over your head	95.7		
$2,000 or more liquid assets	No	94.1	0.86	0.3525
	Yes	94.9		
$3,000 or more liquid assets	No	94.3	0.19	0.6618
	Yes	94.7		

Table D-2. Bivariate NRBA for individual unemployment compensation (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Food Security Score	0	93.9	4.55	0.3631
	1	96.2		
	2	96.3		
	3	94.6		
	4	96.1		
	5	95.3		
	6	95.9		
	7	95.4		
	8	98.2		
	9	90.8		
	10	97.2		
Own/lease car or truck	No	91.3	7.47	0.0304
	Yes, own	94.9		
	Yes, lease	87.3		
	Own and lease	95.1		
Rent or own home	Rent	94.1	0.58	0.7423
	Own	94.7		
	Other, do not pay for housing	93.7		
Costs too much to eat healthy foods	Disagree	94.4	0.12	0.7274
	Agree	94.8		
Too busy to prepare healthy foods	Disagree	94.6	0.05	0.8301
	Agree	94.2		
Respondent thinks healthy foods don't taste good	Disagree	94.4	1.39	0.2389
	Agree	95.5		
Reviews bills	Never	95.7	6.47	0.1498
	Rarely	97.0		
	Sometimes	95.2		
	Usually	95.2		
	Always	93.6		
	Not applicable	91.4		

Table D-2. Bivariate NRBA for individual unemployment compensation (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Pays bills on time	Never	98.0	7.18	0.0661
	Rarely	96.9		
	Sometimes	94.5		
	Usually	96.5		
	Always	93.7		
	Not applicable	87.8		
Pays more than minimum on credit card	Never	93.6	4.40	0.273
	Rarely	96.5		
	Sometimes	96.0		
	Usually	95.6		
	Always	93.6		
	Not applicable	94.7		
Household member changed job in past 3 months	No	94.6	0.11	0.739
	Yes	93.9		
Illness/disability in past 3 months	No	94.6	0.34	0.5574
	Yes	93.3		
Initial Interview language	English	94.8	4.89	0.0511
	Spanish	90.4		
	Korean	84.5		
Final Interview language	English	94.7	3.84	0.14
	Spanish	91.3		
	Korean	81.2		
Sampling target group	NonSNAP household, with income <100% of the Federal Poverty Guideline	86.8	24.11	<.0001
	NonSNAP household, with income >=100% and <185% of the Federal Poverty Guideline	94.2		
	NonSNAP household, with income >=185% of the Federal Poverty Guideline	95.8		
	SNAP household	92.2		

Table D-2. Bivariate NRBA for individual unemployment compensation (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Census region	MidWest	95.4	1.89	0.3184
	Northeast	92.1		
	South	95.3		
	West	93.6		
Food and Nutrition Service Region	Mid-Atlantic	96.2	2.64	0.2674
	Midwest	95.6		
	Mountains/Plains	94.5		
	Northeast	89.0		
	Southeast	95.3		
	Southwest	94.5		
	West	93.7		
Household size	1 person	94.3	1.91	0.8424
	2 persons	94.8		
	3 persons	94.9		
	4 persons	94.4		
	5 persons	92.4		
	6 persons	95.0		
	7 or more persons	95.0		
Is anyone in household receiving benefits from WIC?	No	94.5	0.00	0.9704
	Yes	94.4		

Table D-3. Bivariate NRBA for individual retirement/disability income

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
OVERALL		91.4		
Earnings as a source of household income	No	89.7	1.94	0.1641
	Yes	91.8		
Unemployment compensation as a source of household income	No	91.4	0.04	0.8445
	Yes	91.8		
Retirement/disability as a source of household income	No	93.1	13.96	0.0002
	Yes	88.5		
Investment as a source of household income	No	91.4	0.01	0.9146
	Yes	91.6		
Welfare as a source of household income	No	91.7	2.19	0.1394
	Yes	88.5		
Other source of household income	No	91.8	4.63	0.0313
	Yes	86.9		
Relation	Other	89.4	4.96	0.0259
	Primary respondent or spouse or partner	91.9		
Age group	age ≥ 16 and age ≤ 24	92.1	2.82	0.3366
	age ≥ 25 and age ≤ 44	92.3		
	age ≥ 45 and age ≤ 64	90.5		
	age ≥ 65	91.2		
Sex	Male	92.4	7.84	0.0051
	Female	90.5		
White	No	89.8	2.54	0.1107
	Yes	91.8		
Black	No	91.7	1.56	0.2114
	Yes	89.5		
American Indian/Alaskan Native	No	91.4	0.10	0.7536
	Yes	90.6		
Asian	No	91.5	0.95	0.33
	Yes	88.0		
Hawaiian/Pacific Islander	No	91.4	3.40	0.0653
	Yes	97.5		
Other	No	91.4	0.01	0.9284
	Yes	91.2		

Table D-3. Bivariate NRBA for individual retirement/disability income (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Education level	High school or less	89.2	3.59	0.1427
	High school graduate or equivalent	92.7		
	Some college	92.2		
	College graduate	92.3		
	More than college	87.5		
Hispanic origin	No	91.6	2.07	0.15
	Yes	90.2		
Marital status	Married	92.0	1.26	0.5109
	Divorced, widowed, separated	90.9		
	Never married	90.8		
Health status	Excellent	89.8	1.97	0.6559
	Very good	91.7		
	Good	91.4		
	Fair	92.5		
	Poor	91.5		
BMI weight category	Not overweight	91.0	0.96	0.6054
	Overweight	92.0		
	Obese	91.3		
Smoke/chew tobacco	No	91.0	3.06	0.0801
	Yes	92.8		
Work status	Working at a job or business	93.5	12.69	0.0023
	With a job or business but not at work	88.5		
	Looking for work	85.4		
	Not working at a job or business	89.2		
Urban area type	Urbanized area	91.0	0.58	0.7268
	Urban cluster	91.7		
	Neither	91.9		
Metro- or micro-area	Metro	91.2	2.66	0.2579
	Micro	89.8		
	Neither	93.1		
Household has any financial problems	No	91.1	1.87	0.1712
	Yes	93.0		

Table D-3. Bivariate NRBA for individual retirement/disability income (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Migrant or seasonal farm worker	No	91.4	0.06	0.7993
	Yes	93.2		
Self-employment status	No	91.6	0.80	0.3723
	Yes	90.2		
Past 30 days—Food pantry or bank	No	91.4	0.10	0.7544
	Yes	92.0		
Receiving SNAP at time of survey	No	92.3	9.56	0.002
	Yes	86.5		
Currently receive SNAP?	No	92.0	6.06	0.0138
	Yes	87.2		
Household financial condition	Very comfortable and secure	90.5	2.20	0.6136
	Able to make ends meet without much difficulty	91.4		
	Occasionally have some difficulty making ends meet	91.5		
	Tough to make ends meet but keeping your head above water	93.1		
	In over your head	90.3		
$2,000 or more liquid assets	No	90.3	4.95	0.0261
	Yes	92.4		
$3,000 or more liquid assets	No	90.5	3.45	0.0632
	Yes	92.4		
Food Security Score	0	91.1	4.42	0.6996
	1	93.1		
	2	92.9		
	3	90.0		
	4	88.2		
	5	93.3		
	6	92.8		
	7	94.7		
	8	93.2		
	9	87.0		
	10	91.5		

Table D-3. Bivariate NRBA for individual retirement/disability income (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Own/lease car or truck	No	86.0	11.51	0.0072
	Yes, own	92.0		
	Yes, Lease	85.2		
	Own and lease	91.3		
Rent or own home	Rent	91.4	0.63	0.7198
	Own	91.5		
	Other, do not pay for housing	89.0		
Costs too much to eat healthy foods	Disagree	91.4	0.01	0.9213
	Agree	91.5		
Too busy to prepare healthy foods	Disagree	91.3	0.16	0.6921
	Agree	91.9		
Respondent thinks healthy foods don't taste good	Disagree	91.5	0.07	0.7869
	Agree	91.0		
Reviews bills	Never	92.9	3.71	0.538
	Rarely	92.6		
	Sometimes	92.2		
	Usually	92.2		
	Always	90.5		
	Not applicable	89.5		
Pays bills on time	Never	94.0	6.43	0.1499
	Rarely	94.9		
	Sometimes	88.2		
	Usually	93.5		
	Always	91.0		
	Not applicable	86.5		
Pays more than minimum on credit card	Never	88.1	6.61	0.1518
	Rarely	94.0		
	Sometimes	93.9		
	Usually	92.9		
	Always	91.3		
	Not applicable	90.5		

Table D-3. Bivariate NRBA for individual retirement/disability income (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Household member changed job in past 3 months	No	91.4	0.02	0.8896
	Yes	91.1		
Illness/disability in past 3 months	No	91.6	1.23	0.2683
	Yes	89.1		
Initial Interview language	English	91.7	3.02	0.1434
	Spanish	87.5		
	Korean	60.8		
Final Interview language	English	91.7	2.60	0.1705
	Spanish	88.0		
	Korean	72.2		
Sampling target group	NonSNAP household, with income <100% of the Federal Poverty Guideline	83.3	31.04	<.0001
	NonSNAP household, with income >=100% and <185% of the Federal Poverty Guideline	90.1		
	NonSNAP household, with income >=185% of the Federal Poverty Guideline	93.5		
	SNAP household	86.5		
Census region	Midwest	92.0	1.23	0.5013
	Northeast	89.7		
	South	92.3		
	West	90.3		
Food and Nutrition Service Region	Mid-Atlantic	94.2	2.31	0.3559
	Midwest	92.4		
	Mountains/Plains	90.5		
	Northeast	86.9		
	Southeast	92.4		
	Southwest	90.5		
	West	90.5		

Table D-3. Bivariate NRBA for individual retirement/disability income (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Household size	1 person	91.7	3.04	0.7243
	2 persons	91.6		
	3 persons	91.7		
	4 persons	92.0		
	5 persons	88.5		
	6 persons	92.5		
	7 or more persons	89.5		
Is anyone in household receiving benefits from WIC?	No	91.5	0.49	0.4853
	Yes	90.2		

Table D-4. Bivariate NRBA for individual investment income

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
OVERALL		93.6		
Earnings as a source of household income	No	89.3	9.65	0.0019
	Yes	94.6		
Unemployment compensation as a source of household income	No	93.6	0.03	0.8562
	Yes	94.0		
Retirement/disability as a source of household income	No	92.7	5.54	0.0186
	Yes	95.2		
Investment as a source of household income	No	93.7	0.17	0.6801
	Yes	92.9		
Welfare as a source of household income	No	93.4	3.48	0.0622
	Yes	95.7		
Other source of household income	No	93.7	0.56	0.4537
	Yes	92.5		
Relation	Other	93.4	0.07	0.7844
	Primary respondent or spouse or partner	93.7		
Age group	age ≥ 16 and age ≤ 24	95.1	3.24	0.2775
	age ≥ 25 and age ≤ 44	94.3		
	age ≥ 45 and age ≤ 64	93.0		
	age ≥ 65	92.4		
Sex	Male	94.5	7.83	0.0051
	Female	92.8		
White	No	93.3	0.17	0.679
	Yes	93.7		
Black	No	93.5	0.27	0.6047
	Yes	94.3		
American Indian/Alaskan Native	No	93.6	6.46	0.011
	Yes	96.5		
Asian	No	93.8	2.13	0.1445
	Yes	88.4		
Hawaiian/Pacific Islander	No	93.6	3.95	0.0469
	Yes	98.6		
Other	No	93.7	0.10	0.7497
	Yes	93.1		

Table D-4. Bivariate NRBA for individual investment income (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Education level	High school or less	93.2	5.57	0.0365
	High school graduate or equivalent	95.7		
	Some college	94.2		
	College graduate	94.0		
	More than college	86.7		
Hispanic origin	No	93.8	0.66	0.4153
	Yes	92.8		
Marital status	Married	93.6	0.34	0.8119
	Divorced, widowed, separated	93.2		
	Never married	94.1		
Health status	Excellent	91.6	3.82	0.2807
	Very good	93.4		
	Good	93.8		
	Fair	95.3		
	Poor	95.6		
BMI weight category	Not overweight	92.3	6.92	0.0282
	Overweight	94.4		
	Obese	94.4		
Smoke/chew tobacco	No	93.1	8.74	0.0031
	Yes	95.8		
Work status	Working at a job or business	95.6	11.93	0.0048
	With a job or business but not at work	91.5		
	Looking for work	88.9		
	Not working at a job or business	91.4		
Urban area type	Urbanized area	93.1	4.06	0.1212
	Urban cluster	97.7		
	Neither	93.9		
Metro- or micro-area	Metro	93.2	1.55	0.3513
	Micro	93.4		
	Neither	95.5		
Household has any financial problems	No	92.8	18.95	<.0001
	Yes	97.4		

Table D-4. Bivariate NRBA for individual investment income (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Migrant or seasonal farm worker	No	93.6	–	–
	Yes	100.0		
Self-employment status	No	93.8	1.00	0.3181
	Yes	92.5		
Past 30 days—Food pantry or bank	No	93.6	1.03	0.3106
	Yes	95.0		
Receiving SNAP at time of survey	No	93.9	1.73	0.1888
	Yes	91.9		
Currently receive SNAP?	No	93.8	0.48	0.4886
	Yes	92.8		
Household financial condition	Very comfortable and secure	90.3	12.16	0.005
	Able to make ends meet without much difficulty	93.4		
	Occasionally have some difficulty making ends meet	95.4		
	Tough to make ends meet but keeping your head above water	95.9		
	In over your head	95.6		
$2,000 or more liquid assets	No	93.9	0.22	0.636
	Yes	93.4		
$3,000 or more liquid assets	No	94.2	1.44	0.2302
	Yes	93.0		
Food Security Score	0	92.6	10.10	0.041
	1	95.9		
	2	95.7		
	3	94.4		
	4	95.9		
	5	98.4		
	6	95.4		
	7	95.7		
	8	98.3		
	9	90.8		
	10	96.9		

Table D-4. Bivariate NRBA for individual investment income (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Own/lease car or truck	No	90.4	5.70	0.0916
	Yes, own	94.0		
	Yes, lease	89.4		
	Own and lease	92.2		
Rent or own home	Rent	93.9	0.24	0.8848
	Own	93.5		
	Other, do not pay for housing	93.7		
Costs too much to eat healthy foods	Disagree	93.2	0.96	0.3262
	Agree	94.4		
Too busy to prepare healthy foods	Disagree	93.4	0.54	0.4613
	Agree	94.5		
Respondent thinks healthy foods don't taste good	Disagree	93.5	2.52	0.1121
	Agree	95.1		
Reviews bills	Never	95.2	2.65	0.6538
	Rarely	94.6		
	Sometimes	94.4		
	Usually	93.8		
	Always	93.0		
	Not applicable	91.4		
Pays bills on time	Never	97.1	10.46	0.0169
	Rarely	96.8		
	Sometimes	94.3		
	Usually	96.2		
	Always	92.6		
	Not applicable	87.8		
Pays more than minimum on credit card	Never	91.5	8.82	0.0434
	Rarely	96.5		
	Sometimes	96.2		
	Usually	95.0		
	Always	92.2		
	Not applicable	94.4		

Table D-4. Bivariate NRBA for individual investment income (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Household member changed job in past 3 months	No	93.6	0.01	0.9047
	Yes	93.8		
Illness/disability in past 3 months	No	93.7	0.17	0.6845
	Yes	92.8		
Initial Interview language	English	93.8	3.00	0.1328
	Spanish	90.4		
	Korean	84.5		
Final Interview language	English	93.8	2.78	0.2474
	Spanish	91.3		
	Korean	81.2		
Sampling target group	NonSNAP household, with income <100% of the Federal Poverty Guideline	85.7	21.36	<.0001
	NonSNAP household, with income >=100% and <185% of the Federal Poverty Guideline	93.8		
	NonSNAP household, with income >=185% of the Federal Poverty Guideline	94.8		
	SNAP household	91.9		
Census region	Midwest	95.0	4.06	0.0903
	Northeast	88.8		
	South	95.0		
	West	92.9		
Food and Nutrition Service Region	Mid-Atlantic	95.0	4.09	0.1037
	Midwest	95.2		
	Mountains/Plains	94.0		
	Northeast	84.2		
	Southeast	94.9		
	Southwest	94.4		
	West	93.0		

Table D-4. Bivariate NRBA for individual investment income (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Household size	1 person	92.1	4.20	0.4795
	2 persons	94.3		
	3 persons	93.8		
	4 persons	94.2		
	5 persons	91.8		
	6 persons	94.7		
	7 or more persons	94.6		
Is anyone in household receiving benefits from WIC?	No	93.6	0.39	0.5329
	Yes	94.4		

Table D-5. Bivariate NRBA for individual welfare income

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
OVERALL		94.5		
Earnings as a source of household income	No	91.3	8.80	0.003
	Yes	95.3		
Unemployment compensation as a source of household income	No	94.5	0.05	0.8311
	Yes	94.9		
Retirement/disability as a source of household income	No	93.4	13.88	0.0002
	Yes	96.4		
Investment as a source of household income	No	94.4	0.80	0.3704
	Yes	95.7		
Welfare as a source of household income	No	94.4	0.96	0.3262
	Yes	95.6		
Other source of household income	No	94.6	0.19	0.6589
	Yes	94.0		
Relation	Other	93.6	1.22	0.2689
	Primary respondent or spouse or partner	94.8		
Age group	age ≥ 16 and age ≤ 24	95.1	2.21	0.483
	age ≥ 25 and age ≤ 44	94.7		
	age ≥ 45 and age ≤ 64	93.9		
	age ≥ 65	95.2		
Sex	Male	95.7	11.98	0.0005
	Female	93.5		
White	No	93.7	1.24	0.2652
	Yes	94.8		
Black	No	94.5	0.00	0.9788
	Yes	94.6		
American Indian/Alaskan Native	No	94.5	2.12	0.145
	Yes	96.4		
Asian	No	94.7	2.13	0.1442
	Yes	89.1		
Hawaiian/Pacific Islander	No	94.5	3.33	0.0682
	Yes	98.6		
Other	No	94.6	0.35	0.5542
	Yes	93.7		

Table D-5. Bivariate NRBA for individual welfare income (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Education level	High school or less	93.5	3.48	0.1591
	High school graduate or equivalent	95.9		
	Some college	94.9		
	College graduate	94.7		
	More than college	91.6		
Hispanic origin	No	94.8	2.49	0.1144
	Yes	93.1		
Marital status	Married	94.8	0.32	0.8233
	Divorced, widowed, separated	94.5		
	Never married	94.2		
Health status	Excellent	93.2	2.69	0.4218
	Very good	94.5		
	Good	94.6		
	Fair	95.7		
	Poor	96.1		
BMI weight category	Not overweight	93.6	4.85	0.0788
	Overweight	95.4		
	Obese	94.8		
Smoke/chew tobacco	No	94.3	2.69	0.1009
	Yes	95.6		
Work status	Working at a job or business	96.2	9.78	0.0097
	With a job or business but not at work	91.7		
	Looking for work	89.3		
	Not working at a job or business	92.9		
Urban area type	Urbanized area	93.6	8.02	0.0176
	Urban cluster	97.7		
	Neither	95.3		
Metro- or micro-area	Metro	94.1	2.53	0.222
	Micro	95.5		
	Neither	96.0		
Household has any financial problems	No	94.0	14.60	0.0001
	Yes	97.2		

Table D-5. Bivariate NRBA for individual welfare income (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Migrant or seasonal farm worker	No	94.5	–	–
	Yes	100.0		
Self-employment status	No	94.7	0.41	0.5213
	Yes	93.9		
Past 30 days—Food pantry or bank	No	94.5	0.14	0.7089
	Yes	95.0		
Receiving SNAP at time of survey	No	94.9	3.48	0.062
	Yes	92.3		
Currently receive SNAP?	No	94.8	2.04	0.1528
	Yes	92.9		
Household financial condition	Very comfortable and secure	92.7	6.07	0.1054
	Able to make ends meet without much difficulty	94.4		
	Occasionally have some difficulty making ends meet	95.5		
	Tough to make ends meet but keeping your head above water	96.0		
	In over your head	95.6		
$2,000 or more liquid assets	No	94.4	0.16	0.6909
	Yes	94.7		
$3,000 or more liquid assets	No	94.6	0.02	0.8799
	Yes	94.5		
Food Security Score	0	94.0	3.71	0.4767
	1	96.1		
	2	95.9		
	3	94.4		
	4	95.9		
	5	95.3		
	6	95.9		
	7	95.7		
	8	97.9		
	9	90.8		
	10	96.9		

Table D-5. Bivariate NRBA for individual welfare income (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Own/lease car or truck	No	91.2	6.51	0.052
	Yes, own	94.9		
	Yes, lease	89.4		
	Own and lease	95.1		
Rent or own home	Rent	94.1	0.66	0.7
	Own	94.8		
	Other, do not pay for housing	93.7		
Costs too much to eat healthy foods	Disagree	94.5	0.04	0.8467
	Agree	94.7		
Too busy to prepare healthy foods	Disagree	94.4	0.22	0.6376
	Agree	95.0		
Respondent thinks healthy foods don't taste good	Disagree	94.5	0.79	0.3755
	Agree	95.3		
Reviews bills	Never	95.2	2.80	0.6343
	Rarely	95.5		
	Sometimes	95.2		
	Usually	95.2		
	Always	93.9		
	Not applicable	91.4		
Pays bills on time	Never	97.1	7.02	0.0749
	Rarely	96.7		
	Sometimes	94.6		
	Usually	96.6		
	Always	93.8		
	Not applicable	87.8		
Pays more than minimum on credit card	Never	93.3	5.89	0.1906
	Rarely	96.6		
	Sometimes	96.3		
	Usually	95.4		
	Always	93.7		
	Not applicable	94.8		

Table D-5. Bivariate NRBA for individual welfare income (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Household member changed job in past 3 months	No	94.6	0.00	0.991
	Yes	94.5		
Illness/disability in past 3 months	No	94.6	0.29	0.5874
	Yes	93.5		
Initial Interview language	English	94.8	5.23	0.0441
	Spanish	90.5		
	Korean	84.5		
Final Interview language	English	94.8	3.93	0.1314
	Spanish	91.4		
	Korean	81.2		
Sampling target group	NonSNAP household, with income <100% of the Federal Poverty Guideline	86.4	28.58	<.0001
	NonSNAP household, with income >=100% and <185% of the Federal Poverty Guideline	94.4		
	NonSNAP household, with income >=185% of the Federal Poverty Guideline	95.9		
	SNAP household	92.3		
Census region	Midwest	95.1	1.98	0.4041
	Northeast	93.1		
	South	95.3		
	West	93.6		
Food and Nutrition Service Region	Mid-Atlantic	96.0	3.15	0.3829
	Midwest	95.2		
	Mountains/Plains	94.5		
	Northeast	91.0		
	Southeast	95.3		
	Southwest	94.5		
	West	93.7		

Table D-5. Bivariate NRBA for individual welfare income (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Household size	1 person	93.7	3.07	0.62
	2 persons	95.3		
	3 persons	94.9		
	4 persons	94.4		
	5 persons	92.5		
	6 persons	95.0		
	7 or more persons	95.0		
Is anyone in household receiving benefits from WIC?	No	94.5	0.02	0.8887
	Yes	94.7		

Table D-6. Bivariate NRBA for individual other income

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
OVERALL		94.3		
Earnings as a source of household income	No	91.9	4.70	0.0302
	Yes	94.8		
Unemployment compensation as a source of household income	No	94.2	2.96	0.0854
	Yes	96.5		
Retirement/disability as a source of household income	No	93.4	8.09	0.0044
	Yes	95.9		
Investment as a source of household income	No	94.4	0.05	0.8246
	Yes	93.9		
Welfare as a source of household income	No	94.2	1.24	0.2662
	Yes	95.4		
Other source of household income	No	94.1	6.37	0.0116
	Yes	96.4		
Relation	Other	93.2	2.00	0.157
	Primary respondent or spouse or partner	94.6		
Age group	age ≥ 16 and age ≤ 24	95.7	2.76	0.3606
	age ≥ 25 and age ≤ 44	94.4		
	age ≥ 45 and age ≤ 64	93.6		
	age ≥ 65	94.5		
Sex	Male	95.2	7.77	0.0053
	Female	93.5		
White	No	93.2	1.53	0.2165
	Yes	94.6		
Black	No	94.3	0.04	0.8423
	Yes	94.1		
American Indian/Alaskan Native	No	94.3	1.91	0.1669
	Yes	96.0		
Asian	No	94.4	1.64	0.1997
	Yes	89.8		
Hawaiian/Pacific Islander	No	94.3	3.35	0.0671
	Yes	98.6		
Other	No	94.4	0.61	0.4341
	Yes	93.1		

Table D-6. Bivariate NRBA for individual other income (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Education level	High school or less	93.2	4.73	0.0715
	High school graduate or equivalent	95.8		
	Some college	95.2		
	College graduate	94.4		
	More than college	89.7		
Hispanic origin	No	94.6	1.82	0.1768
	Yes	93.0		
Marital status	Married	94.1	0.29	0.8464
	Divorced, widowed, separated	94.3		
	Never married	94.7		
Health status	Excellent	93.0	3.26	0.3872
	Very good	94.1		
	Good	94.4		
	Fair	95.5		
	Poor	96.6		
BMI weight category	Not overweight	93.2	5.38	0.0588
	Overweight	95.3		
	Obese	94.5		
Smoke/chew tobacco	No	93.8	16.48	<.0001
	Yes	96.4		
Work status	Working at a job or business	95.7	5.97	0.0675
	With a job or business but not at work	92.3		
	Looking for work	90.7		
	Not working at a job or business	92.9		
Urban area type	Urbanized area	93.4	10.18	0.0058
	Urban cluster	98.3		
	Neither	95.0		
Metro- or micro-area	Metro	93.7	3.19	0.146
	Micro	95.3		
	Neither	96.1		
Household has any financial problems	No	93.6	19.08	<.0001
	Yes	97.5		

Table D-6. Bivariate NRBA for individual other income (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Migrant or seasonal farm worker	No	94.3	–	–
	Yes	100.0		
Self-employment status	No	94.6	1.70	0.1928
	Yes	92.9		
Past 30 days—Food pantry or bank	No	94.3	0.63	0.4265
	Yes	95.4		
Receiving SNAP at time of survey	No	94.7	2.65	0.1033
	Yes	92.1		
Currently receive SNAP?	No	94.5	1.65	0.1992
	Yes	92.7		
Household financial condition	Very comfortable and secure	91.4	9.92	0.0136
	Able to make ends meet without much difficulty	94.3		
	Occasionally have some difficulty making ends meet	95.6		
	Tough to make ends meet but keeping your head above water	96.1		
	In over your head	95.9		
$2,000 or more liquid assets	No	94.3	0.01	0.9425
	Yes	94.3		
$3,000 or more liquid assets	No	94.5	0.23	0.632
	Yes	94.1		
Food Security Score	0	93.6	6.77	0.1385
	1	96.4		
	2	95.8		
	3	94.0		
	4	95.8		
	5	98.2		
	6	95.9		
	7	95.7		
	8	98.1		
	9	90.8		
	10	96.9		

Table D-6. Bivariate NRBA for individual other income (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Own/lease car or truck	No	92.0	4.99	0.1337
	Yes, own	94.7		
	Yes, lease	89.4		
	Own and lease	90.1		
Rent or own home	Rent	94.3	0.07	0.9611
	Own	94.3		
	Other, do not pay for housing	93.6		
Costs too much to eat healthy foods	Disagree	94.1	0.27	0.6031
	Agree	94.7		
Too busy to prepare healthy foods	Disagree	94.1	0.93	0.3355
	Agree	95.1		
Respondent thinks healthy foods don't taste good	Disagree	94.2	1.83	0.1758
	Agree	95.5		
Reviews bills	Never	95.4	6.22	0.1846
	Rarely	96.1		
	Sometimes	94.8		
	Usually	94.9		
	Always	93.5		
	Not applicable	98.6		
Pays bills on time	Never	97.1	–	–
	Rarely	96.2		
	Sometimes	94.5		
	Usually	97.0		
	Always	93.3		
	Not applicable	100.0		
Pays more than minimum on credit card	Never	94.5	9.20	0.0509
	Rarely	96.3		
	Sometimes	96.4		
	Usually	95.4		
	Always	92.9		
	Not applicable	94.8		

Table D-6. Bivariate NRBA for individual other income (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Household member changed job in past 3 months	No	94.4	0.19	0.6647
	Yes	93.6		
Illness/disability in past 3 months	No	94.4	0.32	0.573
	Yes	93.2		
Initial Interview language	English	94.6	4.31	0.0712
	Spanish	90.7		
	Korean	84.5		
Final Interview language	English	94.5	3.35	0.1839
	Spanish	91.9		
	Korean	84.2		
Sampling target group	NonSNAP household, with income <100% of the Federal Poverty Guideline	87.5	18.35	0.0002
	NonSNAP household, with income >=100% and <185% of the Federal Poverty Guideline	93.8		
	NonSNAP household, with income >=185% of the Federal Poverty Guideline	95.6		
	SNAP household	92.1		
Census region	Midwest	95.5	3.69	0.171
	Northeast	91.6		
	South	95.2		
	West	93.0		
Food and Nutrition Service Region	Mid-Atlantic	95.4	3.83	0.2055
	Midwest	95.8		
	Mountains/Plains	94.6		
	Northeast	88.9		
	Southeast	95.1		
	Southwest	94.8		
	West	93.0		

Table D-6. Bivariate NRBA for individual other income (continued)

Subgroup	Value	Weighted response rate	Chi-square	
			Statistic	p-value
Household size	1 person	94.6	4.67	0.4201
	2 persons	95.0		
	3 persons	94.0		
	4 persons	94.4		
	5 persons	91.2		
	6 persons	93.8		
	7 or more persons	94.8		
Is anyone in household receiving benefits from WIC?	No	94.3	0.19	0.6664
	Yes	94.8		

Figure D-1. Distribution of reported and imputed values for each income component

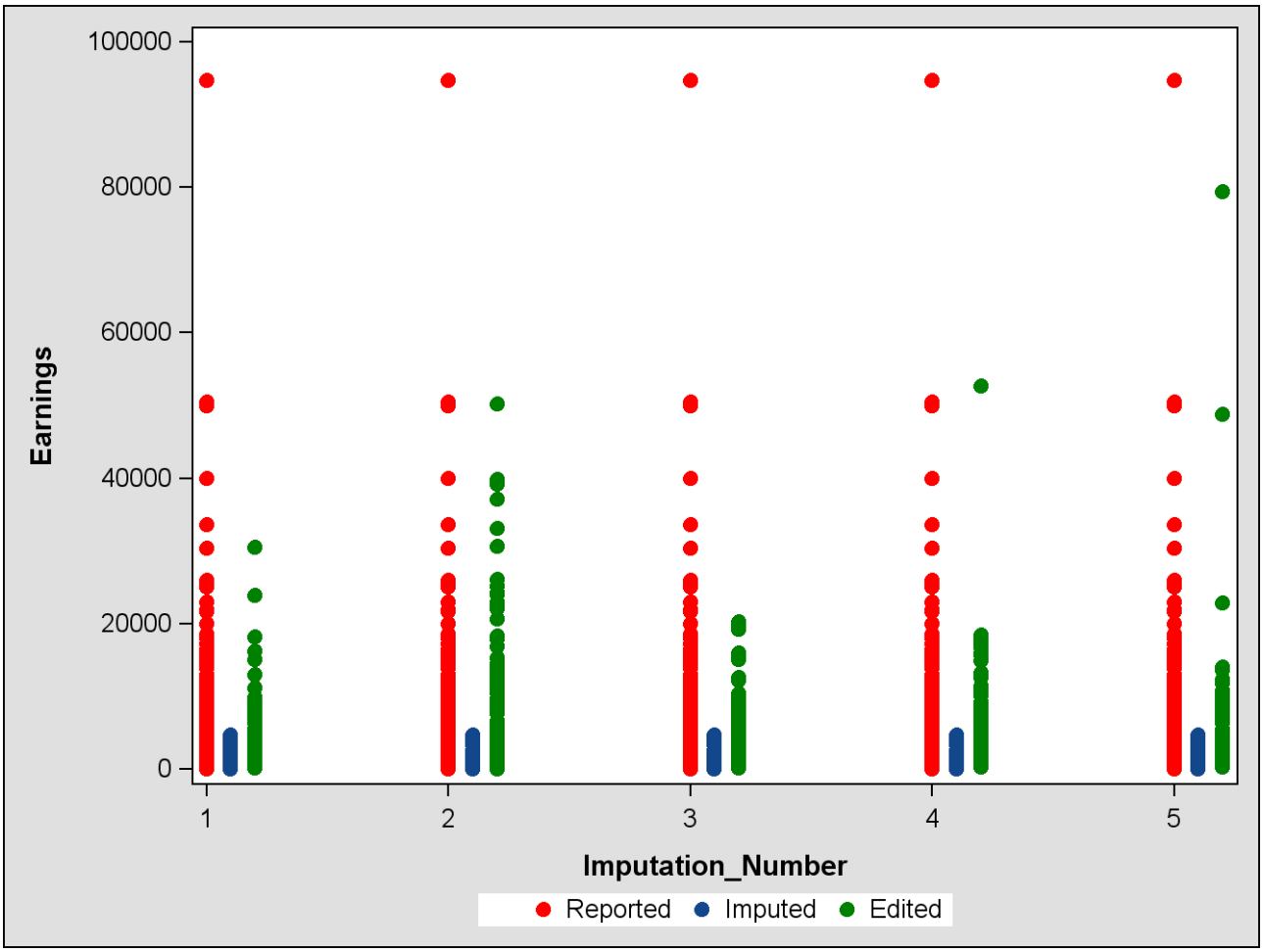


Figure D-1. Distribution of reported and imputed values for each income component (continued)

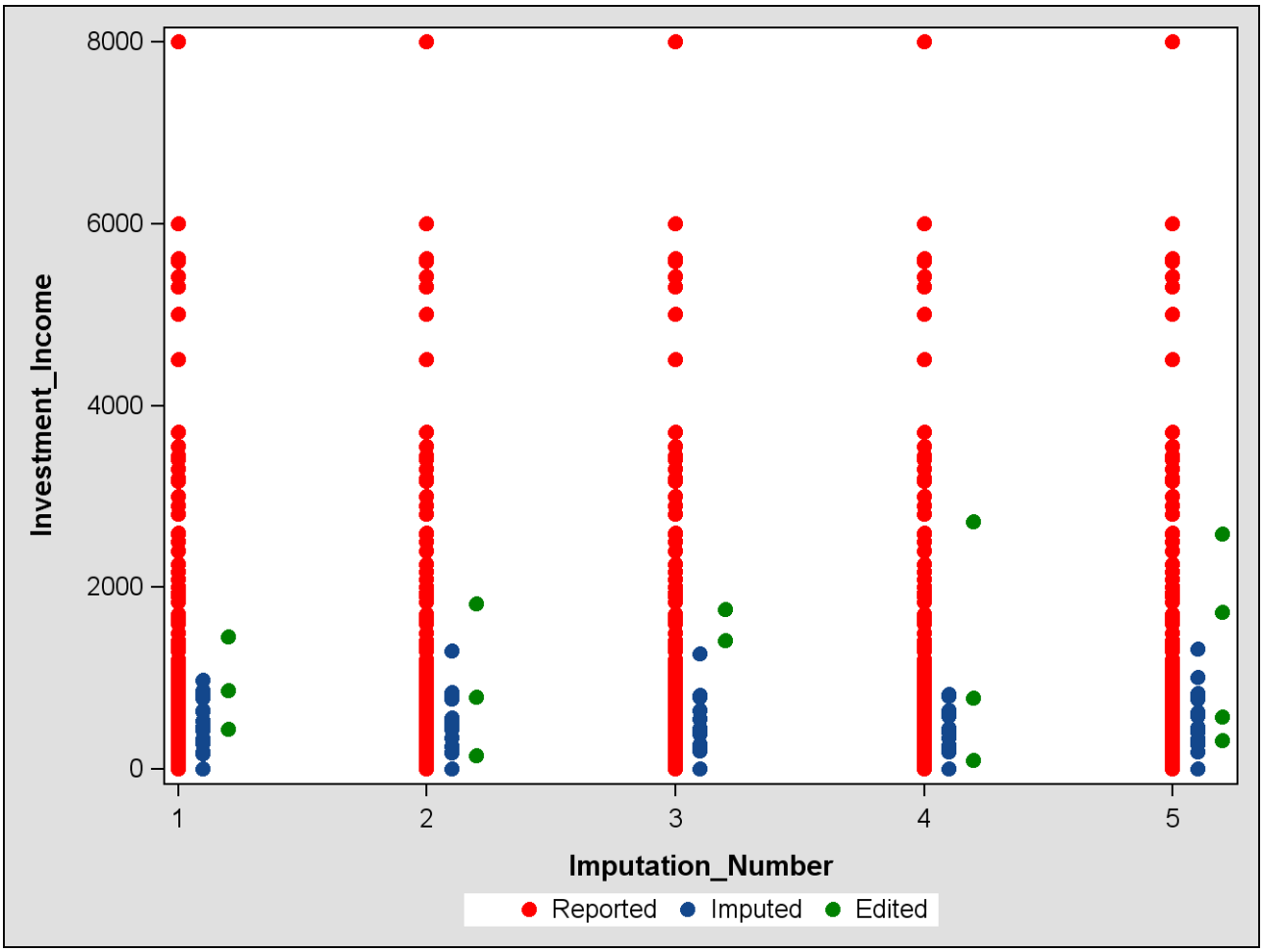


Figure D-1. Distribution of reported and imputed values for each income component (continued)

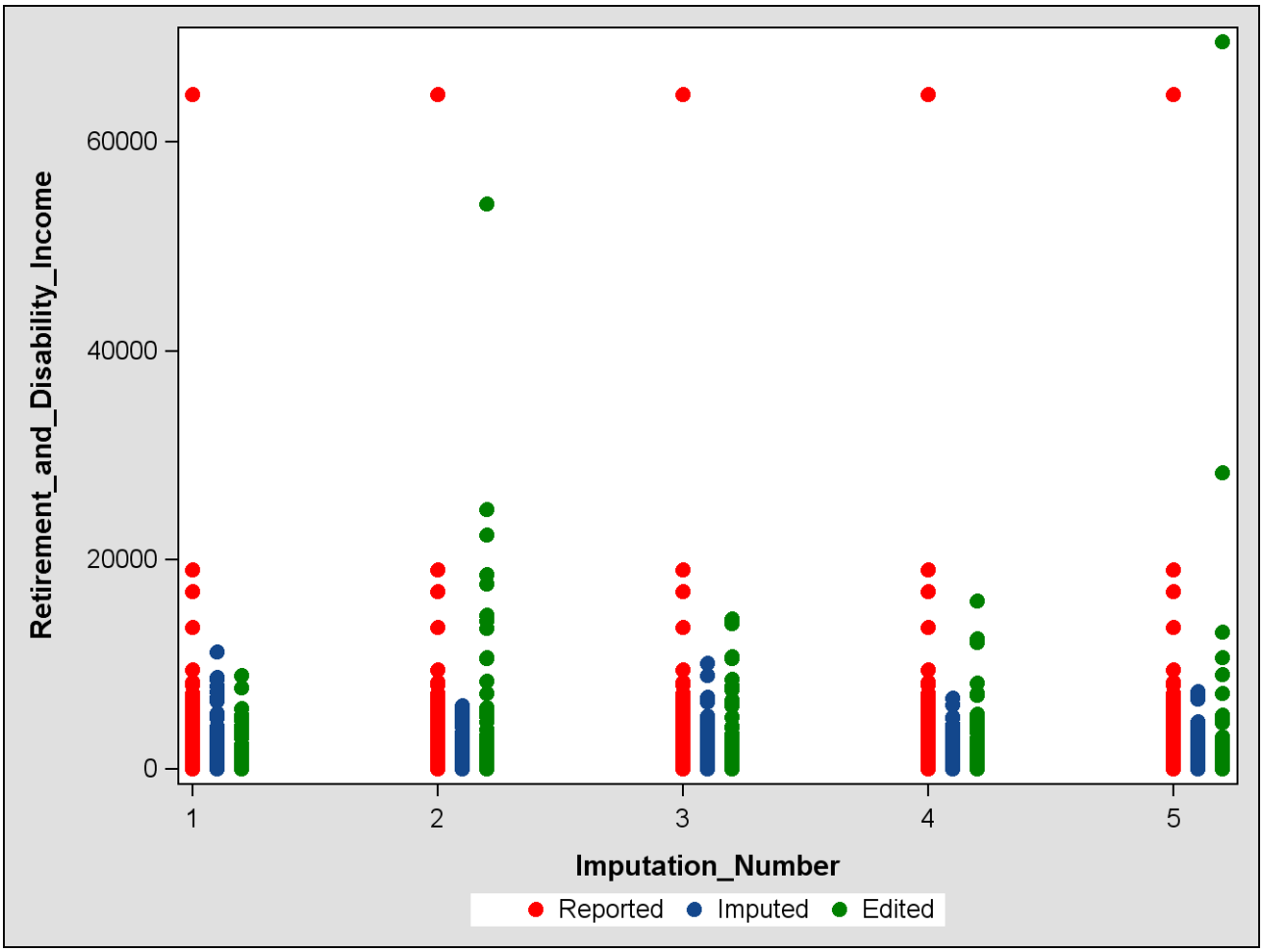


Figure D-1. Distribution of reported and imputed values for each income component (continued)

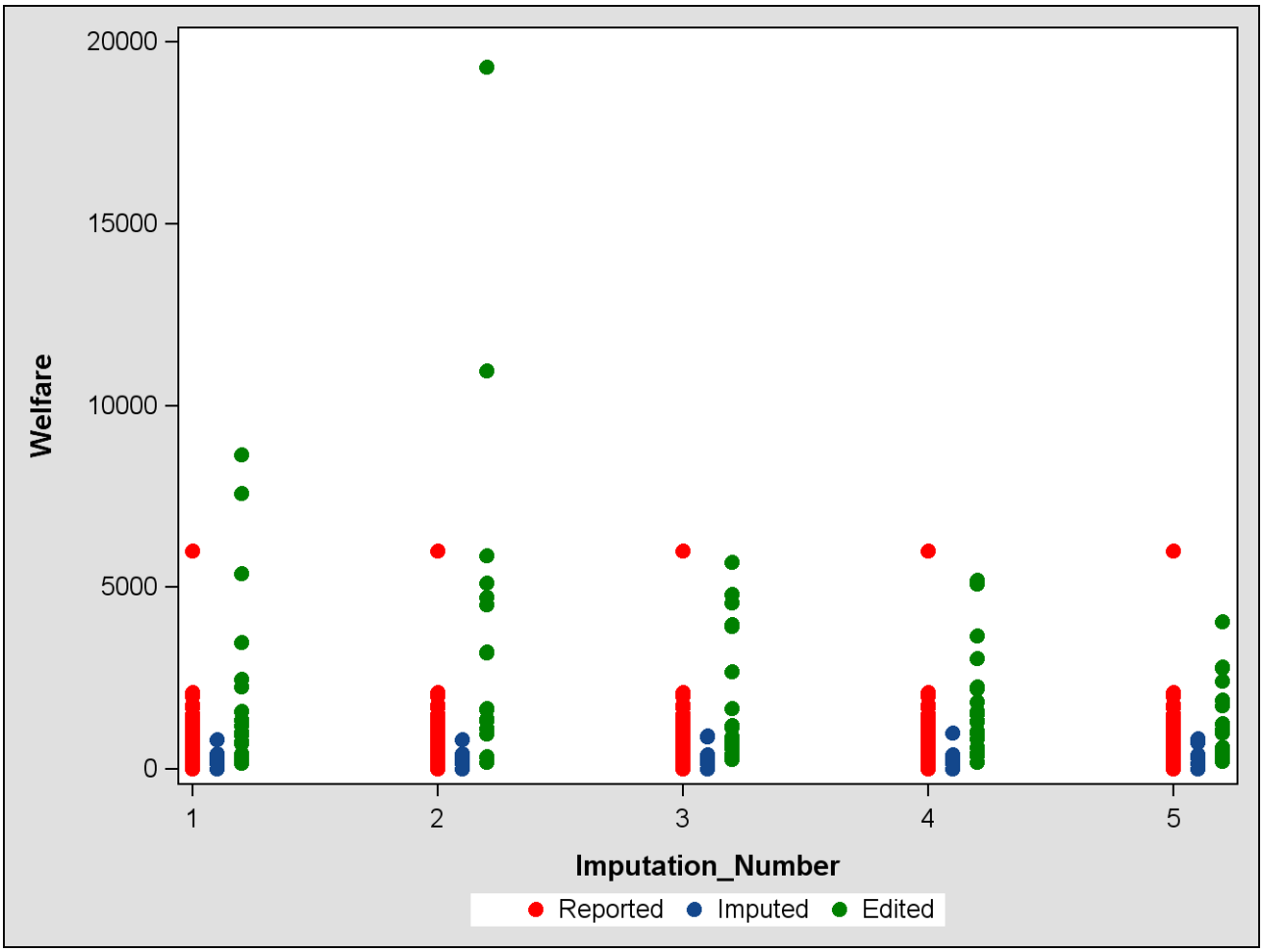


Figure D-1. Distribution of reported and imputed values for each income component (continued)

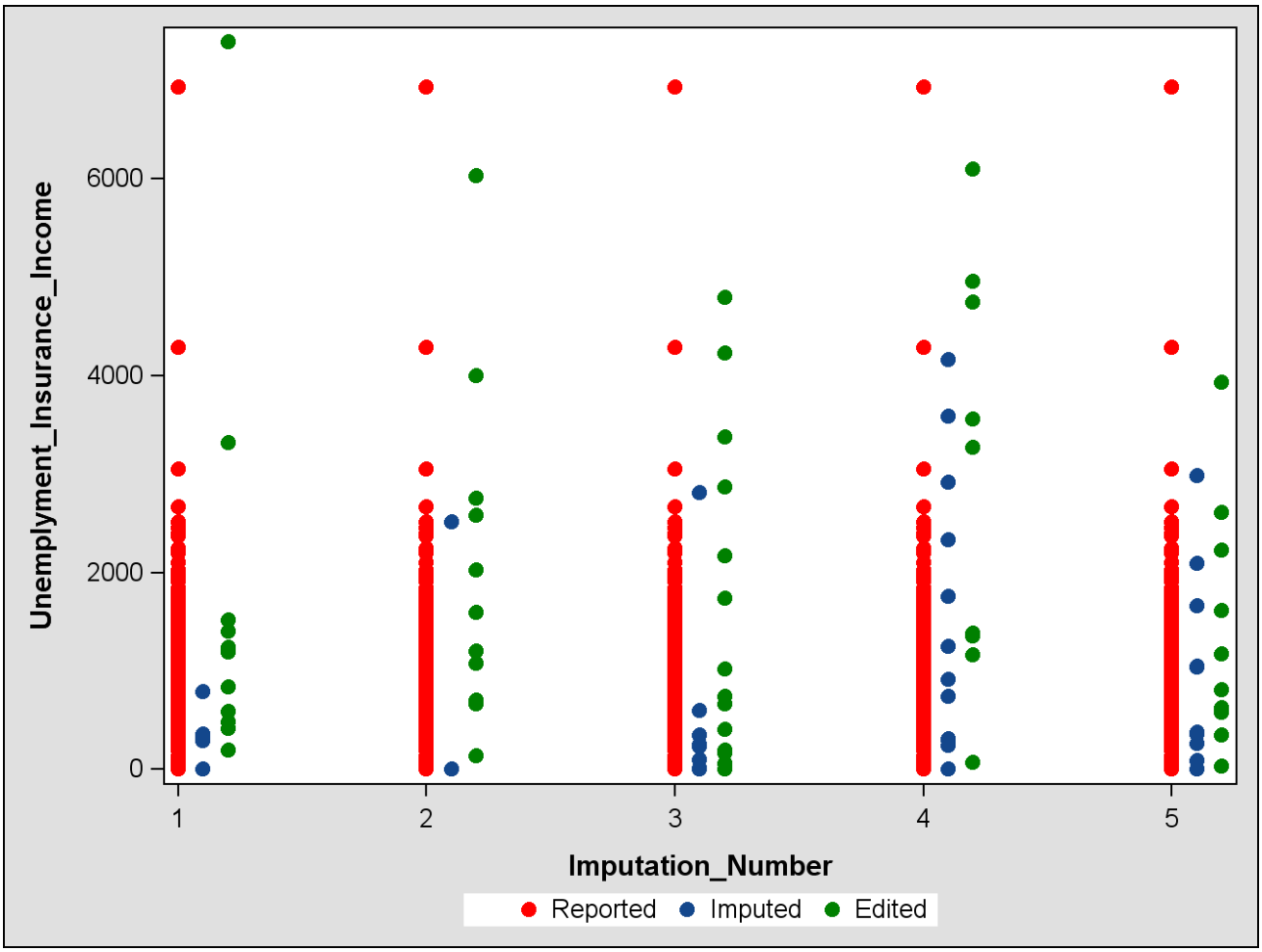


Figure D-1. Distribution of reported and imputed values for each income component (continued)

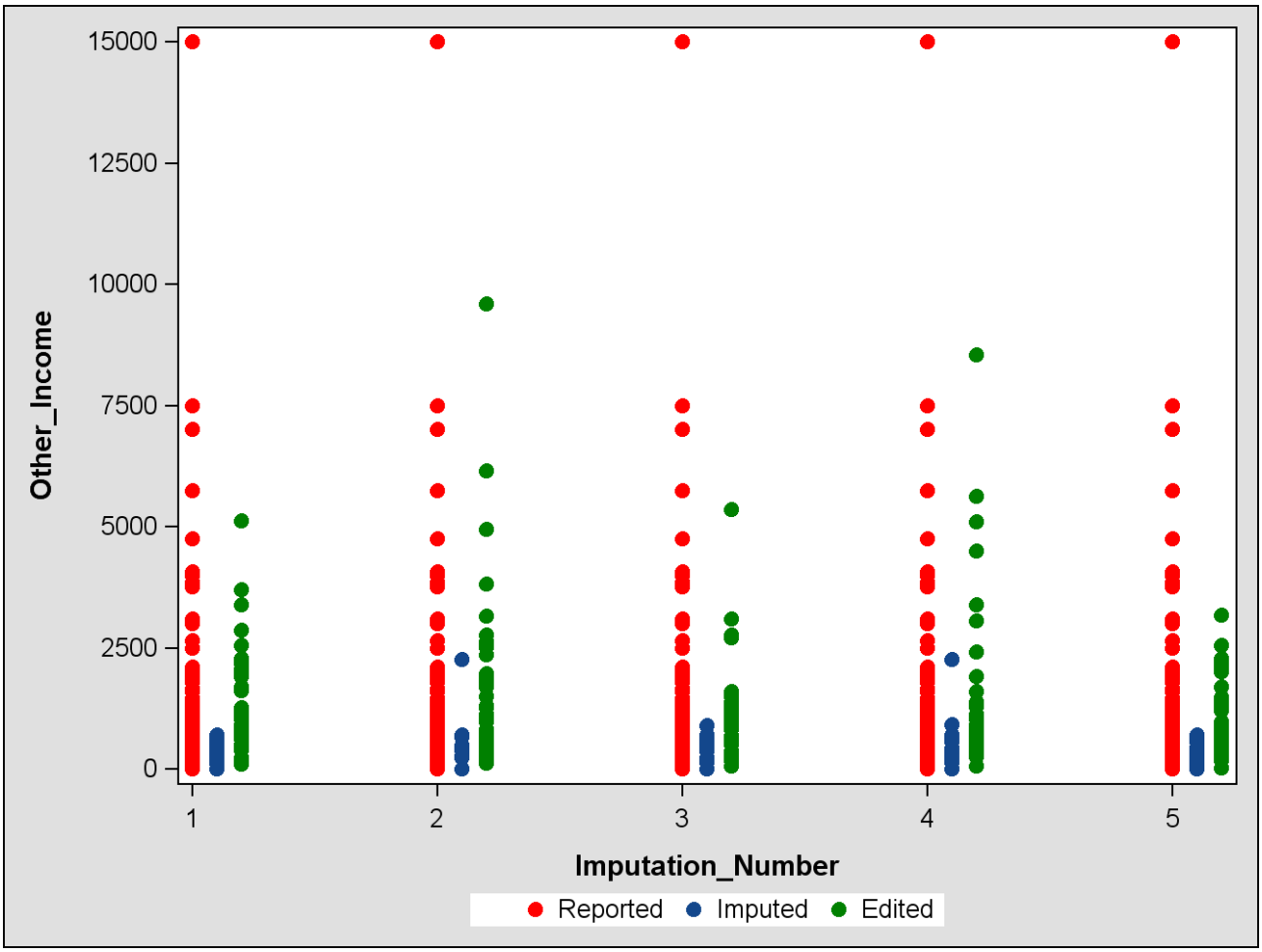
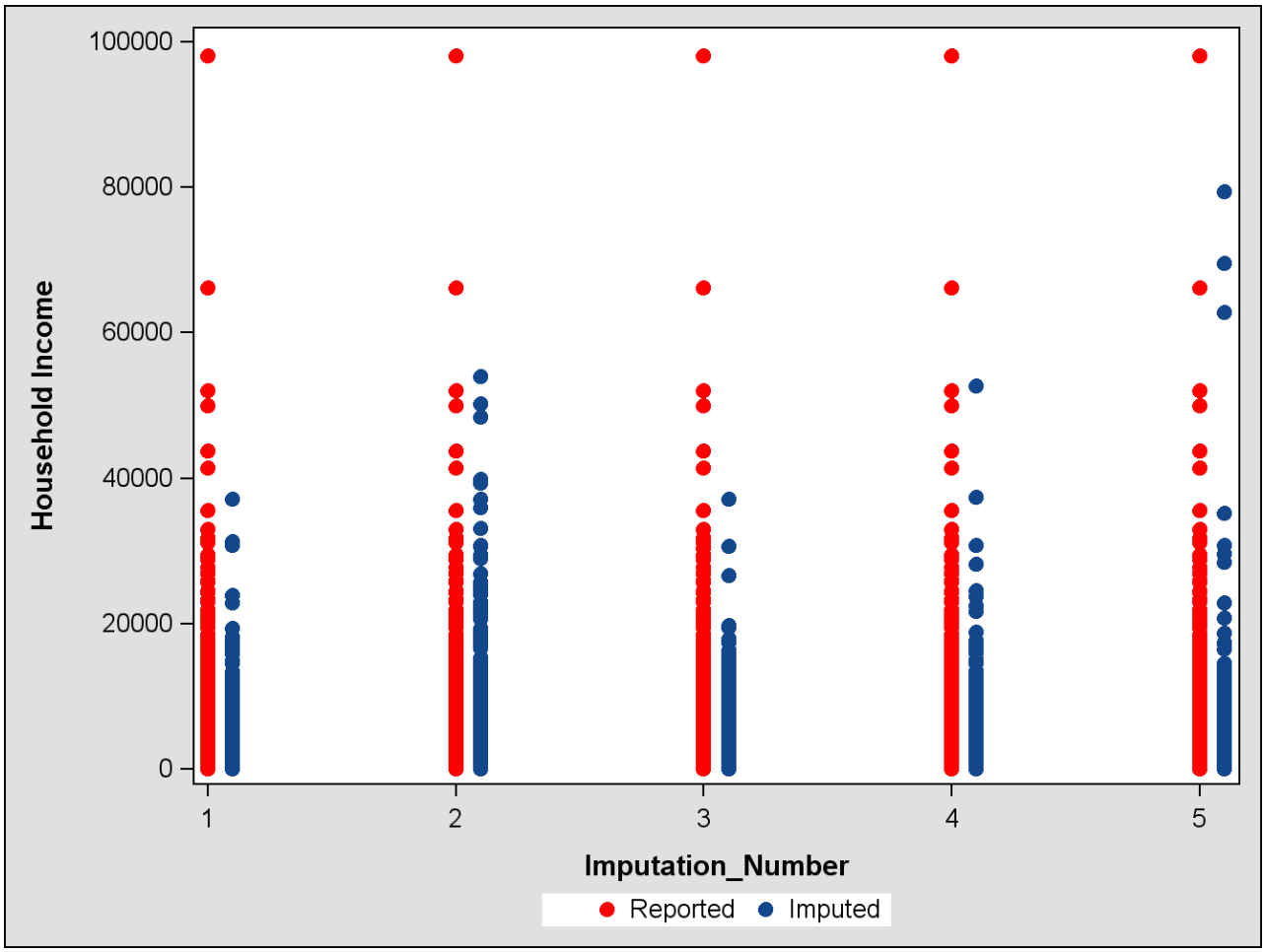


Figure D-2. Reported and imputed household income



Appendix E

Tables for FAH Price Imputation

Appendix E
Tables for FAH Price Imputation

Table E-1. Bivariate NRBA for FAH item price

Subgroup		Sample size	Weighted response rate	Chi-square[1]	
				Statistic	p-value
OVERALL		138,855	95.4		
Place type	Superstore	61,758	98.0	72.81	<.0001
	Supermarket	55,037	96.9		
	Convenience store or small grocery	4,737	87.4		
	Other food stores	16,413	84.6		
	Eating places	239	67.1		
	All other places	499	76.8		
	Unknown	172	59.3		
Package size unit	OZ, LITER	83,758	96.4	30.22	<.0001
	LBS, DRYOZ, GRAM	1,386	89.7		
	COUNT, PIECE	1,894	94.8		
	OTHER	1,247	90.8		
	NOT APPLICABLE	10,090	98.2		
	MISSING	40,480	93.0		
IRI department	DEPT-BEVERAGES	14,398	94.8	12.52	0.0019
	DEPT-FROZEN	13,421	96.6		
	DEPT-GENERAL FOOD	55,192	96.3		
	DEPT-HEALTH	258	98.0		
	DEPT-LIQUOR	1,252	89.1		
	DEPT-REFRIGERATED	51,407	95.4		
	Missing	2,927	82.0		
Quantity	One	133,586	95.7	26.49	<.0001
	More than one	5,269	87.8		
Barcode source	No barcode	32,681	94.9	10.07	0.0024
	Scanned UPC on item	79,922	97.0		
	Scanned other barcode on item	2,107	73.9		
	Scanned Food Book barcode	8,572	88.8		
	Assigned Food Book barcode	13,200	94.8		
	Extracted UPC from item description	2,373	100.0		
Number of items	1 to 2	5,068	80.3	172.03	<.0001
	3 to 5	12,732	88.6		
	6 to 12	28,786	93.5		
	13 to 25	36,560	96.7		
	26 or more	55,709	98.6		
Loyalty card	No	89,694	95.2	15.43	<.0001
	Yes	38,340	97.4		
	Missing	10,821	88.9		

Table E-1. Bivariate NRBA for FAH item price (continued)

Subgroup		Sample size	Weighted response rate	Chi-square[1]	
				Statistic	p-value
Paid with cash	No	92,947	97.4	107.43	<.0001
	Yes	44,844	92.5		
	Missing	1,064	10.4		
Paid with SNAP EBT	No	104,787	95.8	19.54	<.0001
	Yes	33,004	98.2		
	Missing	1,064	10.4		
Household size	1	16,623	96.2	7.53	0.0503
	2	34,535	95.9		
	3 or 4	52,141	94.3		
	5 or 6	26,039	96.7		
	7 or more	9,517	95.3		
Target group	Income < 100% poverty guideline	9,705	91.9	7.67	0.0312
	Income 100-185% of poverty guideline	22,514	94.3		
	Income > 185% of poverty guideline	57,003	95.9		
	SNAP	49,633	95.2		
WIC household	Yes	16,754	93.5	3.88	0.0489
	No or don't know	122,101	95.5		
Financial condition	Very comfortable and secure	18,934	95.4	5.03	0.1029
	Able to make ends meet without much difficulty	39,281	96.4		
	Occasionally have some difficulty making ends meet	42,983	95.2		
	Tough to make ends meet but keeping your head above water	29,893	93.4		
	In over your head	7,585	94.8		
	Missing	179	98.6		
FNS region	Mid-Atlantic	12,610	95.6	5.01	0.2837
	Midwest	23,762	94.9		
	Mountains/plains	9,531	96.2		
	Northeast	13,011	95.5		
	Southeast	30,142	96.7		
	Southwest	19,385	95.7		
	West	30,414	94.2		
Low access	1st quartile	31,788	94.9	2.22	0.4367
	2nd quartile	48,179	95.8		
	3rd quartile	35,565	96.1		
	4th quartile	23,323	94.3		
Low income-low access	1st quartile	43,190	95.0	0.90	0.7384
	2nd quartile	48,229	95.7		
	3rd quartile	37,545	95.7		
	4th quartile	9,891	95.2		

[1] Test of independence between response status and the subgroup variable. Observations with a subgroup value of "Missing" or "Unknown" are excluded from the test.

Table E-2. Mean FAH item price before and after hot deck imputation

Subgroup	Reported (n=132,000)		Reported + Imputed (n=135,927)		Relative difference[1]
	Mean	S.E.	Mean	S.E.	
Overall	2.92	0.044	2.91	0.042	-0.145
Place type					
Superstore	2.65	0.032	2.65	0.031	0.014
Supermarket	2.77	0.047	2.77	0.046	-0.029
Convenience store or small grocery	2.96	0.097	2.89	0.108	-0.679
Other food stores	4.56	0.257	4.41	0.239	-0.570
Eating places	5.24	0.597	4.94	0.530	-0.506
All other places	3.70	0.835	3.56	0.739	-0.167
Unknown	4.13	1.400	5.08	1.473	0.677
Package size unit					
OZ, LITER	2.69	0.037	2.69	0.036	-0.075
LBS, DRYOZ, GRAM	3.46	0.130	3.52	0.139	0.469
COUNT, PIECE	2.64	0.112	2.62	0.108	-0.143
OTHER	3.18	0.188	3.18	0.174	0.021
NOT APPLICABLE	2.52	0.073	2.52	0.074	0.001
MISSING	3.50	0.085	3.50	0.085	0.000
IRI department					
DEPT-BEVERAGES	2.74	0.073	2.73	0.070	-0.138
DEPT-FROZEN	3.93	0.090	3.93	0.088	-0.101
DEPT-GENERAL FOOD	2.36	0.038	2.36	0.038	-0.125
DEPT-HEALTH	9.04	1.022	8.92	1.005	-0.121
DEPT-LIQUOR	8.92	0.448	8.51	0.478	-0.915
DEPT-REFRIGERATED	2.98	0.051	2.97	0.050	-0.051
Missing	3.83	0.169	3.83	0.169	0.000
Quantity					
One	2.95	0.044	2.94	0.043	-0.177
More than one	1.95	0.090	2.03	0.081	0.907
Barcode source					
No barcode	3.86	0.114	3.84	0.112	-0.231
Scanned UPC on item	2.68	0.036	2.67	0.035	-0.076
Scanned other barcode on item	5.25	0.268	5.24	0.271	-0.024
Scanned Food Book barcode	1.90	0.060	1.91	0.059	0.107
Assigned Food Book barcode	2.67	0.059	2.66	0.059	-0.082
Extracted UPC from item description	2.72	0.173	2.72	0.173	0.001
Number of items					
1 to 2	4.45	0.202	4.23	0.179	-1.102
3 to 5	3.39	0.098	3.35	0.095	-0.450
6 to 12	3.13	0.072	3.10	0.070	-0.297
13 to 25	2.80	0.065	2.80	0.063	-0.004
26 or more	2.67	0.053	2.67	0.052	-0.004
Loyalty card					
No	2.83	0.041	2.83	0.040	-0.056
Yes	3.03	0.066	3.02	0.067	-0.115
Missing	3.22	0.251	3.18	0.237	-0.184

Table E-2. Mean FAH item price before and after hot deck imputation (continued)

Subgroup	Reported (n=132,000)		Reported + imputed (n=135,927)		Relative difference[1]
	Mean	S.E.	Mean	S.E.	
Paid with cash					
No	3.02	0.045	3.02	0.044	-0.146
Yes	2.64	0.080	2.65	0.077	0.046
Missing	3.36	0.930	2.72	0.193	-0.686
Paid with SNAP EBT					
No	2.92	0.043	2.92	0.042	-0.131
Yes	2.86	0.160	2.85	0.159	-0.009
Missing	3.36	0.930	2.72	0.193	-0.686
Household size					
1	2.85	0.111	2.84	0.109	-0.052
2	2.94	0.050	2.93	0.048	-0.136
3 or 4	2.96	0.063	2.95	0.061	-0.127
5 or 6	2.90	0.074	2.89	0.073	-0.040
7 or more	2.81	0.097	2.81	0.094	-0.035
Target group					
Income < 100% poverty guideline	2.70	0.095	2.69	0.093	-0.061
Income 100-185% of poverty guideline	2.53	0.074	2.54	0.071	0.120
Income > 185% of poverty guideline	3.05	0.054	3.04	0.053	-0.151
SNAP	2.64	0.043	2.64	0.044	-0.043
WIC household					
Yes	2.59	0.074	2.60	0.073	0.140
No or don't know	2.94	0.047	2.93	0.045	-0.148
Financial condition					
Very comfortable and secure	3.15	0.068	3.15	0.068	-0.081
Able to make ends meet without much difficulty	2.95	0.057	2.94	0.055	-0.119
Occasionally have some difficulty making ends meet	2.84	0.082	2.84	0.080	-0.020
Tough to make ends meet but keeping your head above water	2.58	0.054	2.57	0.052	-0.202
In over your head	2.60	0.113	2.61	0.112	0.060
Missing	3.54	0.487	3.52	0.487	-0.042
FNS region					
Mid-Atlantic	2.72	0.165	2.72	0.161	0.004
Midwest	2.83	0.079	2.82	0.076	-0.133
Mountains/Plains	2.71	0.031	2.73	0.029	0.431
Northeast	3.16	0.272	3.16	0.272	-0.013
Southeast	2.89	0.063	2.89	0.063	-0.070
Southwest	2.71	0.126	2.70	0.126	-0.100
West	3.26	0.116	3.25	0.112	-0.088
Low access					
1st quartile	2.94	0.134	2.93	0.129	-0.033
2nd quartile	2.93	0.086	2.93	0.085	0.013
3rd quartile	2.88	0.059	2.86	0.060	-0.213
4th quartile	2.94	0.142	2.92	0.137	-0.101

Table E-2. Mean FAH item price before and after hot deck imputation (continued)

Subgroup	Reported (n=132,000)		Reported + imputed (n=135,927)		Relative difference[1]
	Mean	S.E.	Mean	S.E.	
Low income-low access					
1st quartile	3.07	0.094	3.07	0.091	-0.049
2nd quartile	2.90	0.079	2.90	0.076	-0.013
3rd quartile	2.80	0.059	2.79	0.060	-0.208
4th quartile	2.77	0.178	2.75	0.176	-0.084

[1] The relative difference is the difference between the means divided by the standard error of the mean for reported items.

Appendix F

Tables for FAFH Price Imputation

Appendix F
Tables for FAFH Price Imputation

Table F-1. Bivariate NRBA for FAFH item price

Subgroup	Value	Weighted		Chi-square[1]	
		Sample size	Response rate	Statistic	p-value
OVERALL	OVERALL	59,893	92.0		
Menu group	Beverages (excl. specialty beverages from top restaurant)	17,103	93.1	42.04	<.0001
	Food items not obtained from a top restaurant	22,870	90.0		
	Food items or specialty beverages from a top restaurant	19,920	93.6		
Place type	Top fast food restaurants	23,496	94.6	29.18	<.0001
	Top non-fast food restaurants	3,846	91.6		
	Food store	7,441	94.3		
	Other, restaurant	17,843	89.6		
	Other, non-restaurant	5,324	91.2		
	Missing	1,943	85.9		
Bundle indicator	Not part of a bundle	46,804	90.7	86.72	<.0001
	Part of a bundle	13,089	96.4		
Relative beverage size	Not a beverage	42,790	91.5	26.86	<.0001
	X-small or small	4,037	93.4		
	Medium	4,867	94.5		
	Large	3,144	95.7		
	X-large	549	92.5		
	Missing	4,506	89.9		
Quantity	One	50,109	92.4	0.93	0.3353
	More than one	9,354	91.7		
	Missing	430	34.5		
Book type	Adult	16,066	91.4	2.30	0.2867
	Primary	41,240	92.2		
	Youth	2,587	89.3		

Table F-1. Bivariate NRBA for FAFH item price (continued)

Subgroup	Value	Weighted		Chi-square[1]	
		Sample size	Response rate	Statistic	p-value
Number HH members that shared meal	1	34,031	93.4	8.08	0.0228
	2	14,812	90.6		
	3 to 4	8,743	91.8		
	5+	1,798	85.6		
	Missing	509	57.4		
Breakfast	No	50,299	92.3	0.01	0.9223
	Yes	7,964	92.5		
	Missing	1,630	75.2		
Lunch	No	36,415	92.4	0.01	0.9094
	Yes	21,848	92.3		
	Missing	1,630	75.2		
Dinner	No	38,216	92.9	3.91	0.0481
	Yes	20,047	91.4		
	Missing	1,630	75.2		
Snack or drink	No	48,823	92.1	4.89	0.0270
	Yes	9,440	93.9		
	Missing	1,630	75.2		
Paid with cash	No	23,193	92.0	0.16	0.6868
	Yes	35,055	92.4		
	Missing	1,645	77.3		
Target group	Income < 100% poverty guideline	4,215	88.5	4.67	0.1338
	Income 100-185% of poverty guideline	9,591	92.5		
	Income > 185% of poverty guideline	31,107	92.3		
	SNAP	14,980	90.6		
WIC household	Yes	6,050	90.7	0.94	0.3314
	No or don't know	53,843	92.0		

Table F-1. Bivariate NRBA for FAFH item price (continued)

Subgroup	Value	Weighted		Chi-square[1]	
		Sample size	Response rate	Statistic	p-value
Financial condition	Very comfortable and secure	9,263	92.6	1.67	0.6881
	Able to make ends meet without much difficulty	18,033	91.7		
	Occasionally have some difficulty making ends meet	17,768	91.7		
	Tough to make ends meet but keeping your head above water	12,000	92.3		
	In over your head	2,773	89.0		
	Missing	56	92.4		
FNS region	Mid-Atlantic	5,374	93.2	14.53	0.0114
	Midwest	9,677	93.1		
	Mountains/Plains	5,251	93.7		
	Northeast	4,718	89.8		
	Southeast	14,363	91.2		
	Southwest	8,372	88.8		
	West	12,138	93.2		

[1] Test of independence between response status and the subgroup variable. Observations with a subgroup value of "Missing" or "Unknown" are excluded from the test.

Table F-2. Mean FAFH item price of purchased, non-school items before and after imputation

Subgroup	Reported (n=54,441)		Reported + Imputed (n=59,094)		Relative difference[1]
	Mean	S.E.	Mean	S.E.	
Overall	2.81	0.069	2.89	0.064	1.148
Menu group					
Beverages (excl. specialty beverages from top restaurant)	1.58	0.070	1.60	0.068	0.388
Food items not obtained from a top restaurant	3.65	0.095	3.73	0.086	0.879
Food items or specialty beverages from a top restaurant	2.90	0.068	2.99	0.066	1.349
Place type					
Top fast food restaurants	2.21	0.047	2.25	0.045	0.956
Top non-fast food restaurants	3.46	0.185	3.68	0.167	1.186
Food store	2.13	0.079	2.11	0.076	-0.195
Other, restaurant	3.56	0.132	3.65	0.116	0.661
Other, non-restaurant	2.25	0.115	2.26	0.111	0.139
Missing	3.38	0.212	3.43	0.206	0.212
Bundle indicator					
Not part of a bundle	3.07	0.078	3.10	0.073	0.438
Part of a bundle	1.95	0.079	2.14	0.082	2.409
Relative beverage size					
Not a beverage	3.33	0.075	3.42	0.070	1.230
X-small or small	1.27	0.074	1.30	0.075	0.317
Medium	1.40	0.056	1.44	0.052	0.678
Large	1.58	0.103	1.60	0.096	0.158
X-large	2.46	0.375	2.41	0.353	-0.130
Missing	1.98	0.153	1.99	0.141	0.096
Quantity					
One	2.95	0.077	3.02	0.073	0.907
More than one	2.07	0.054	2.14	0.052	1.374
Missing[2]	0.48	0.235	3.33	0.502	12.132
Book type					
Adult	2.61	0.088	2.69	0.086	0.957
Primary	2.90	0.074	2.98	0.068	1.069
Youth	2.17	0.072	2.25	0.063	1.109
Number HH members that shared meal					
1	2.58	0.069	2.62	0.065	0.622
2	3.07	0.116	3.17	0.104	0.819
3 to 4	3.23	0.123	3.31	0.122	0.678
5+	3.75	0.639	4.03	0.512	0.430
Missing	3.31	0.556	3.98	0.597	1.199
Breakfast					
No	2.95	0.077	3.02	0.073	0.953
Yes	1.91	0.052	1.96	0.049	0.859
Missing	2.93	0.285	3.35	0.293	1.473
Lunch					
No	2.93	0.089	2.99	0.084	0.657
Yes	2.60	0.068	2.69	0.068	1.315
Missing	2.93	0.285	3.35	0.293	1.473

Table F-2. Mean FAFH item price of purchased, non-school items before and after imputation (continued)

Subgroup	Reported (n=54,441)		Reported + imputed (n=59,094)		Relative difference[1]
	Mean	S.E.	Mean	S.E.	
Dinner					
No	2.35	0.051	2.42	0.051	1.197
Yes	3.63	0.130	3.70	0.122	0.531
Missing	2.93	0.285	3.35	0.293	1.473
Snack or drink					
No	2.93	0.074	3.01	0.070	1.083
Yes	2.14	0.079	2.14	0.075	0.016
Missing	2.93	0.285	3.35	0.293	1.473
Paid with cash					
No	3.35	0.108	3.43	0.100	0.799
Yes	2.40	0.069	2.47	0.064	0.951
Missing	2.42	0.157	2.61	0.196	1.241
Target group					
Income < 100% poverty guideline	2.89	0.155	2.93	0.136	0.305
Income 100-185% of poverty guideline	2.37	0.076	2.46	0.073	1.275
Income > 185% of poverty guideline	2.91	0.083	2.99	0.077	0.867
SNAP	2.34	0.067	2.47	0.061	1.985
WIC household					
Yes	2.30	0.078	2.44	0.073	1.763
No or don't know	2.84	0.070	2.91	0.066	1.080
Financial condition					
Very comfortable and secure	3.21	0.140	3.32	0.130	0.770
Able to make ends meet without much difficulty	2.75	0.063	2.81	0.057	0.915
Occasionally have some difficulty making ends meet	2.60	0.064	2.69	0.061	1.413
Tough to make ends meet but keeping head above water	2.58	0.120	2.66	0.112	0.610
In over your head	2.65	0.291	2.67	0.264	0.060
Missing	4.50	0.394	4.43	0.372	-0.157
FNS region					
Mid-Atlantic	2.73	0.077	2.80	0.078	0.925
Midwest	2.63	0.189	2.71	0.171	0.432
Mountains/Plains	2.65	0.207	2.69	0.207	0.196
Northeast	3.49	0.163	3.53	0.150	0.285
Southeast	2.62	0.131	2.70	0.126	0.576
Southwest	2.72	0.090	2.87	0.081	1.604
West	3.19	0.132	3.26	0.118	0.524

[1] The relative difference is the difference between the means divided by the standard error of the mean for reported items.

[2] Less than 1 percent of items have a missing value for QUANTITY, and approximately two-thirds of items with a missing value of QUANTITY had a missing price prior to imputation.