OXFORD UNIVERSITY PRESS | Journal of Survey Statistics and Methodo

Estimation of Underreporting in Diary Surveys: An Application using the National Household Food Acquisition and Purchase Survey

Journal:	*Journal of Survey Statistics and Methodology*
Manuscript ID	JSSAM-2018-065.R2
Manuscript Type:	Survey Methodology
Date Submitted by the Author:	n/a
Complete List of Authors:	Hu, Mengyao; University of Michigan, Survey Research Center Kirlin, John; USDA Economic Research Service, United States Department of Agriculture West, Brady; University of Michigan-Ann Arbor, Institute for Social Research He, Wenyi; University of Michigan, Survey Research Center Ong, Ai Rene; University of Michigan, Survey Research Center Zhang, Shiyu; University of Michigan, Institute for Social Research Zhang, Xingyou; USDA Economic Research Service, Department of Agriculture
Keywords:	diary surveys, National Household Food Acquisition and Purchase Survey (FoodAPS), item non-response error, imputation

SCHOLARONE™ Manuscripts

Estimation of Underreporting in Diary Surveys: An Application using the National Household Food Acquisition and Purchase Survey[1]

Mengyao Hu[2]* (maggiehu@umich.edu)

John A. Kirlin[3] (johnkir516@hotmail.com)

Brady T. West[2] (bwest@umich.edu)

Wenyi He[2] (hwenyi@umich.edu)

Ai Rene Ong[2] (aireneo@umich.edu)

Shiyu Zhang[2] (zsy@umich.edu)

Xingyou Zhang[4] (xingyou.zhang@gmail.com)

For Review Only

- Word count for the entire manuscript: 7,272
- Number of tables: 6
- Number of figures:1
- Online supplementary material**:**
    - Appendix A. Description of the covariates.
    - Appendix B. Descriptive statistics for the sample.
    - Appendix C. Heckman Selection Model Results Details.
    - Appendix D. Estimation and prediction of number of items using the negative binomial regression model.
    - Appendix E. Stata and R code illustrating the imputation methodology.
    - Supplemental Tables 1-7

---

[1] This work was supported via a Cooperative Assistance Grant from the Economic Research Service of the United States Department of Agriculture [59-5000-5-0008]. The findings and conclusions in this preliminary publication have not been formally disseminated by the U.S. Department of Agriculture and should not be construed to represent any agency determination or policy.

[2] Survey Research Center, Institute for Social Research, University of Michigan-Ann Arbor, Michigan

[3] Kirlin Analytic Services, Fredericksburg, VA, previously with the Economic Research Service, United States Department of Agriculture, Washington, DC

[4] Economic Research Service, United States Department of Agriculture, Washington, DC

* To whom correspondence should be addressed. Email: maggiehu@umich.edu; mailing address: 426 Thompson Street, Ann Arbor, MI 48104-1248.

ABSTRACT

Diary surveys are used to collect data on a variety of topics, including health, time use, nutrition and expenditures. The U.S. National Household Food Acquisition and Purchase Survey (FoodAPS) is a nationally representative diary survey, providing an important data source for decision makers to design policies and programs for promoting healthy lifestyles. Unfortunately, a multi-day diary survey like the FoodAPS can be subject to various survey errors, especially item nonresponse error occurring at the day level. The FoodAPS public-use data set provides survey weights that adjust only for unit nonresponse. Due to the lack of day-level weights (which could possibly adjust for the item nonresponse that arises from refusals on particular days), the adjustments for unit nonresponse are unlikely to correct any bias in estimates arising from households that initially agree to participate in FoodAPS, but then fail to report on particular days. This paper develops a general methodology for estimating the extent of underreporting due to this type of item nonresponse errors in diary surveys, using FoodAPS as a case study. We describe a methodology combining bootstrap replicate sampling for complex samples and imputation based on a Heckman selection model to predict food expenditures for person-days with missing expenditures. We estimated the item nonresponse error by comparing weighted estimates according to 1) only reported expenditures and 2) both reported expenditures and predictions for missing values. Results indicate that ignoring the missing data would lead to consistent over-estimation of the mean expenditures and events per person per day and under-estimation of the total expenditures and events. Our study suggests that the household-level weights which generally account for unit nonresponse may not be entirely sufficient for addressing the nonresponse occurring at the day level in diary surveys, and proper imputation methods will be important for estimating the size of the underreporting.

For Review Only

Keywords: diary surveys; National Household Food Acquisition and Purchase Survey (FoodAPS); item non-response error; imputation

For Review Only

INTRODUCTION

The U.S. National Household Food Acquisition and Purchase Survey (FoodAPS) is a nationally representative diary survey of the continental United States that provides information on all food acquisitions and expenditures of all household members. As the first detailed national survey on household food acquisitions, FoodAPS is important for tracking food expenditures and understanding how food access, food choice and federal nutrition assistance programs are related to food security and various health outcomes. Data from FoodAPS are also important for policy and decision makers, enabling them to design effective policies and develop programs that will promote healthy eating and healthy lifestyles. Given the importance of these data, it is important that FoodAPS provides accurate and complete data that can be used effectively by researchers, practitioners and policymakers.

The FoodAPS collects information about all types of food acquisition and expenditure data using a seven-day diary survey. The broad range of food items collected and the requirement to record food acquisitions for everyone in the household pose high burdens on respondents. Similar to other diary surveys such as expenditure, time use, activity and media diary surveys (Fricker et al. 2015; Groves et al. 2009; Krizek et al. 2009; Mathiowetz et al. 2011), the FoodAPS faces various types of nonresponse error – sampled households may refuse to participate in the survey in any way (unit nonresponse error), or a household member/respondent may agree to participate but fail to provide information on a particular day of the seven-day data collection period (item nonresponse error).

The purpose of the current study is to develop a general methodology that can estimate the extent of these item nonresponse errors (i.e., occurring at the person-day level) in diary surveys. We apply our methodology using FoodAPS in this study. Note that while we focus on a food expenditure diary survey in our application, we aim to develop a methodology that can be applied to repair this type of item nonresponse error in any diary survey asking respondents to record multiple outcomes over a short period of time.

Nonresponse and Underreporting in Diary Surveys

Even a well-designed and well-conducted diary survey can be subject to various threats that jeopardize the validity of the self-reported data. One major threat to the validity of diary survey data in general is nonresponse error. This type of error can be introduced when survey respondents either report for some but not all days of the diary period (item nonresponse) or simply refuse to participate on any of the days (unit nonresponse), and those who do participate are systematically different from those who do not in terms of the measures of interest.

These errors can pose major threats to data quality for an expenditure diary survey like FoodAPS, not only leading to biased estimates of food expenditures, but also distorting the relationships between food acquisitions and selected measures of health (e.g., body mass index (BMI) or self-reported health). In addition, since expenditure surveys like FoodAPS are important for policy-making (e.g., the drafting and modification of programs to promote healthy eating), substantial bias in the survey estimates may lead to inaccurate policy decisions, which would have a series of adverse consequences for the promotion of healthy lifestyles and assistance to those who are food insecure.

Nonresponse and Underreporting Evaluation in FoodAPS

The nonresponse problem in FoodAPS has been documented in two previous studies. Hu et al. (2017) modeled respondents' probability of reporting food acquisition events during the diary process and found that reporting of food events declined over the course of the diary week, suggesting increased underreporting as the diary period progressed. Using paradata from FoodAPS, these authors also found that the likelihood of refusal on a given day of the week increased over the diary period, especially for those who started their diaries on a Sunday or Friday. Besides the starting day of the diary, other factors such as race, relationship to the primary respondent (PR), participation in the Supplemental Nutrition Assistance Program (SNAP) and household size are related to the likelihood of food-acquisition-event reporting – African American respondents, non-PR members, respondents in a household with a current SNAP member and those from larger households are more likely to not respond. Maitland and Li (2016) predicted the probability that a respondent did not report any expenditure on a given day and found that the probability of reporting days without any expenditure increases across days of the data collection week. This finding is consistent with Hu et al. (2017)'s study that nonresponse at the day level increased across the diary period. Factors like income and SNAP status were found to be significantly associated with the probability of respondents reporting – higher income and non-SNAP households are more likely to report expenditures.

While these previous studies have contributed to our understanding of the non-response errors in FoodAPS, one noteworthy weakness in these studies is that they focus only on evaluating whether there are nonresponse errors, instead of quantifying the *sizes* of these errors. Knowing the extent of underreporting is important to both survey researchers and those who would like to use these data, in that it can help researchers to decide whether to impute the missing data, or to use the data as is if there is no substantial nonresponse bias.

Previous Efforts to Evaluate the Magnitude of Underreporting in Diary Surveys

Diary surveys mainly rely on self-reported data. For example, for expenditure diary surveys, no studies of national populations have collected both self-reported data on expenditures and actual measurements of expenditures. The same is true for many other diary surveys (e.g. time use diaries). In other words, the true values for the measured entity are usually not available, preventing researchers from evaluating the magnitude of underreporting by directly comparing reported and actual values.

Given this lack of validation data, a common approach used in previous studies is to compare survey estimates to external benchmarks, such as estimates from other national household surveys and national accounts (e.g., Fricker et al. 2015). For example, estimates from the U.S. Consumer Expenditure Survey (which, similar to FoodAPS, has both an interview survey component and a diary survey component) have been compared to multiple external benchmarks, including estimates from the Panel Survey on Income Dynamics (PSID; Creech 2015; Li et al. 2010), the American Community Survey (ACS; Bureau of Labor Statistics 2010) and the Personal Consumption Expenditures (PCE) of the National Income and Product Accounts (NIPA; e.g., Passero et al. 2013). The success of this approach largely depends on at least two factors. The first is how well the benchmark survey matches with the survey of interest in terms of "target population, definitions of expenditure categories and reference periods" (Fricker et al. 2015). It is uncommon to find another data source which can perfectly match the survey of interest, and direct comparisons without fully understanding or adjusting for the differences can lead to invalid comparisons. The second factor involves the accuracy of the external benchmarks. As discussed by Fricker et al (2015), in most cases, the errors in the benchmark estimates are usually not fully understood themselves, posing great challenges to teasing out the reporting error from other error sources when comparing the diary survey of interest and the benchmark survey.

An ideal benchmark for FoodAPS does not exist. Clay et al. (2016) compared the food spending estimated in FoodAPS to several benchmarks, including the National Health and Nutrition Examination Survey (NHANES), CE and the Information Resources Consumer Network Panel (IRI), and found that estimated food spending from FoodAPS exceeds the estimates from the CE and IRI but falls below the NHANES estimate. While such comparisons provide valuable insight into FoodAPS estimates, it is still quite difficult to know which estimate is closer to true food spending, given that the error sources in these surveys are not fully known, and they differ from FoodAPS in many ways. For example, IRI uses a different sampling procedure, where households were selected using quota sampling in the Consumer Network panel (Clay et al., 2016). NHANES uses a set of seven questions on food spending based on past 30-day recall, instead of a diary survey. Although the CE includes a diary survey similar to that of FoodAPS, the CE prompts respondents with specific examples of food items. This may help respondents with the recall of the food items included in the examples, but not those not in the examples, also known as "part-category cuing" (Alba and Chattopadhyay 1985). Also, the CE has a slightly different reference period (the year of 2012) compared to FoodAPS (April 2012 to January 2013).

Given that there is no ideal benchmark for the FoodAPS survey, this study does not rely on using benchmarks to evaluate underreporting. Instead, we leverage the rich paradata in FoodAPS to identify which household members refused to participate in the survey and for which fielding days. Because interviewers in FoodAPS were required to confirm the reports of respondents on each day of the diary period through phone interviews, we also are able to examine the incidence of unconfirmed "reports" of no acquisitions. We then model the food expenditures on individual days using the reported data (which have also been confirmed with the interviewers), and use these models to predict expenditures on those days when respondents refused or failed to have their reports confirmed with the interviewers. Finally, we estimate the magnitude of underreporting (due to failures to report expenditures at day levels) by comparing the expenditures solely based on confirmed respondent reports with the expenditures including predictions of expenditures for those who refused or failed to confirm their non-expenditures.

Most surveys use adjustments for *unit* nonresponse (e.g., weighting adjustments based on response propensity) to account for households that do not respond to initial survey requests when computing estimates. In this particular survey, the FoodAPS public-use data set provides *household-level* weights that have been adjusted for the probability that a sampled household agreed to participate in the survey. However, no individual- or day-level weights are provided, which could possibly be used to adjust for the item nonresponse that arises from refusals on particular days. Due to this lack of day-level weights, adjustments using these household-level weights are unlikely to correct any bias in estimates of total expenditures or mean expenditures arising from households that do respond, but then fail to report expenditures on particular days. In short, these "typical" adjustments to survey weights will not be a sufficient nonresponse bias correction mechanism in this type of diary survey where estimation of expenditures is so important.

Featuring the use of paradata and the inclusion of a comprehensive set of predictors of food acquisition event reporting (Hu et al. 2017) and expenditures (Maitland and Li 2016), our modeling approach will allow us to estimate the magnitude of underreporting at the individual–day level, in the absence of the true data. Careful assessment of model fit and predictive ability is clearly essential, however, for this approach to be effective.

METHODS

Overview of the FoodAPS Design

Fielded between April 2012 and January 2013, FoodAPS was a nationally representative, seven-day, cross-sectional diary survey designed to provide information on food acquisitions and expenditures for everyone in the household. FoodAPS collected several types of food acquisition information, including both at-home and away-from-home purchases. Food-at-home purchases (FAH) refer to "food and drinks brought into the home" and used to prepare meals. Food-away-from-home (FAFH) acquisitions are defined as "meals, snacks, and drinks got outside the home" (FoodAPS User's Guide 2016). In total, 14,317 individuals living in 4,826 households participated in the FoodAPS survey. One can access public-use FoodAPS data at https://www.ers.usda.gov/foodaps. The FoodAPS screener response rate was 70.9%, and the overall FoodAPS response rate was 41.5% (Response Rate 3 of the American Association for Public Opinion Research; FoodAPS User's Guide 2016). More details on the response rates are available upon request.

Designed to provide information about the weekly food acquisitions of American households (the target population of FoodAPS), the FoodAPS sampling procedure identified four target groups of households, of which three groups were oversampled: 1) households with at least one member receiving SNAP benefits; 2) non-SNAP households with income below the poverty threshold; and 3) non-SNAP households with income between 100 and 185% of poverty. The fourth group included non-SNAP households with income equal to or greater than 185% of poverty. In this paper, we also evaluate the extent of underreporting for each of these subgroups.

The survey selected eligible households through a multistage sampling design. Before sampling, primary sampling units (PSUs) were defined as counties or groups of contiguous counties. In the first stage, a stratified sample of 50 PSUs was selected from two strata using probability proportionate to size (PPS) selection in each stratum. In the second stage, again using PPS selection, eight secondary sampling units (SSUs) were selected within each of the 50 PSUs. In the third stage, households were selected based on an address-based sampling frame, which combined the United States Postal Service Delivery Sequence File and a list of SNAP addresses. For analytical purposes, the USDA provides pseudo-strata and pseudo-PSU codes for variance estimation in FoodAPS (Heeringa et al., 2017, Chapter 4). In total, there are 25 unique pseudo-stratum codes and 57 unique pseudo-PSU codes in the data set.

A diary book was provided to respondents age 11 years and above to record their food purchases and acquisitions for seven days. The diary book collected information on the following items: who got the food, the place where the food items were obtained, food item details and expenditure details (e.g., amount paid and payment method). Prior to the diary survey data collection, the person reported as doing the most shopping for each household was identified as the household's primary respondent (PR), and the PR was responsible for reporting any food acquisitions made by children under 11 years old (e.g., meals at school if not brought from home). The PR also participated in two in-person interviews (before and after the one-week diary survey) and up to three telephone interviews. Figure 1 shows the overview of the planned FoodAPS data collection process. On days 2, 5, and 7 of the data collection week, the PR was supposed to call the telephone interview center, and the telephone interviewers were instructed to confirm with the PR each household member's food acquisitions and purchases on each day since the previous contact. When a PR failed to call in on a specified day, a telephone interviewer tried to call the PR the following morning.

For Review Only

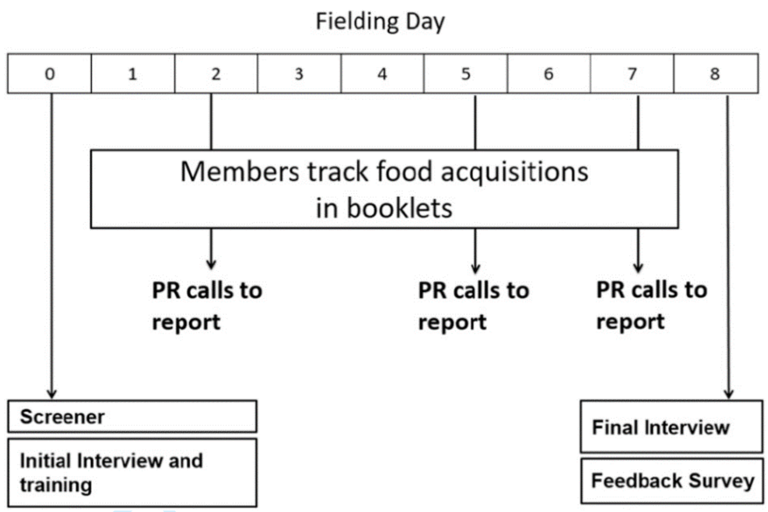

Figure 1. Overview of the planned data collection week of FoodAPS.

FoodAPS collected a rich set of paradata, including contact data, respondent feedback data and detailed records of daily reporting status. The feedback data come from a short, mail-in survey provided to the PR at the end of the final interview. The feedback survey asked questions related to ease of participating in FoodAPS. Daily reporting status information was collected through the three planned telephone interviews, where interviewers recorded and classified each individual's daily reporting of food purchases or acquisitions as 1) confirmed status of food acquisition (a food acquisition was reported to the interviewer – confirmed yes, or no food acquisitions were reported and the PR confirmed that the member did not have any FAH or FAFH events that day – confirmed no); 2) refusal (the PR reported that the household member refused to participate on that day); or 3) unconfirmed (no food acquisitions were reported by a member, but the PR could not confirm that no acquisitions occurred). In this study, we use the daily reporting statuses of individuals to identify those who are non-respondents (refusals and unconfirmed cases).

For Review Only

FoodAPS unconditionally offered a $5 incentive to all households contacted during the screening interviews. Interviewers offered eligible respondents additional incentives to encourage participation: a $100 base incentive for the PR; a $10 gift card for each of the three scheduled telephone calls by the PR; a $10 gift card for children aged 11-14; and a $20 gift card for household members older than 14.

Data

This study used data from both the FoodAPS public-use data file and the restricted data file. The measures evaluated in this study include mean FAH and FAFH expenditures and events per person per day, and the total weekly FAH and FAFH expenditures at the population level for those who are at least 11 years old, resulting in 11,552 individuals in our analysis. Those who are 10 or younger are not included in the analysis since they do not have any FAH expenditures and have only minimal FAFH expenditures. Our data set is constructed at the person-day level. Table 1 presents an illustration of the data set using hypothetical values on the variables of interest.

An important variable that we used in this study is the status variable, providing paradata on whether the PR confirmed the acquisition for each household member for a given day. For those person-days with a confirmed status, respondents reported at least one FAH or FAFH event (with valid expenditure amounts recorded in the data set) or they confirmed that they did not have any FAH / FAFH events that day (zero recorded expenditures for both FAH and FAFH events). For those person-days that are reported as being refusals or unconfirmed non-events, the expenditure amounts for both FAH and FAFH events are missing. There are also a handful of confirmed events with missing expenditure values. In evaluating the extent of expenditure underreporting, we aim to develop a method for imputing expenditures for those person-days

For Review Only

with missing expenditures. Of the 80,864 person-days in our sample (i.e., 11,552 × 7), 84.4% had a confirmed status with a reported or confirmed $0 expenditure, 1.2% had a confirmed status with a missing expenditure, 4.5% had a refusal status, and 9.9% had an unconfirmed status.

Table 1. A hypothetical illustration of the data set analyzed in this study, including predicted values.

Household ID	Respondent ID	Fielding day	Status	FAH dollars	FAH dollars including prediction	FAFH dollars	FAFH dollars including prediction
100001	1	1	confirmed	18.5	18.5	1.5	1.5
100001	1	2	confirmed	0	0	4	4
100001	1	3	confirmed	14	14	0	0
100001	1	4	confirmed	0	0	0	0
100001	1	5	confirmed	missing	8	missing	7
100001	1	6	refusal	missing	5	missing	3
100001	1	7	refusal	missing	9	missing	0
100001	2	1	confirmed	5	5	4.5	4.5
100001	2	2	confirmed	34	34	14	14
100001	2	3	confirmed	0	0	6.8	6.8
100001	2	4	confirmed	5	5	0	0
100001	2	5	refusal	missing	3	missing	3
100001	2	6	unconfirmed	missing	3	missing	2
100001	2	7	unconfirmed	missing	15	missing	5
100002	3	1	confirmed	34	34	9	9
100002	3	2	confirmed	0	0	14	14
100002	3	3	confirmed	0	0	24	24
100002	3	4	confirmed	0	0	3	3
100002	3	5	unconfirmed	missing	18	missing	3
100002	3	6	unconfirmed	missing	5	missing	5
100002	3	7	unconfirmed	missing	21	missing	11

The key covariates included in the analysis are presented in Table 2 in the next section. Both time-varying (i.e., day-varying) and time-invariant covariates are used in the models predicting FAH / FAFH events and expenditures. Most of these variables have been reported as predictive of the reporting of FAH / FAFH events in a prior study of FoodAPS data (Hu et al.

For Review Only

2017). Additional details on the covariates can be found in the supplementary materials (Appendices A and B).

Statistical Methodology

To evaluate the magnitude of underreporting in FAH (/ FAFH) expenditures, we aim to impute the missing expenditures at the *person-day* level, and then compare the *weighted* mean expenditures (recall that the FoodAPS weights adjust for unit nonresponse) between 1) only reported expenditures for days with confirmed status (which is what FoodAPS data users would be analyzing) and 2) reported *and* imputed expenditures. To fully account for the uncertainties in a single imputation of the missing values, we first select several hundred replicate bootstrap samples from the original FoodAPS data, following the resampling techniques outlined by Rao and Wu (1988), Girard (2009), Kolenikov (2010) and Valliant, Dever and Kreuter (2013) for complex samples. We then impute the missing expenditures using a Heckman selection model approach (Heckman 1979) for each bootstrap sample, and compare the expenditure estimates before and after imputation. Besides accounting for the uncertainties in a single imputation, another important advantage of this bootstrap approach is that it allows us to obtain an empirical standard error for the estimated difference between the two aforementioned expenditure estimates. We provide more detail below on each aspect of our methodological approach.

Bootstrap Resampling. Our bootstrap resampling technique can be summarized with the following steps. First, suppose that for each stratum h based on the FoodAPS sample design, there are n_h PSUs. For all strata, there are in total $n = \sum_{h=1}^{H} n_h$ PSUs. In each stratum, we draw a simple random sample *with replacement* of m_h PSUs from the initial n_h PSUs (available in the FoodAPS data set). For a given PSU i, m_{hi}^* denotes the number of times that this PSU is selected

For Review Only

from the stratum h, and $\sum_{h=1}^{n_h} m_{hi}^* = m_h$. For all sample PSUs, a replicate weight for each sampling unit k (household) within PSU i from stratum h can be created as below:

$$w_{hik}^* = w_{hik}\left(\left\{1 - \sqrt{\frac{m_h}{(n_h - 1)}}\right\} + \sqrt{\frac{m_h}{(n_h - 1)}}\frac{n_h}{m_h} m_{hi}^*\right) \qquad (1)$$

where w_{hik} is the original weight provided in the FoodAPS data file – i.e., the household weight adjusted for unit nonresponse for household k within PSU i from stratum h.

In our analysis, we use $m_h = n_h - 1$ (Rao and Wu 1988). We then repeat this step 500 times, resulting in a set of 500 replicate weights. Each re-sampled PSU in a given bootstrap sample includes all households, individuals, and person-days that were associated with that PSU in the original FoodAPS dataset. The resulting replicate weights for each individual are *constant* for each fielding day.

Heckman selection model. To predict the missing expenditures for person-days that were refusals or unconfirmed reports in each of the replicate bootstrap samples, we use a Heckman selection model, in which the estimates of the expenditures are conditioned on the selection process of whether a respondent has at least one FAH (/ FAFH) event that day. There are two equations in the Heckman selection model – a (substantive) regression model and a selection model. The regression equation models the relationship between the outcome, here the daily expenditure amounts, and a vector of covariates **X**:

$$FAH\ (/FAFH)\ expenditures = \mathbf{X}\beta + u_1 \qquad (2)$$

For Review Only

In (2), β is a vector of regression coefficients for the covariates, and u_1 is the residual of the regression model and is assumed to follow a normal distribution with mean 0 and standard deviation σ:

$$u_1 \sim N(0,\sigma^2)$$

The dependent variable in the expenditure model, food expenditures, is not always observed. Whether it will be censored or not depends on the selection model, as shown below:

$$FAH\ (/FAFH)\ Reporting\ Status = I\{\mathbf{Z}\gamma + u_2 > 0\} \qquad (3)$$

In (3), $\mathbf{Z}$ is the vector of covariates used to predict the dichotomous FAH or FAFH reporting status based on person-days with confirmed status (1 = yes, an event was reported; 0 = no event was reported), γ is a vector of regression coefficients, and u_2 is a random error, which is assumed to follow a standard normal distribution: $u_2 \sim N(0,1)$. $\mathbf{Z}\gamma + u_2$ can be interpreted as the propensity to have a FAH / FAFH event on a given day, and $FAH\ (/FAFH)$ $Reporting\ Status = 1$ if $\mathbf{Z}\gamma + u_2 > 0$. The dependent variable measuring expenditures is observed only when $FAH\ (/FAFH)\ Reporting\ Status = 1$.

The correlation between the two error terms, u_1 and u_2, is denoted as ρ: $corr(u_1, u_2) = \rho$. If ρ is not equal to zero, this would suggest that there is correlation between the two error terms, and thus using the Heckman selection model is appropriate (i.e., there is an association between selection and the substantive responses, potentially introducing bias in the substantive model). The parameter $\lambda = \sigma\rho$ is known as the coefficient on the inverse Mills ratio. A significant λ would suggest that there is sample selection bias and that the Heckman selection model should be used, instead of the ordinary least squares (OLS) – based regression model. When fitting the Heckman models, we accounted for the complex sampling features of the

For Review Only

FoodAPS, including the adjusted household-level weights for each bootstrap replicate. To account for the correlations between days within the same person, we specified the individual ID as the cluster variable when running the Heckman models.

The main independent variables for both equations include both individual- and household-level variables. In an effort to explain between-person and between-household variance in the outcomes (which were large before adding any covariates), we added a rich set of covariates in the model, including variables describing cumulative changes in relevant correlates across the days for each household member. By controlling for these cumulative measures in our models, we are explaining outcomes in a way that accounts for what everyone else in the household has done up to this point in time. Table 2 shows all the variables initially considered in the models, following previous literature presenting analyses of the FoodAPS reporting data (Hu et al. 2017; Ong et al. 2018). Based on initial assessments of the models using the full FoodAPS sample, we removed some insignificant covariates and covariates that caused model convergence issues. In Table 2, the "X" labels indicate the final sets of covariates selected for each model. Appendix A provides definitions for all the covariates.

For Review Only

Table 2. Descriptions of the variables analyzed.

	FAH model		FAFH model	
	Selection Model	Substantive Model	Selection Model	Substantive Model
Time-Varying Covariates				
Individual-level variables:				
• Cumulative number of FAH events	X	X	X	X
• Cumulative number of FAFH events	X	X	X	X
• Cumulative FAH expenditures	X	X	X	X
• Cumulative FAFH expenditures	X	X	X	X
• Cumulative reporting status counts	X	X	X	X
• Interaction between age group and weekend	X	--	X	--

Household-level variables:				
• Fielding day	X	--	X	--
• Interaction between start date and fielding day	X	--	X	--
• Days-since-issuance of SNAP benefits	X	X	X	X
• Weekend	X	X	X	X
Time-invariant Covariates				
Individual-level variables:				
• Body Mass Index (BMI)	X	X	X	
• Gender	X	X	X	X
• Age	X	X	X	X
• Individual income	X	X	X	X
• Education	X	X	X	X
• Marital status	X	X	X	X
• Race	X	X	X	X
• Relationship with PR	X	X	X	X
• Employment status	X	X	X	X
Household-level variables:				
• Time gap between the initial interview and the start date of the survey	X	X	X	
• Start date cohort	X	X	X	X
• Feedback on the completion of meals and snacks forms	X	X	X	
• Feedback on member participation	X	X	X	
• Feedback on ease of keeping track of food	X	X	X	
• Financial condition	X	X	X	X
• How often bills are reviewed for accuracy	X	X		
• Household size	X	X	X	X
• Total inbound calls	X	X	X	
• Total outbound calls	X	X	X	
• Summer indicator	X	X	X	X
• Interaction between age group and summer indicator	X	--	X	--
• Total initial interview time	X	X	X	X
• The total number of contacts before the diary survey	X	X	X	X

For Review Only

• Ever refused before diary survey data collection	X	X	X	X

Note. "X" indicate the final sets of covariates selected for each model; "--" indicates a variable was never considered in the substantive model.

Prediction. After fitting the selection model and generating estimates of the parameters defined above for each replicate bootstrap sample, and accounting for the complex sampling features of the FoodAPS when fitting each model (via the svy: heckman command in Stata), we applied this model to those person-days with missing expenditures and predicted the amount of money spent on FAH (/FAFH) events on those specific days for each person with missing expenditures. For confirmed "yes" person-days with missing expenditures, refusals, and unconfirmed reports, predictions of the missing expenditure values were calculated as below (all confirmed "no" observations had the expenditure amounts set to zero):

$$y_{it}{}^{*} = \Pr(\mathrm{y}_{it}\text{ observed})E(\mathrm{y}_{it}\,|\mathrm{y}_{it}\text{ observed}),$$

where $\Pr(\mathrm{y}_{it}\text{ observed})$ is the probability that person i will have a FAH (/FAFH) event on a given day t (based on the available covariates), and $E(\mathrm{y}_{it}\,|\mathrm{y}_{it}\text{ observed})$ is the expected expenditure given that this event is observed (again based on the available covariates). Note that the missing values for confirmed "yes" cases also had a predicted probability of an event used in this imputation, given the uncertainty in whether an event actually occurred in light of the missing value for the dollar amount. Carefully-specified imputation models will minimize the prediction uncertainty for any given item nonresponse rate, making the careful selection of covariates in these models (per Table 2) an important part of this process.

For Review Only

In the next step, we calculated both 1) the mean FAH / FAFH expenditures per person per day based on the reported data for confirmed person-days only, and 2) the mean FAH / FAFH expenditures when including predictions for the missing person-days, using the bootstrap replicate weights w^*_{hik} instead of w_{hik} in each case. For example, the mean expenditure based on *only non-missing confirmed days*, say $\hat{\theta}_{obs}$, was calculated using the replicate weights w^*_{hik} for each replicate sample.

$$\hat{\theta}_{obs} = \frac{\sum_{h=1}^{h=H}\sum_{i=1}^{i=I_h}\sum_{k=1}^{k=K_{hi}}\sum_{p=1}^{p=P_{hik}}\sum_{t=1}^{t=T_{hikp}} w^*_{hik} y_{hikpt}}{\sum_{h=1}^{h=H}\sum_{i=1}^{i=I_h}\sum_{k=1}^{k=K_{hi}}\sum_{p=1}^{p=P_{hik}}\sum_{t=1}^{t=T_{hikp}} w^*_{hik}}$$

where H is the total number of strata, I_h indicates the total number of PSUs within stratum h, K_{hi} indicates the total number of households within PSU i and stratum h, P_{hik} indicates the total number of household members within household k in PSU i and stratum h, and T_{hikt} indicates the total number of days where individual p (in household k in PSU i and stratum h) had non-missing expenditures. We repeated this process 500 times, and thus obtained 500 bootstrap sample estimates, $\hat{\theta}_{obs(1)}, \hat{\theta}_{obs(2)}, \dots, \hat{\theta}_{obs(500)}$, for each mean expenditure estimated.

Similarly, the mean for the final estimate when *including* imputed expenditures, say $\hat{\theta}_{imp}$, was also calculated using the replicate weights w^*_{hik} for each replicate sample.

$$\hat{\theta}_{imp} = \frac{\sum_{h=1}^{h=H}\sum_{i=1}^{i=I_h}\sum_{k=1}^{k=K_{hi}}\sum_{p=1}^{p=P_{hik}}\sum_{t=1}^{t=7} w^*_{hik} \hat{y}_{hikpt}}{\sum_{h=1}^{h=H}\sum_{i=1}^{i=I_h}\sum_{k=1}^{k=K_{hi}}\sum_{p=1}^{p=P_{hik}}\sum_{t=1}^{t=7} w^*_{hik}}$$

where

$$\hat{y}_{hikpt} = \begin{cases} observed\ values, & if\ observed \\ 0, & if\ confirmed\ no\ status \\ predicted\ values, & if\ missing \end{cases}$$

We repeated this process for each of the 500 replicate samples and thus obtained 500 bootstrap sample estimates, $\hat{\theta}_{imp(1)}$, $\hat{\theta}_{imp(2)}$,… , $\hat{\theta}_{imp(500)}$, for each mean expenditure estimated, which includes imputed expenditures for missing person-days.

We then computed the difference between the two mean expenditure estimates, $\hat{\theta}_{imp} - \hat{\theta}_{obs}$, for each of the 500 bootstrap samples. We evaluated the mean of the differences and its empirical standard error based on the 500 bootstrap samples. Note that this imputation method captures the prediction uncertainty associated with the imputation models in the standard errors of the resulting estimates. Following the same approach, we estimated the differences in *total* weekly expenditures based on reported expenditures from confirmed days only and then also including predictions for missing days. The statistical code for this methodology is illustrated in Appendix E. When presenting group-level estimates, we combine the two non-SNAP household groups with incomes below 185% poverty, yielding a total of three subgroups defined by SNAP participation status and poverty level.

We also estimated the extent of underreporting for the number of FAH / FAFH events using a similar approach as described above for expenditures. Given space constraints, we include detailed information on evaluations of FAH / FAFH events in Appendix D.

Global day-adjustment method. To evaluate the results of our imputation model, we employed an alternative adjustment approach which aimed to account for increased underreporting in later fielding days (referred to as the *global day-adjustment method*), as initially proposed by Ren, Yan and Krenzke (2017).

We computed the adjustment factor at the national level. Specifically, for the first two days, the adjustment factors are both 1. For Day 3, the factor is the average of the national total

https://mc.manuscriptcentral.com/jssam

Day 1 and Day 2 expenditures to the national total Day 3 expenditure. Similarly, the adjustment factor for Day 4 is the average of Day 2 and Day 3 expenditures to the total Day 4 expenditures, and so on throughout the week. We adjusted the initial household weights with the day-level adjustment factor, and estimated expenditures accounting for these day-level factors. Suppose w_i is the weight (at household level) for respondent i and δ_j is the factor for Day j. The mean expenditure $\overline{\mathrm{y}}$ is calculated as below:

$$\overline{\mathrm{y}} = \left(\sum_i \sum_j w_i \delta_j y_{ij}\right) / \left(\sum_i \sum_j w_i \delta_j\right)$$

where y_{ij} is reported expenditure for respondent i for Day j. We then compared this adjustment approach with our imputation results.

Results

The Results section contains two parts. The first describes estimates from the Heckman selection models fitted to the original FoodAPS sample (see also Appendix C). The second describes the results of our bootstrap / imputation procedure, where we compared the differences between reported and predicted mean and total weekly expenditures (accounting for non-response across the fielding days).

Heckman Selection Model Results

Specific estimates from the Heckman selection models predicting FAH and FAFH expenditures at the person-day level can be found in Appendix C. We remind readers that the

For Review Only

models discussed below were fitted to *each bootstrap replicate sample* in our imputation procedure; this section provides an overview of important findings based on the original FoodAPS sample only.

For both FAH and FAFH models, we found that ρ is significantly different from zero (FAH: $\rho = -0.11$, $p < 0.001$; FAFH: $\rho = -0.05$, $p = 0.013$), indicating that we can reject the null hypothesis that there is no correlation between the error terms of the selection equation and substantive regression equation. We also found that λ was significantly different from zero for both models, indicating that there is differential sample selection; that is, those who report FAH / FAFH events are systematically different from those who do not. This supports our use of Heckman selection models and indicates that a standard regression model fitted using ordinary least squares (OLS) could provide biased estimates of relationships and predictions.

For Review Only

Prediction and Imputation Results

The FAH (/FAFH) Heckman models accounting for the complex sample design of the FoodAPS successfully converged for 472 (/479) of the 500 replicate bootstrap samples within 100 iterations. The results presented below are based on the predictions computed using the 472 (/ 479) converged models.

Table 3 include estimates based on our method and the global day-adjustment method. For our method, the table present 1) the weighted FAH mean expenditures per person per day according to the available FoodAPS weights and the original FoodAPS respondents (in bold), which is what FoodAPS users would be analyzing; 2) the mean of all of the weighted FAH mean estimates coming from the bootstrap replicates for respondents only ($\hat{\theta}_{obs}$), 3) the mean of all of the weighted estimates from the bootstrap replicates *after* imputation of missing values ($\hat{\theta}_{imp}$), and 4) the difference between $\hat{\theta}_{obs}$and $\hat{\theta}_{imp}$, for both the overall sample and for specific SNAP and income subgroups respectively. The results indicate that, if one ignores the missing data and analyzes expenditures based only on FAH reports with a confirmed status, one would over-estimate the mean FAH expenditure per person per day by \$0.15, or about 1.8 percent ($p < 0.001$).We also examined the differences between reported and predicted mean expenditures for three groups of households defined by SNAP participation status and poverty level – SNAP, non-SNAP with income at or below 185% of the federal poverty level, and non-SNAP with income above 185% of the federal poverty level. As shown in Table 3, if one ignores the missing data and analyzes expenditures based only on FAH reports with a confirmed status, one would over-estimate the mean FAH expenditure per person per day for all three groups.

Similarly, as shown in Table 4, one would over-estimate the mean FAFH expenditures per person per day by \$0.08, or about 1.8 percent ($p <0.001$). The positive differences imply that events not being reported are, on average, less costly than events that are reported. For FAFH mean expenditures across the three SNAP and income groups, one would over-estimate the mean FAFH expenditures per person per day for SNAP and non-SNAP with incomes above 185% of the federal poverty level but not for non-SNAP with income at or below 185% of the federal poverty level.

Table 3. Mean FAH expenditure differences and bootstrap imputation results by target group.

	Bootstrap samples		**Global day-adjustment method**	
	Mean (\$)	se	Mean (\$)	se

For Review Only

Overall (Weighted Respondent Mean = $8.26, SE = 0.22)				
FAH dollar amount each person spent per day	8.26	0.22		
FAH dollar amount after imputation	8.11	0.20	8.21	0.22
Difference between the two means	0.15***	0.04		
SNAP (Weighted Respondent Mean = $7.30, SE = 0.32)				
FAH dollar amount each person spent per day	7.33	0.33		
FAH dollar amount after imputation	7.21	0.29	7.24	0.33
Difference between the two means#	0.11**	0.09		
Non-SNAP below 185% (Weighted Respondent Mean = $6.46, SE = 0.29)				
FAH dollar amount each person spent per day	6.47	0.28		
FAH dollar amount after imputation	6.51	0.26	6.43	0.28
Difference between the two means	-0.04	0.08		
Non-SNAP 185 or higher (Weighted Respondent Mean = $8.83, SE = 0.28)				
FAH dollar amount each person spent per day	8.83	0.28		
FAH dollar amount after imputation	8.66	0.26	8.78	0.28
Difference between the two means	0.17***	0.05		

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.0001$ # note that due to rounding, the difference between the two means may differ slightly from $7.33-$7.21.

Table 4. Mean FAFH expenditure differences and bootstrap results by target group.

	Bootstrap samples		**Global day-adjustment method**	
	Mean ($)	se	Mean ($)	se
Overall (Weighted Respondent Mean = $4.54, SE = 0.14)				
FAFH dollar amount each person spent per day	4.53	0.16		
FAFH dollar amount after imputation	4.45	0.16	4.52	0.14
Difference between the two means	0.08***	0.03		
SNAP (Weighted Respondent Mean = $2.55, SE = 0.16)				

FAFH dollar amount each person spent per day	2.56	0.16		
FAFH dollar amount after imputation	2.50	0.16	2.53	0.159
Difference between the two means	0.06**	0.03		
Non-SNAP below 185% (Weighted Respondent Mean = $3.07, SE = 0.21)				
FAFH dollar amount each person spent per day	3.05	0.21		
FAFH dollar amount after imputation	3.11	0.19	3.06	0.208
Difference between the two means	-0.06	0.04		
Non-SNAP 185 or higher (Weighted Respondent Mean = $5.24, SE = 0.19)				
FAFH dollar amount each person spent per day	5.24	0.21		
FAFH dollar amount after imputation	5.17	0.20	5.21	0.189
Difference between the two means	0.07***	0.03		

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.0001$

Weighted estimates based on the *global day-adjustment method* are presented in Tables 3 and 4 in the final two columns. In general, adjustments using either method move estimates in the same direction. In comparison, our method adjusts the estimates to a larger extent. Taking the mean FAH expenditures of all households as an example (the top section in Table 3), the original estimate of FAH dollars per person per day is $8.26. After imputation, the estimate decreases to $8.11, suggesting that, without imputation, the original estimate would be an overestimate. Using the alternative global day-adjustment method, the adjusted estimate is $8.21, which also suggests that the original estimate is an overestimate but adjusts the original estimate to a lesser extent. This indicates that our method is picking up more information from the microdata and the model-based approach. Similar findings were found for the imputation of number of events based on a negative binomial regression model (see Appendix D).

Table 5 presents results based on total weekly FAH expenditures. For overall FAH expenditures regardless of SNAP and income group, if one ignores the missing data and analyzes

For Review Only

only the reported FAH expenditures, one would under-estimate total weekly FAH expenditures by 9.4%. For FAH expenditures across the three SNAP and income groups, the results suggest that ignoring the missing data issue leads to under-estimation of total weekly expenditures for all three SNAP and income groups. The levels of under-estimation for FAH total expenditures are greater for the two low-income household groups (15.6% and 14.1%) than for the high-income group (7.5%).

Similarly, as shown in Table 6 for total weekly FAFH expenditures, if one ignores the missing data and analyzes only the reported FAFH expenditures, one would under-estimate FAFH total expenditures by 9.4%. For FAFH total weekly expenditures across the three SNAP and income groups, one would under-estimate FAFH total expenditures for all three SNAP and income groups. Similar to the FAH results, the levels of under-estimation for FAFH total expenditures are greater for the two low-income household groups (14.9% and 15.3%) than for the high-income group (8.0%). The global day-adjustment method provides similar shifts in Tables 5 and 6, although not as large as our method in magnitude.

Recall from our earlier description of the FoodAPS data that about 15.6 percent of all person-days were either refusals, unconfirmeds, or confirmeds with missing expenditure data. This 15.6 percent is greater than the overall underestimation rates for FAH and FAFH since it is roughly equal to the group-specific rates for low-income households while greater than for higher-income households. The disparities of the results across SNAP and income groups indicate the importance of taking group membership into account when imputing missing values for expenditure data (which was done in this case study). It is also a harbinger of how the final household survey weights in the FoodAPS data, which do account for group membership, are not successful in fully correcting for missing data at the person-day level.

For Review Only

Table 5. Weighted total weekly FAH expenditures and differences (in billions) and bootstrap imputation results, overall and by target group.

	Bootstrap samples		**Global day-adjustment method**	
	Total ($)	se	Total ($)	se
Overall (Weighted population total = 12.97, SE = 0.84)				
Total FAH dollar amount	12.97	0.83		
Total FAH dollar amount after imputation	14.31	0.84	14.00	0.90
Difference between the two total amounts	-1.34 ***	0.10		
% change	-9.39%			
SNAP (Weighted population total = 1.58, SE = 0.11)				
Total FAH dollar amount	1.58	0.11		
Total FAH dollar amount after imputation	1.87	0.13	1.70	0.11
Difference between the two total amounts	-0.29***	0.04		
% change	-15.56%			
Non-SNAP below 185% (Weighted population total = 1.55, SE = 0.12)				
Total FAH dollar amount	1.56	0.12		
Total FAH dollar amount after imputation	1.81	0.13	1.68	0.13
Difference between the two total amounts	-0.26***	0.03		
% change	-14.14%			
Non-SNAP 185 or higher (Weighted population total = 9.84, SE = 0.79)				
Total FAH dollar amount	9.83	0.78		
Total FAH dollar amount after imputation	10.63	0.79	10.62	0.85
Difference between the two total amounts	-0.80***	0.08		
% change	-7.50%			

* p < 0.05; ** p < 0.01; *** p < 0.0001

For Review Only

Table 6. Weighted total weekly FAFH expenditures and differences (in billions) and bootstrap results, overall and by target group.

	Bootstrap samples		**Global day-adjustment method**	
	Total ($)	se	Total ($)	se
Overall (Weighted population total = 7.12, SE = 0.39)				
Total FAFH dollar amount	7.14	0.41	7.54	0.41
Total FAFH dollar amount after imputation	7.88	0.40		
Difference between the two total amounts	-0.74***	0.07		
% change	-9.40%			
SNAP (Weighted population total = 0.55, SE = 0.05)				
Total FAFH dollar amount	0.55	0.04	0.58	0.05
Total FAFH dollar amount after imputation	0.65	0.05		
Difference between the two total amounts	-0.10***	0.02		
% change	-14.90%			
Non-SNAP below 185% (Weighted population total = 0.74, SE = 0.06)				
Total FAFH dollar amount	0.73	0.06	0.78	0.07
Total FAFH dollar amount after imputation	0.87	0.07		
Difference between the two total amounts	-0.13***	0.02		
% change	-15.26%			
Non-SNAP 185 or higher (Weighted population total = 5.83, SE = 0.40)				
Total FAFH dollar amount	5.85	0.42	6.18	0.43
Total FAFH dollar amount after imputation	6.37	0.42		
Difference between the two total amounts	-0.51***	0.06		
% change	-8.04%			

* p < 0.05; ** p < 0.01; *** p < 0.0001

For Review Only

Discussion

In this paper, we used a methodology combining bootstrap replicate sampling for complex samples and imputation based on a Heckman selection model to predict FAH and FAFH expenditures for person-days in the FoodAPS data set with refusals, unconfirmed status or otherwise missing expenditures. We then estimated the item nonresponse error associated with the mean FAH and FAFH expenditures per person per day, and the total FAH and FAFH expenditures at the population level, by comparing weighted estimates according to 1) only reported expenditures and 2) both reported expenditures and predictions for missing values. Results indicate that ignoring the missing data leads to over-estimation of the mean expenditures per person per day by $0.15 for FAH and $0.08 for FAFH, and under-estimations of the total weekly expenditures by 9.4% for both FAH and FAFH.

This study provides several key contributions. First, most previous literature studying the magnitude of underreporting in expenditure diary surveys compare the survey estimates to external benchmarks (e.g., from another similar survey) (e.g., Fricker et al. 2015; Passero et al. 2013). Our research presents a useful imputation method to quantify the size and significance of the over- and under- reporting. Comparing to the global day-adjustment method, our imputation method is picking up more information from the microdata and the model-based approach. An additional benefit of our method is that it allows analysts to employ standard bootstrap resampling and imputation modeling for handling item-missing data. The blending of these standard and widely-used approaches in this methodology facilitates this type of approach for analysts, whereas the global day-adjustment method would require custom coding depending on the specific design of the diary survey. Second, one might expect that the use of household survey weights (including unit nonresponse adjustments) would be sufficient to correct for

For Review Only

nonresponse error in diary surveys like the FoodAPS. This was not the case in this study. Our results indicate that the weighted estimates of the total expenditures prior to imputation are clearly not compensating for the biases in estimation of the totals due to failures of respondents to report expenditures on particular days. This is because the standard household weighting procedures for nonresponse adjustment are performed at the *household* level. In this type of diary survey across seven days, using household-level weights does not correct for refusals and unconfirmeds that are occurring at the person-day level among those who did participate but then fail to report expenditures on particular days. We demonstrate that sound imputation of these missing values is clearly an important step in this process. Third, this study has important implications for future waves of FoodAPS and future designs of similar expenditure diary surveys. For example, given the higher level of underreporting among the low-income subgroups, future iterations of FoodAPS could provide some tailored interventions (e.g., increased incentives and / or modified introductions to address their unique circumstances) to the low-income households following the principles of responsive survey design (Groves and Heeringa 2006).

It is also important to note that this study is limited in several ways. First, we have only addressed a couple possible sources of underreporting (primarily due to item non-response error occurring at the day level) in this study. Without knowing the true values of expenditures, we are not able to investigate possible non-reporting of events on days when the member reported at least one event, nor are we able to examine incomplete reporting of the expenditures of acquired food items conditional on an event being reported. Future studies could consider general methodologies to simultaneously tackle misreporting (where someone responds, but with a value not equal to the true value) *and* item nonresponse occurring at the day level, in the spirit of

For Review Only

Ghosh-Dastidar and Schafer (2003). Second, one of the assumptions we made in the imputation approach is that the Heckman models fitted based on the respondents are applicable to those who refused or were unconfirmed for some days of the week. Future studies could further test this assumption using administrative data [e.g., using the Food and Nutrition Service's ALERT (Anti-Fraud Locator Using Electronic Benefit Transfer Retailer Transactions) data file (which monitors SNAP recipient transactions) when analyzing SNAP subgroups and purchases only], if available, to see whether similar Heckman models would emerge for both groups. Third, this paper did not impute where the implied event occurred (e.g., supermarket, convenience store, farmers' market), which affects both expenditure amounts and the nutritional value of foods acquired (Mancino et al 2018; Todd and Scharadin 2016). Future research could replicate our analyses by store type, and could improve estimates of underreporting and provide insight into the nutritional characteristics of acquired, but unreported, food.

Our research identifies several important directions for future research. First, future studies could apply our approach to other expenditure diary surveys to determine whether similar findings on the level of underreporting among different income groups may appear. Second, future studies could explore other imputation / prediction methods. For example, in a longitudinal study, the information from previous waves may be used in the imputation process to better predict the missing values across fielding days. Third, to address the item nonresponse error occurring at the day level in diary surveys, besides the imputation method as introduced in this paper, future studies could explore the construction of day-specific weights accounting for item nonresponse. Fourth, prior to the diary component of the survey, respondents could be asked about how often they shop for food or dine out, on average, and how much they usually spend. The next round of FoodAPS plans to capture such contextual information to provide more

For Review Only

information about how representative the data collection week is for sampled households, but the information also may be used to identify households or individuals which report fewer acquisitions than would be expected from this auxiliary information.

In conclusion, our study suggests that the household-level weights typically computed for diary surveys, which generally account for unit nonresponse among households, may not be entirely sufficient for addressing the nonresponse occurring on each day in an diary survey such as FoodAPS, and appropriate imputation methods will be important for estimating the size of the underreporting. Future implementations of this or other expenditure diary surveys can apply this method and use the findings of this study to improve the designs of these future surveys.

For Review Only

REFERENCE

Alba, J. W., and Chattopadhyay, A. (1985), “Effects of Context and Part-Category Cues on Recall of Competing Brands,” *Journal of Marketing Research*, 22, 340.

Bureau of Labor Statistics (2010), “Consumer Expenditure Survey compared with the American Community Survey.”, available at https://www.bls.gov/cex/twoyear/200607/csxacs.pdf.

CDC (2015), “About Child & Teen BMI”, available at http://www.cdc.gov/healthyweight/assessing/bmi/childrens_bmi/about_childrens_bmi.html/.

Clay, M., Ver Ploeg, M., Coleman-Jensen, A., Elitzak, H., Gregory, C., Levin, D., Newman, C., Rabbitt, M. P., and others (2016), “Comparing National Household Food Acquisition and Purchase Survey (FoodAPS) Data With Other National Food Surveys’ Data.”, EIB-157, U.S. Department of Agriculture, Economic Research Service, July 2016, available at https://www.ers.usda.gov/webdocs/publications/79893/eib-157.pdf.

Creech, B. J. (2015), “A Comparison of Consumer Expenditure Survey and Panel Study of Income Dynamics Expenditure Data: 1999–2011.”, *Bureau of Labor Statistics Discussion Paper*, available at https://www.bls.gov/cex/ce_psid_199911.pdf.

FoodAPS (2016), “National Household Food Acquisition and Purchase Survey: User’s Guide to Survey Design, Data Collection, and Overview of Datasets.”, available at: https://www.ers.usda.gov/media/8804/0_foodaps-user-guide-puf.pdf.

Fricker, S., Kopp, B., Tan, L., and Tourangeau, R. (2015), “A review of measurement error assessment in a u.s. household consumer expenditure survey,” *Journal of Survey Statistics and Methodology*, 3, 67–88.

Ghosh-Dastidar, B., & Schafer, J. L. (2003). Multiple Edit/Multiple Imputation for Multivariate Continuous Data. *Journal of the American Statistical Association*, 98(464), 807–817.

Girard, C. (2009), “The Rao-Wu rescaling bootstrap: From theory to practice,” In *Proceedings of the Federal Committee on Statistical Methodology Research Conference* (pp. 2-4), available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.178.7134&rep=rep1&type=pdf

Groves, R. M., Bowers, A., Kreuter, F., Casas-Cordero, C., Miller, P., and Excellence, C. for R. (2009), “An Independent Analysis of the Nielsen Meter and Diary Nonresponse Bias Studies.” *A Report to the Council for Research Excellence*, available at: http://www.researchexcellence.com/files/pdf/2015-02/id157_cre_report_diary_03_09_09_.pdf.

Groves, R. M., and Heeringa, S. G. (2006), “Responsive design for household surveys: tools for actively controlling survey errors and costs,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169, 439–457.

Heckman, J. J. (1979), *Statistical models for discrete panel data*, Department of Economics and Graduate School of Business, University of Chicago Chicago, IL.

Heeringa, S. G., West, B. T., & Berglund, P. A. (2017). *Applied survey data analysis*. Chapman

For Review Only

and Hall/CRC.

Hu, M., Gremel, G. W., Kirlin, J. A., and West, B. T. (2017), "Nonresponse and Underreporting Errors Increase over the Data Collection Week Based on Paradata from the National Household Food Acquisition and Purchase Survey," *Journal of Nutrition*, 147(5), 964-975.

Kolenikov, S. (2010). Resampling variance estimation for complex survey data. *Stata Journal*, 10(2), 165-199.

Krizek, K. J., Handy, S. L., and Forsyth, A. (2009), "Explaining changes in walking and bicycling behavior: Challenges for transportation research," *Environment and Planning B: Planning and Design*, 36, 725–740.

Li, G., Schoeni, R. F., Danziger, S., and Charles, K. K. (2010), "New Expenditure Data in the PSID: Comparisons with the CE," *Monthly Labor Review*, JSTOR, 133, 29–39.

Maitland, A., and Li, L. (2016), "Review of the Completeness and Accuracy of FoodAPS 2012 Data.", availale at: https://www.ers.usda.gov/media/9069/completenessaccuracy.pdf.

Mancino, Lisa, Joanne Guthrie, Michele Ver Ploeg, and Biing-Hwan Lin (2018), "Nutritional Quality of Foods Acquired by Americans: Findings From USDA's National Household Food Acquisition and Purchase Survey.", availale at: https://www.ers.usda.gov/webdocs/publications/87531/eib-188.pdf?v=43151.

Mathiowetz, N., Olson, K., and Kennedy, C. (2011), "Redesign Options for the Consumer Expenditure Survey.", available at: https://www.bls.gov/cex/redwrkshp_pap_abtsrbirecommend.pdf.

Ong, A. R., Hu, M., West, B. T., and Kirlin, J. A. (2018), "Interviewer effects in food acquisition surveys," *Public Health Nutrition*, 1–13.

Todd, J. E., & Scharadin, B. (2016). "Where households get food in a typical Week: Findings from USDA's FoodAPS (No. 242450)". *United States Department of Agriculture, Economic Research Service,* available at: https://www.ers.usda.gov/webdocs/publications/80542/eib-156.pdf?v=42663.

WHO (2017), "BMI classification." Available at: http://apps.who.int/bmi/index.jsp?introPage=intro_3.html.

Passero, W., Garner, T. I., and McCully, C. (2013), "*Understanding the Relationship: CE Survey and PCE*.", available at: https://www.bls.gov/osmr/pdf/ec130020.pdf.

Rao, J., and Wu, C. (1988), "'Resampling Inference With Complex Survey Data': TM," *Journal of the American Statistical Association*, 83, 231.

Ren, W., Yan, T., and Krenzke, T. (2017), "Tasks 16 and 21: ADCM Recall Interview Analysis" *Westat*. Available upon request Xingyou Zhang (xingyou.zhang@ers.usda.gov).

Valliant, R., Dever, J. A., and Kreuter, F. (2013), *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

For Review Only

Online Supplementary Materials

Appendix A. Description of the covariates.

Fielding day. Given that the nonresponse pattern varies across the fielding days (Hu et al. 2017), we constructed fielding day as the actual numeric day of the seven-day period, (coded 1, 2, …, 7), centered at the first day (i.e., the actual codes are 0, 1, 2, …, 6).

Days since issuance of SNAP benefits. We hypothesize that participation in SNAP will be correlated with food spending, but that the relationship may vary throughout the monthly issuance cycle. To capture these effects, we constructed a days-since-issuance (DSI) SNAP status indicator based on a SNAP status variable and DSI information, which is the number of days between the last time SNAP benefits were received by the household and the fielding day. This time-varying indicator includes five categories: SNAP household with DSI≤10, 11≤DSI≤20, DSI≥21, SNAP household but DSI unknown, and non-SNAP household.

Cumulative change variables. These covariates describe cumulative changes in relevant correlates across the days for each household member. Specifically, we computed the cumulative number of FAH and FAFH events, the cumulative FAH and FAFH expenditures, and the cumulative reporting status counts for each fielding day and each household member.

Take the cumulative expenditure as an example. For a one-person household, the cumulative measures are set to be 0 for the first fielding day. For households with at least two members, for each household member, the average expenditure for the other members (excluding the person whose value is to be calculated) is computed for the first fielding day. Similarly, for subsequent fielding days, we calculate the average cumulative expenditure based on other members' expenditures until that day and all prior days for the specific member whose value is

For Review Only

to be calculated (excluding the expenditure for the current to-be-calculated fielding day of the specific member). For example, in a household with two members – A and B, the cumulative expenditure variable for A on Day Two is calculated as $\frac{(total\ household\ expenditures\ on\ Day\ 1\ +\ B's\ expenditure\ on\ Day\ 2)}{(number\ of\ fielding\ days * number\ of\ household\ member\ -\ 1)}$. The denominator is simply the number of person-day expenditures which have been added in the numerator, which is 3 in this example. By controlling for these new cumulative measures in our models, we are explaining expenditure outcomes in a way that accounts for what everyone else in the household has done up to this point in time.

Body Mass Index (BMI). BMI was calculated for each respondent based on reported heights and weights collected during the final interview. It was categorized into four categories: not overweight, overweight, obese, and not applicable / missing / error. The BMI categories for adults were based on WHO guidelines: BMI<25 kg/m^2 for not overweight, 25≤BMI<30 kg/m^2 for overweight and BMI≥30kg/m^2 for obese (WHO 2017). BMI categorizations for children and teens are based on CDC age- and sex-specific growth charts (CDC 2015). To maximize the number of respondents in our model, we have treated missing as a category.

Time gap between the initial interview and the start date of the survey. Although each household was instructed to start the diary survey right after the first face-to-face interview, when the PRs were provided with detailed instructions, this did not always occur. This time-invariant, household-level variable is constructed based on the time difference (measured in days) between the initial interview and the start date of the diary week.

Start data cohort. As noted by Hu et al. (2017), the start day of the survey is associated with the nonresponse pattern. We therefore constructed start date based on the actual day of the week (e.g., Monday) each household started the survey. This classifies households into seven start date cohorts (Sunday to Saturday).

Feedback variables. As noted by Hu et al. (2017), households' feedback on the survey process is associated with respondents' propensity to have a FAH / FAFH event on a given day. We therefore included three feedback variables in our analysis – feedback on the completion frequency of meals and snacks forms, feedback on the ease to get other household member to participate and feedback on the ease of keeping track of food for the PR. These variables were collected during a mail-in feedback survey left with the PR after the final interview. To maximize the number of respondents in our model, we have treated missing as a category.

Financial management variables. Household-level financial management variables were obtained during the final interview and include questions on how often bills are reviewed for accuracy, how often bills are paid on time and whether a household has not paid rent/mortgage or utility bills on time in the previous six months. To test whether these household-level indicators of possible financial distress or management are correlated with food spending by household members, our models include both household "financial condition" and how often bills are reviewed for accuracy. The former indicator was measured using a five-point ordinal scale, ranging from "very comfortable and secure" to "in over your head". The latter indicator used a five-point Likert scale ranging from "Never" to "Always".

Call history paradata. Paradata of this type were collected during the telephone interviews and include the number of inbound calls (where the PR calls the survey's telephone center to report acquisition information) and outbound calls (where the telephone center calls the PR) for the entire week. Outbound calls were made by the telephone center if a scheduled inbound call was not received the night before. As the willingness or ability of the PR to initiate

For Review Only

the scheduled call may be related to expenditure patterns or respondents' propensities to have a reported FAH / FAFH event, we include both the counts of inbound and outbound calls as two separate covariates.

Additional covariates. Additional covariates were obtained from the interviews and during the survey process. For each member, we include gender, age in years (categorized as 11-15, 16-20, 21-30, 31-55, and 56+), monthly gross income (log transformed), highest education level (high school grad, college grad, other), marital status (married, widowed, divorced, separated, never married), race (African American, Caucasian, Other, Multiple), and relationship to the PR (self, partner or child, other relative, non-relative). We include an indicator for whether the household member is employed. Household size is included, as is an indicator for weekend days. Finally, a "summer" indicator (i.e., fielding week started during months of June, July or August) is included to capture possible seasonal differences in food acquisition patterns (e.g., when children are generally not in school). See Table 2 in the manuscript for details.

For Review Only

Appendix B. Descriptive statistics for the sample.

Supplemental Table 1. Descriptive statistics for the sample.

	n	Weighted Mean or %	SE
Time-varying Covariates (n = 80,864 person-days)			
Days-since-issuance of SNAP benefits	80,836		
Non-SNAP household		85.2	0.98
10 or less days since last SNAP issuance		5.9	0.46
11-20 days since last SNAP issuance		4.6	0.29
21-31 days since last SNAP issuance		4.0	0.39
32+ days since last SNAP issuance		0.2	0.08
Weekend	80,864	28.6	<0.01
Cumulative number of FAH events	80,864	0.2	0.01
Cumulative number of FAFH events	80,864	0.6	0.01
Cumulative FAH expenditures	80,864	8.2	0.31
Cumulative FAFH expenditures	80,864	4.3	0.15
Cumulative reporting status counts	80,864	0.9	0.01
Time-invariant Covariates			
Individual level (n = 11,552 individuals)			
BMI	11,552		
Not overweight (<24.9)		39.0	1.01
Overweight (25.0–29.9)		32.0	0.67
Obese (>30.0)		27.5	1.06
NA, missing, or error		1.6	0.28
Gender	11,552		
Male		47.9	0.54
Female		52.1	0.54
Age	11,552		
11-15		8.3	0.52
16-20		7.2	0.46
21-30		15.5	0.83
31-55		39.4	0.87
>55		29.7	1.18
Individual Income (log-transformed)	11,479	6.0	0.07
Education	11,499		
Below high school		21.5	1.12
High school grads, below college		40.6	1.43

For Review Only

College grads and above		37.9	1.92
Marital Status	11,539		
Married		44.3	1.28
Widowed		5.7	0.34
Divorced		13.0	0.94
Separated		2.3	0.24
Never married		34.7	1.09
Race	11,537		
White		74.4	2.13
Black/African American		12.8	1.84
Other Races		11.1	1.33
Multiple Races		1.6	0.17
Relationship with PR	11,551		
Primary respondent		48.1	0.85
Partner or children		41.9	0.76
Other relatives		7.4	0.65
Non-relatives		2.7	0.26
Employment status	11,479	0.6	0.91
Household level (n = 4,822 households)			
Time gap between the initial interview and the start date of the survey	4,822	1.1	0.02
Start date cohort	4,822		
Sunday		11.8	0.98
Monday		6.9	0.46
Tuesday		17.7	0.75
Wednesday		18.6	0.91
Thursday		16.6	1.07
Friday		16.2	0.82
Saturday		12.3	0.86
Feedback on the completion of meals and snacks forms	4,822		
Every day		72.1	1.51
>1, not every day		15.6	1.16
Once before end		1.0	0.17
Once at end		3.8	0.50
Did not complete		2.7	0.41
Missing		4.8	0.76
Feedback on member participation	4,822		
Very easy/ Easy		59.9	1.12
Neutral		12.7	0.84

For Review Only

Very difficult / Difficult		6.5	0.72
Not applicable / Missing		20.9	1.22
Feedback on ease of keeping track of food	4,822		
Very easy/ Easy		80.6	0.94
Neutral		11.9	0.81
Very difficult / Difficult		3.9	0.52
Missing		3.6	0.78
Financial condition	4,815		
Very comfortable and secure		24.6	1.58
Without much difficulty		33.7	1.25
Occasionally difficulty		24.7	1.20
Tough to make ends meet		13.8	0.76
In over your head		3.3	0.35
How often bills are reviewed for accuracy	4,822		
Never		7.4	0.67
Rarely		5.8	0.51
Sometimes		15.5	0.74
Usually		17.0	0.86
Always		53.6	1.24
Not applicable		0.7	0.14
Household size	4,822	2.4	0.04
Total inbound calls	4,803	3.0	0.03
Total outbound calls	4,803	1.4	0.04
Summer indicator	4,822	52.0	1.64
Total initial interview time	4,822		
0-4.2 min		19.0	1.42
4.2 min - 6.3min		26.4	1.24
>6.3 min		54.6	2.14
The total number of contacts before the diary survey	4,822		
0-3		26.6	2.16
4-5		32.0	1.64
5-7		18.7	0.90
>7		22.7	1.51
Ever refused before diary survey data collection	4,822	2.4	0.47

For Review Only

Appendix C. Heckman Selection Model Results: Details

FAH (Supplemental Table 2)

First, the selection model shows that BMI, relationship to the PR, feedback on ease of keeping track of food, days-since-issuance of SNAP status, inbound calls, household size and all cumulative variables except cumulative FAFH expenditures have significant relationships with the probability that a household member has a FAH event for a given day. Comparing with those who are not overweight, those who are obese are less likely to have any FAH event on a particular day. Household members other than the PRs are less likely to report any FAH events for a given day. Households in which the PRs reported the ease of keeping track food as neither easy nor difficult are more likely to report any FAH events for a given day. SNAP households which have the last SNAP issued in last 10 days or between 11 to 20 days are positively associated with the probability of having any FAH events on that day. The number of inbound calls is positively related to the likelihood of having a FAH on a given day. The larger the household, the more likely they will have at least a FAH event for a given day. Both FAH and FAFH cumulative number of events are positively related to the likelihood of having a FAH on a given day. Cumulative FAH expenditure and cumulative reporting status counts are negatively related to the likelihood of having a FAH on a given day. In addition, several demographic and socio-economic status variables (including gender, age, education and race) also have significant associations with the probability that a household member has a FAH event for a given day. Male household members and respondents in younger age groups are less likely to report any FAH events for a given day. Comparing with those who do not have a high school degree, higher education groups are more likely to report FAH events for a given day. Comparing with white respondents, those who identified themselves as African Americans are less likely to report FAH events.

For Review Only

Second, the substantive regression model shows that BMI, relationship to the PR, feedback on the completion of meals and snacks forms, days-since-issuance of SNAP status, financial condition, household size, whether the day reported is a weekend day, cumulative number of events and cumulative FAFH expenditure all have significant relationships with expenditures on FAH events for a given day. Compared to those with a normal weight or those who are underweight, those with missing BMI information reported less expenditure on FAH events. Household members other than the PRs spend less on FAH events (as expected, given how PRs were selected for the survey). Compared to the households in which the PRs complete meals and snacks forms every day, those households where the PRs did not complete the meals and snacks forms spend more on FAH events. Compared to non-SNAP households, SNAP households with last SNAP issuance in the past 10 days spend more on FAH events. SNAP households with a longer DSI, on the other hand, spend less on FAH events comparing with non-SNAP households. Those households whose PR reported being "in over their head" financially tend to spend less on FAH events on a given day, compared to those who are financially very comfortable and secure. Not surprisingly, the larger the household, the more that respondents spend on FAH events. More spending are reported on weekend days. Both FAH and FAFH cumulative number of events are negatively related to FAH expenditure on a given day. Cumulative FAFH expenditure is positively related to the FAH expenditure on a given day. In addition, several demographic and socio-economic status variables also have significant relationships with FAH expenditure. Males spend less on FAH events than females, as do African Americans and those who identified themselves as multiple races or non-White, non-Black and not multiple races. Compared to those who are over 55 years old, those who are 31-55

For Review Only

years old spend more on FAH events, as do those with higher education and those who are married.

FAFH (Supplemental Table 3)

First, the selection model indicates that members of households who started the diary survey on Friday have a lower probability of reporting a paid FAFH event with each additional fielding day. Note that BMI, relationship to the PR, days-since-issuance of SNAP status, household size, employment status and PR's feedbacks on meals and snacks forms completion, member participation, the ease of keeping track of food and cumulative reporting status counts all have significant relationships with the probability that a household member has a paid FAFH event for a given day. Contrary to the FAH model results, those who are overweight or obese are more likely to have paid FAFH events on a given day. Consistent with the reporting of FAH events, household members other than the PRs are less likely to report any paid FAFH events for a given day. Compared to non-SNAP households, SNAP households which had their last SNAP benefit issued in the last 10 days, or between 11 to 20 or 21 to 31 days, are less likely to have any paid FAFH events on a given day. Employed household members are more likely to report paid FAFH events for a given day. Household members belonging to households that did not complete the meals and snacks form every day are less likely to report a paid FAFH event than households that completed the form every day. Members of a household in which the PR reported that the ease of keeping track of foods was very difficult or difficult or neither easy nor difficult were more likely to report a paid FAFH event than did members in households who found it very easy or easy to keep track of foods. Cumulative reporting status counts variable is negatively related to the likelihood of having a FAFH on a given day.

For Review Only

With regard to sociodemographic predictors, household members who are male, with higher income, younger age groups (compared with the reference group of >55 years) or have college or higher levels of education are more likely to report a paid FAFH event. Compared to those who have never been married, those who are married or divorced are less likely to report a paid FAFH event.

Second, consistent with the substantive FAH model, all household members other than the PRs spend less on FAFH events. Among the seven start-date cohorts, the Sunday start-date cohort spends the least on FAFH events. PRs who reported that it was not too difficult to make ends meet, occasionally difficult or tough to make ends meet spend less on FAFH events than do households that were financially very comfortable and secure. Members of larger households spend more on FAFH events. Household members spend more on FAFH events on weekends and when the data collection took place in the summer. Both FAH and FAFH cumulative number of events are negatively related to FAFH expenditure on a given day. Cumulative FAFH expenditure is positively related to the FAFH expenditure on a given day. In addition, several demographic and socio-economic status variables also demonstrate significant relationships with FAFH expenditures. Compared with those who are over 55, those who are between 11 and 20 spend less on FAFH events, as do African Americans and those who identified themselves as non-White, non-Black and not multiple races (compared with the reference group of White household members). Compared to those who have low education, those with a college degree or higher education spend more on FAFH events. Compared to those who have never been married, married household members spend more on FAFH events, and widowed household members spend less.

For Review Only

Supplemental Table 2. FAH Heckman full model results.

	Estimate	95% Confidence Intervals	
Substantive regression model			
Days-since-issuance of SNAP benefits (reference = Non-SNAP household)			
10 or less days since last SNAP issuance	12.75	7.626	17.879
11-20 days since last SNAP issuance	-6.43	-10.387	-2.481
21-31 days since last SNAP issuance	-5.85	-10.837	-0.871
32+ days since last SNAP issuance	-16.97	-24.226	-9.716
Weekend (reference = Not weekend)	8.05	4.727	11.365
Cumulative number of FAH events	-16.67	-21.172	-12.171
Cumulative number of FAFH events	-6.18	-9.098	-3.261
Cumulative FAH expenditures	0.04	-0.078	0.164
Cumulative FAFH expenditures	0.51	0.230	0.786
Cumulative reporting status counts	1.91	-3.700	7.522
BMI (reference = Not overweight (<24.9))			
Overweight (25.0–29.9)	3.74	-0.128	7.610
Obese (>30.0)	3.00	-0.488	6.486
NA, missing, or error	-12.58	-21.743	-3.414
Gender (reference = Female)			
Male	-5.43	-9.028	-1.829
Age (reference = >55)			
11-15	19.43	-12.176	51.036
16-20	4.28	-5.182	13.747
21-30	4.08	-1.066	9.224
31-55	7.57	3.351	11.798
Individual income (log-transformed)	-0.45	-1.284	0.387
Education (reference = Below high school)			
High school grads, below college	6.42	2.875	9.965
College grads and above	9.42	5.363	13.482
Marital status (reference = Never married)			
Married	8.99	5.144	12.837
Widowed	1.49	-3.660	6.646
Divorced	-1.20	-5.785	3.387
Separated	2.21	-6.778	11.202
Race (reference = White)			
Black/African American	-12.02	-16.432	-7.610
Other races	-11.19	-15.263	-7.124
Multiple races	-2.00	-12.098	8.102
Relationship with PR (reference = Primary respondent)			
Partner or children	-13.19	-18.322	-8.060
Other relatives	-10.27	-18.528	-2.013
Non-relatives	-9.63	-17.605	-1.658
Employed (reference = Not employed)	2.42	-1.666	6.501
Time gap	1.10	-0.050	2.251

For Review Only

Start date (reference = Friday)			
Sunday	4.81	-1.530	11.154
Monday	9.77	1.458	18.082
Tuesday	-0.13	-5.767	5.501
Wednesday	1.38	-4.059	6.816
Thursday	0.87	-4.674	6.411
Saturday	3.99	-2.027	9.998
Feedback on the completion of meals and snacks forms (reference = Every day)			
>1, not every day	2.81	-1.027	6.640
Once before end	-4.77	-13.965	4.428
Once at end	4.87	-3.788	13.535
Did not complete	8.00	0.282	15.726
Missing	12.55	-5.292	30.398
Feedback on member participation (reference = Very easy / Easy)			
Neutral	1.32	-2.717	5.354
Very difficult / Difficult	0.06	-7.510	7.640
Not applicable / Missing	4.37	-0.662	9.410
Feedback on ease of keeping track of food (reference = Very easy / Easy)			
Neutral	-2.06	-6.739	2.610
Very difficult / Difficult	1.03	-6.317	8.386
Missing	-7.68	-27.907	12.551
Financial condition (reference = Very comfortable and secure)			
Without much difficulty	0.47	-3.997	4.936
Occasionally difficulty	-2.05	-6.593	2.502
Tough to make ends meet	-2.72	-7.900	2.461
In over your head	-11.75	-18.231	-5.276
How often bills are reviewed for accuracy (reference = Never)			
Rarely	-0.69	-7.595	6.216
Sometimes	-1.42	-7.061	4.230
Usually	1.94	-4.211	8.081
Always	1.67	-3.587	6.925
Not applicable	-4.44	-19.762	10.891
Household size	2.41	0.856	3.963
Total inbound calls	-0.83	-2.048	0.382
Total outbound calls	-0.17	-1.137	0.798
Summer indicator (reference = Not summer)	-0.28	-3.407	2.842
Total initial interview time (reference = 0 - 4.2 min)			
4.2 min - 6.3 min	-1.63	-6.416	3.157
>6.3 min	-2.73	-7.873	2.412
The total number of contacts before the diary survey (reference = 0-3)			

4-5	0.74	-2.918	4.402
5-7	-1.18	-5.750	3.388
>7	2.75	-2.116	7.615
Ever refused before diary survey data collection	-5.66	-13.146	1.830
Intercept	30.20	14.858	45.541
Selection model			
Fielding day	-0.02	-0.049	0.003
Interaction between start date and fielding day (reference = Friday)			
Sunday	0.00	-0.037	0.039
Monday	-0.02	-0.065	0.022
Tuesday	-0.03	-0.061	0.011
Wednesday	0.00	-0.034	0.036
Thursday	0.00	-0.038	0.029
Saturday	-0.04	-0.070	0.000
Days-since-issuance of SNAP benefits (reference = Non-SNAP household)			
10 or less days since last SNAP issuance	0.17	0.096	0.238
11-20 days since last SNAP issuance	0.14	0.060	0.221
21-31 days since last SNAP issuance	0.01	-0.075	0.086
32+ days since last SNAP issuance	0.09	-0.205	0.378
Weekend (reference = Not weekend)	-0.02	-0.106	0.073
Interaction between age group and weekend			
11-15#Weekend	-0.10	-0.412	0.214
16-20#Weekend	0.09	-0.108	0.294
21-30#Weekend	0.08	-0.058	0.208
31-55#Weekend	0.08	-0.023	0.190
Cumulative number of FAH events	0.31	0.228	0.402
Cumulative number of FAFH events	0.06	0.017	0.109
Cumulative FAH expenditures	-0.01	-0.011	-0.005
Cumulative FAFH expenditures	0.00	-0.005	0.002
Cumulative reporting status counts	-0.39	-0.491	-0.290
BMI (reference = Not overweight (<24.9))			
Overweight (25.0–29.9)	-0.02	-0.076	0.034
Obese (>30.0)	-0.07	-0.131	-0.018
NA, missing, or error	-0.01	-0.235	0.211
Gender (reference = Female)			
Male	-0.15	-0.197	-0.095
Age (reference = >55)			
11-15	-1.29	-1.551	-1.024
16-20	-0.57	-0.734	-0.400
21-30	-0.18	-0.287	-0.065
31-55	-0.01	-0.097	0.077
Individual income (log-transformed)	0.00	-0.012	0.008
Education (reference = Below high school)			

For Review Only

High school grads, below college	0.10	0.039	0.164
College grads and above	0.18	0.108	0.249
Marital status (reference = Never married)			
Married	0.12	0.051	0.184
Widowed	0.00	-0.106	0.097
Divorced	0.02	-0.063	0.101
Separated	0.07	-0.061	0.193
Race (reference = White)			
Black/African American	-0.10	-0.182	-0.017
Other races	0.04	-0.022	0.103
Multiple races	0.01	-0.127	0.141
Relationship with PR (reference = Primary respondent)			
Partner or children	-0.92	-0.987	-0.856
Other relatives	-0.94	-1.123	-0.752
Non-relatives	-0.82	-0.974	-0.664
Employed (reference = Not employed)	-0.04	-0.101	0.019
Time gap	-0.01	-0.027	0.006
Start date (reference = Friday)			
Sunday	-0.04	-0.172	0.092
Monday	-0.01	-0.159	0.136
Tuesday	0.03	-0.095	0.147
Wednesday	-0.07	-0.185	0.055
Thursday	-0.07	-0.182	0.041
Saturday	0.01	-0.109	0.125
Feedback on the completion of meals and snacks forms (reference = Every day)			
>1, not every day	-0.01	-0.064	0.049
Once before end	0.11	-0.098	0.316
Once at end	0.06	-0.066	0.183
Did not complete	0.01	-0.152	0.166
Missing	0.00	-0.189	0.182
Feedback on member participation (reference = Very easy / Easy)			
Neutral	0.02	-0.062	0.097
Very difficult / Difficult	-0.02	-0.109	0.064
Not applicable / Missing	-0.01	-0.079	0.066
Feedback on ease of keeping track of food (reference = Very easy / Easy)			
Neutral	0.08	0.009	0.156
Very difficult / Difficult	0.10	-0.013	0.213
Missing	-0.01	-0.234	0.204
Financial condition (reference = Very comfortable and secure)			
Without much difficulty	0.04	-0.026	0.098
Occasionally difficulty	0.05	-0.020	0.112
Tough to make ends meet	0.02	-0.056	0.095

For Review Only

In over your head	0.02	-0.074	0.122
How often bills are reviewed for accuracy (reference = Never)			
Rarely	0.00	-0.103	0.110
Sometimes	0.08	-0.012	0.176
Usually	0.00	-0.089	0.087
Always	0.05	-0.029	0.131
Not applicable	-0.16	-0.409	0.082
Household size	0.05	0.028	0.067
Total inbound calls	0.02	0.003	0.037
Total outbound calls	0.00	-0.012	0.017
Summer indicator (reference = Not summer)	0.06	-0.020	0.135
Interaction between age group and summer indicator			
11-15#Summer	0.12	-0.219	0.458
16-20#Summer	-0.03	-0.260	0.208
21-30#Summer	-0.05	-0.175	0.066
31-55#Summer	-0.05	-0.154	0.050
Total initial interview time (reference = 0 - 4.2 min)			
4.2 min - 6.3 min	0.01	-0.044	0.070
>6.3 min	0.10	0.034	0.158
The total number of contacts before the diary survey (reference = 0-3)			
4-5	0.00	-0.059	0.052
5-7	-0.01	-0.069	0.057
>7	0.02	-0.048	0.091
Ever refused before diary survey data collection	0.06	-0.069	0.188
Intercept	-0.38	-0.596	-0.166

Supplemental Table 3. FAFH Heckman full model results

	Estimate	**95% Confidence Intervals**	
Substantive regression model			
Days-since-issuance of SNAP benefits (reference = Non-SNAP household)			
10 or less days since last SNAP issuance	-0.39	-1.867	1.096
11-20 days since last SNAP issuance	-0.58	-3.063	1.903
21-31 days since last SNAP issuance	-0.71	-2.984	1.560
32+ days since last SNAP issuance	-0.60	-3.013	1.806
Weekend (reference = Not weekend)	5.84	4.428	7.252
Cumulative number of FAH events	-3.27	-5.113	-1.420
Cumulative number of FAFH events	-2.16	-3.378	-0.949
Cumulative FAH expenditures	-0.02	-0.052	0.021
Cumulative FAFH expenditures	0.35	0.233	0.460

For Review Only

Cumulative reporting status counts	1.09	-1.400	3.578
Gender (reference = Female)			
Male	-0.54	-1.774	0.694
Age (reference = >55)			
11-15	-7.16	-10.097	-4.215
16-20	-3.32	-5.937	-0.694
21-30	-1.90	-3.825	0.021
31-55	-1.67	-3.445	0.106
Individual income (log-transformed)	0.14	-0.237	0.510
Education (reference = Below high school)			
High school grads, below college	0.78	-0.920	2.478
College grads and above	3.45	1.415	5.479
Marital status (reference = Never married)			
Married	2.73	1.204	4.257
Widowed	-3.28	-6.184	-0.368
Divorced	-1.19	-3.127	0.756
Separated	0.18	-3.378	3.731
Race (reference = White)			
Black/African American	-1.53	-2.749	-0.318
Other races	2.87	0.319	5.425
Multiple races	-1.45	-4.085	1.191
Relationship with PR (reference = Primary respondent)			
Partner or children	-4.92	-6.642	-3.205
Other relatives	-5.22	-7.312	-3.137
Non-relatives	-4.00	-6.342	-1.648
Employed (reference = Not employed)	-2.22	-4.525	0.083
Start date (reference = Friday)			
Sunday	-2.74	-4.867	-0.606
Monday	-1.20	-3.667	1.258
Tuesday	-1.05	-3.193	1.103
Wednesday	-1.31	-3.698	1.077
Thursday	-1.44	-3.671	0.784
Saturday	-1.06	-3.626	1.498
Financial condition (reference = Very comfortable and secure)			
Without much difficulty	-1.98	-3.745	-0.211
Occasionally difficulty	-3.29	-5.203	-1.381
Tough to make ends meet	-3.13	-5.022	-1.230
In over your head	-1.36	-5.278	2.566
Household size	0.46	0.047	0.883
Summer indicator (reference = Not summer)	1.80	0.683	2.920
Total initial interview time (reference = 0 - 4.2min)			
4.2 min - 6.3 min	-0.53	-1.844	0.789
>6.3 min	-0.26	-1.921	1.402

For Review Only

The total number of contacts before the diary survey (reference = 0-3)			
4-5	-0.52	-2.062	1.026
5-7	0.20	-1.461	1.867
>7	1.17	-0.704	3.041
Ever refused before diary survey data collection	-0.91	-3.971	2.142
Intercept	16.86	11.896	21.821
Selection model			
Fielding day	-0.01	-0.031	0.010
Interaction between start date and fielding day (reference = Friday)			
Sunday	0.02	-0.013	0.046
Monday	-0.04	-0.081	-0.002
Tuesday	-0.03	-0.054	0.002
Wednesday	-0.03	-0.058	-0.004
Thursday	-0.02	-0.050	0.007
Saturday	-0.01	-0.049	0.021
Days-since-issuance of SNAP benefits (reference = Non-SNAP household)			
10 or less days since last SNAP issuance	-0.31	-0.385	-0.231
11-20 days since last SNAP issuance	-0.25	-0.326	-0.172
21-31 days since last SNAP issuance	-0.18	-0.266	-0.103
32+ days since last SNAP issuance	0.21	-0.104	0.529
Weekend (reference = Not weekend)	-0.15	-0.232	-0.065
Interaction between age group and weekend			
11-15#Weekend	-0.33	-0.542	-0.113
16-20#Weekend	-0.05	-0.202	0.096
21-30#Weekend	0.09	-0.036	0.210
31-55#Weekend	-0.07	-0.174	0.034
Cumulative number of FAH events	0.07	-0.009	0.156
Cumulative number of FAFH events	0.45	0.397	0.502
Cumulative FAH expenditures	0.00	-0.005	-0.001
Cumulative FAFH expenditures	0.00	0.000	0.008
Cumulative reporting status counts	-0.46	-0.565	-0.360
BMI (reference = Not overweight (<24.9))			
Overweight (25.0–29.9)	0.06	0.003	0.121
Obese (>30.0)	0.12	0.063	0.184
NA, missing, or error	0.01	-0.218	0.233
Gender (reference = Female)			
Male	0.13	0.074	0.181
Age (reference = >55)			
11-15	0.63	0.426	0.829
16-20	0.36	0.179	0.536
21-30	0.11	-0.006	0.236
31-55	0.14	0.042	0.236

For Review Only

Individual income (log-transformed)	0.02	0.005	0.030
Education (reference = Below high school)			
High school grads, below college	0.07	-0.004	0.150
College grads and above	0.14	0.054	0.222
Marital status (reference = Never married)			
Married	-0.10	-0.168	-0.030
Widowed	-0.07	-0.185	0.054
Divorced	-0.11	-0.203	-0.020
Separated	-0.10	-0.264	0.058
Race (reference = White)			
Black/African American	0.02	-0.047	0.096
Other races	-0.04	-0.112	0.027
Multiple races	0.06	-0.126	0.248
Relationship with PR (reference = Primary respondent)			
Partner or children	-0.55	-0.611	-0.480
Other relatives	-0.58	-0.693	-0.470
Non-relatives	-0.62	-0.785	-0.461
Employed (reference = Not employed)	0.12	0.045	0.190
Time gap	0.00	-0.014	0.018
Start date (reference = Friday)			
Sunday	0.04	-0.089	0.172
Monday	0.16	0.004	0.320
Tuesday	0.11	-0.002	0.229
Wednesday	0.15	0.033	0.262
Thursday	0.18	0.058	0.296
Saturday	0.10	-0.036	0.234
Feedback on the completion of meals and snacks forms (reference = Every day)			
>1, not every day	-0.09	-0.147	-0.024
Once before end	-0.06	-0.241	0.120
Once at end	-0.03	-0.156	0.093
Did not complete	-0.04	-0.166	0.083
Missing	0.07	-0.168	0.311
Feedback on member participation (reference = Very easy / Easy)			
Neutral	0.10	0.036	0.172
Very difficult / Difficult	0.01	-0.074	0.101
Not applicable / Missing	-0.17	-0.260	-0.077
Feedback on ease of keeping track of food (reference = Very easy / Easy)			
Neutral	0.12	0.046	0.198
Very difficult / Difficult	0.13	0.024	0.230
Missing	0.11	-0.167	0.392
Financial condition (reference = Very comfortable and secure)			

For Review Only

Without much difficulty	-0.01	-0.077	0.055
Occasionally difficulty	0.01	-0.064	0.092
Tough to make ends meet	-0.05	-0.139	0.032
In over your head	-0.06	-0.162	0.039
Household size	0.02	0.005	0.041
Total inbound calls	0.00	-0.017	0.022
Total outbound calls	0.00	-0.021	0.019
Summer indicator (reference = Not summer)	-0.01	-0.101	0.081
Interaction between age group and summer indicator			
11-15#Summer	-0.50	-0.704	-0.292
16-20#Summer	-0.09	-0.291	0.107
21-30#Summer	0.02	-0.117	0.165
31-55#Summer	0.04	-0.072	0.160
Total initial interview time (reference = 0 - 4.2min)			
4.2 min - 6.3 min	0.07	0.004	0.132
>6.3 min	0.02	-0.050	0.088
The total number of contacts before the diary survey (reference = 0-3)			
4-5	0.02	-0.044	0.084
5-7	0.02	-0.050	0.094
>7	0.11	0.035	0.180
Ever refused before diary survey data collection	0.09	-0.030	0.220
Intercept	-0.66	-0.877	-0.451

Appendix D. Estimation and prediction of number of items using the negative binomial regression model.

In the manuscript, we presented the results of imputing missing data for the continuous expenditures using Heckman selection model. We also investigated the size of underreporting for a related categorical variable: number of FAH / FAFH events.

To predict the number of events, we tested Poisson regression and negative binomial regression models for both FAH and FAFH events. The negative binomial regression models all had significantly larger-than-zero over-dispersion parameters (alpha – for FAH events: 0.17, CI: 0.11, 0.26; for FAFH events: 0.01, CI: 0.0003, 0.3220), suggesting that the distribution of the

For Review Only

number of events did not have equal mean and variance. Our data seem to violate the assumption of the Poisson distribution, so we proceeded with negative binomial regressions. In each of the 500 bootstrapped samples, we fitted a negative binomial model to a given event count variable, and generated estimates of the parameters in this model; we then applied the model to predict the missing counts of events. In fitting the negative binomial models, we used the same set of covariates as in the Heckman model for expenditure. This procedure was done for FAH and FAFH events separately, and the complex sampling features of FoodAPS were accounted for when we fitted the negative binomial model. To account for the correlations between days within the same person, we specified individual id as the cluster variable when running the models.

The results are presented below in Supplemental Table 4 through Supplemental Table 7. These findings for the number of events are consistent with the findings on expenditures in the sense that the imputations adjusted the original estimate in the same directions. In other words, when the imputation-adjusted estimate of expenditure suggests that the original estimate of expenditure was an underestimation (overestimation), the imputation-adjusted estimate of number of events would also suggest that the original estimate of number of events was an underestimation (overestimation). Take the mean FAH number of events per person per day as an example (the top section of Supplemental Table 4); the imputation adjusted the estimate from 0.27 down to 0.25, and this adjustment was statistically significant, suggesting that the original estimate was an overestimation. The same conclusion could be reached by looking at the top section of Table 3 (in the manuscript); the imputation adjusted the mean FAH expenditure from 8.26 down to 8.02, suggesting that the original estimate was a significant overestimation. Such a consistent pattern indicates that we achieved the same conclusions for different indicators of expending actions (expenditures in dollar and number of spending events).

For Review Only

Supplemental Table 4. Mean number of FAH events per person per day and bootstrap imputation results by target group.

	Bootstrap samples	
	Mean	se
Overall (Weighted Respondent Mean = 0.27, SE = 0.01)		
Number of FAH events each person reported per day	0.27	0.01
Number of FAH events after imputation	0.25	0.01
Difference between the two means	0.02***	<0.01
SNAP (Weighted Respondent Mean = 0.28, SE = 0.01)		
Number of FAH events each person reported per day	0.28	0.01
Number of FAH events after imputation	0.27	0.01
Difference between the two means	0.02***	<0.01
Non-SNAP below 185% (Weighted Respondent Mean = 0.29, SE = 0.01)		
Number of FAH events each person reported per day	0.29	0.01
Number of FAH events after imputation	0.27	0.01
Difference between the two means	0.02***	<0.01
Non-SNAP 185 or higher (Weighted Respondent Mean = 0.26, SE = 0.01)		
Number of FAH events each person reported per day	0.26	0.01
Number of FAH events after imputation	0.25	0.01
Difference between the two means	0.02***	<0.01

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.0001$

Supplemental Table 5. Mean number of FAFH events per person per day and bootstrap imputation results by target group.

	Bootstrap samples	
	Mean	se
Overall (Weighted Respondent Mean = 0.62, SE = 0.01)		

For Review Only

Number of FAFH events each person reported per day	0.62	0.01
Number of FAFH events after imputation	0.61	0.01
Difference between the two means	<0.01***	<0.01
SNAP (Weighted Respondent Mean = 0.48, SE = 0.02)		
Number of FAFH events each person reported per day	0.48	0.02
Number of FAFH events after imputation	0.48	0.02
Difference between the two means	<0.01***	<0.01
Non-SNAP below 185% (Weighted Respondent Mean = 0.52, SE = 0.02)		
Number of FAFH events each person reported per day	0.52	0.03
Number of FAFH events after imputation	0.53	0.02
Difference between the two means	-0.01***	<0.01
Non-SNAP 185 or higher (Weighted Respondent Mean = 0.66, SE = 0.02)		
Number of FAFH events each person reported per day	0.66	0.02
Number of FAFH events after imputation	0.66	0.02
Difference between the two means	0.01***	<0.01

* p < 0.05; ** p < 0.01; *** p < 0.0001

Supplemental Table 6. Weekly population total number of FAH events (in millions) and bootstrap imputation results, overall and by target group.

	Bootstrap samples	
	Total	se
Overall (Weighted population total = 393.41, SE = 25.86)		
Total number of FAH events	394.28	26.11
Total number of FAH events after imputation	448.52	27.53
Difference between the two total amounts	-54.24***	3.29
% change	-11.11%	
SNAP (Weighted population total = 57.75, SE = 4.71)		
Total number of FAH events	57.88	4.57
Total number of FAH events after imputation	69.07	5.41
Difference between the two total amounts	-11.18***	1.28

For Review Only

% change	-14.29%	
Non-SNAP below 185% (Weighted population total = 64.95, SE = 4.89)		
Total number of FAH events	65.05	4.71
Total number of FAH events after imputation	74.38	5.09
Difference between the two total amounts	-9.33***	0.91
% change	-14.29%	
Non-SNAP 185 or higher (Weighted population total = 270.71, SE = 21.02)		
Total number of FAH events	271.34	21.46
Total number of FAH events after imputation	305.07	22.69
Difference between the two total amounts	-33.73***	2.62
% change	-9.68%	

* p < 0.05; ** p < 0.01; *** p < 0.0001

For Review Only

Supplemental Table 7. Weekly population total number of FAFH events (in millions) and bootstrap results, overall and by target group.

	Bootstrap samples	
	Total	se
Overall (Weighted population total = 898.32, SE = 56.96)		
Total number of FAFH events	899.73	58.46
Total number of FAFH events after imputation	1082.94	61.99
Difference between the two total amounts	-183.21***	9.74
% change	-16.67%	
SNAP (Weighted population total =98.36, SE = 7.44)		
Total number of FAFH events	98.69	7.09
Total number of FAFH events after imputation	125.80	9.12
Difference between the two total amounts	-27.11***	3.32
% change	-23.08%	
Non-SNAP below 185% (Weighted population total = 117.53, SE = 7.91)		
Total number of FAFH events	117.19	7.84

Total number of FAFH events after imputation	146.40	10.02
Difference between the two total amounts	-29.21***	3.00
% change	-20.00%	
Non-SNAP 185 or higher (Weighted population total = 682.43, SE = 54.44)		
Total number of FAFH events	683.85	55.79
Total number of FAFH events after imputation	810.73	59.98
Difference between the two total amounts	-126.89***	8.77
% change	-16.05%	

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.0001$

For Review Only

Appendix E. Stata and R code for the imputation methodology.

Since the code for FAH and FAFH follows the same procedures, we present the code using FAH as an illustration of our methods. The code includes two components – the bootstrap code using R and the Heckman model code using Stata.

1. FAH Bootstrap: R Code

```
###########################
#   FAH Bootstrap R Code  #
###########################


#removes all objects from the current workspace (R memory)
rm(list = ls())

#change working directory
setwd("K:/Research/Umich_West/datasets")

#read dataset
fah0 <- read.table("K:/Research/Umich_West/datasets/bootstrapfah.csv", header = TRUE, sep = ",")

#load household level dataset with the target group variable
#recode the variable
load("faps_household.rdata")
table(faps_household$targetgroup)
faps_household$targetgrp3cat <- NA
faps_household$targetgrp3cat[faps_household$targetgroup==1] <- 2
faps_household$targetgrp3cat[faps_household$targetgroup==2] <- 2
faps_household$targetgrp3cat[faps_household$targetgroup==3] <- 3
faps_household$targetgrp3cat[faps_household$targetgroup==4] <- 1

#merge the target group variable to the dataset
fah<-merge(fah0,data.frame("hhnum"=faps_household$hhnum,"targetgrp3cat"=faps_household$targetgrp3cat))

#the survey design features
require(survey)
design <- svydesign(ids = ~tspsu,
                    strata = ~tsstrata,
```


For Review Only

```
                    nest = FALSE,
                    data = fah,
                    weights = ~hhwgt
)


# bootstrap following Rao and Wu(1988)
fah_uid.RW <- as.svrepdesign(design = design,
                             type = "subbootstrap",
                             replicates = 500)

#save replicate weights
rwfah <- weights(fah_uid.RW, type = "analysis")


library(dplyr) # need this package to merge the dataset using left_join

#merge the bootstrap replicate weights with id
person_uid <- fah$person_uid
day_ind <- fah$day_ind
repwts_withID <- cbind.data.frame(person_uid, day_ind, as.data.frame(as.matrix(rwfah)))

#merge the bootstrap replicate weights data file with the FAH data file
fahall2 <- left_join(fah,repwts_withID, by=c("person_uid"="person_uid", "day_ind"="day_ind"))


#save the boostrap results
setwd("K:/Research/Umich_West/datasets")
save(fah_uid.RW, fah, fahall2, file="bootstrapfah_final.Rdata")

#use stattransfer to save fahall2 as a stata data file
```

2. FAH Heckman Model and Final Estimates: Stata code

```
/*************************************************************/
/* merge bootstrap data and original data with predicators */
/*************************************************************/

*prepare the data with bootstrap replicate weights for merging
use "K:\Research\...\fahall2.dta", clear
```


For Review Only

```
tostring person_uid, replace
sort person_uid day_ind
save fahall4merge, replace

*prepare the response data for merging
use "K:\Research\...\fah_fafh_10plus_0.dta", clear
sort person_uid day_ind
save fah_fafh_10plus_0_4merge, replace

*merge the two datasets together
use fahall4merge, clear
merge 1:1 person_uid day_ind using fah_fafh_10plus_0_4merge

*save the merged dataset
save "K:\Research\...\fahall_weights.dta", replace


/********************************************************/
/* Create an overall dataset to append the estimates    */
/********************************************************/

*create a dataset to append estimates later - overallfah.dta;
*need to delete this first row of data later;
clear
input converged orig_mean updt_mean orig_sum updt_sum
. 1 1 1 1
end

save "K:\Research\...\overallfah.dta", replace

*create another dataset to append target-group estimates later - overallfah_tgtgrp.dta;
*need to delete this first row of data later;
clear
input converged targetgrp3cat orig_mean updt_mean orig_sum updt_sum
. 1 1 1 1 1
end
save "K:\Research\...\overallfah_tgtgrp.dta", replace


/**********************/
/* Heckman models      */
```


For Review Only

```
/**********************/

*run Heckman model, save estimates using each of the 500 replicate weights;
foreach i of num 1/500{
            use "K:\Research\...\fahall_weights.dta", clear

            * set up the design features;
            svyset person_uid [pweight=V`i'], strata(tsstrata) singleunit(centered)

            *run the Heckman model using the ith bootstrap replicate weight;
            capture noisily svy: heckman fahdollar  i.bmi_ind_4cat ib2.sex ib6.agecat log_income i.edu  ///
            ib5.marital i.race i.relation16 time_gap ib6.start_date  i.ff1_6cat i.ff2_4cat i.ff3_4cat ///
            i.snapwhenhh i.fincondition i.billrevfreq  hhsize  inboundcalls outboundcalls i.employed  ///
            i.weekend i.jun_aug cum_status cum_fafhdollar cum_fahdollar i.init_time_cat i.num_all_cat  ///
            i.refuseTotal cum_fafhfreq cum_fahfreq, ///
            select (fah_indicator=day_ind_centered i.bmi_ind_4cat ib2.sex ib6.agecat log_income i.edu  ///
            ib5.marital i.race i.relation16 time_gap ib6.start_date i.start_date#c.day_ind_centered ///
            i.ff1_6cat i.ff2_4cat i.ff3_4cat i.snapwhenhh outboundcalls inboundcalls ///
            i.fincondition i.billrevfreq  hhsize  i.employed i.weekend i.agecat#i.weekend i.jun_aug ///
            i.agecat#i.jun_aug  cum_status cum_fafhdollar cum_fahdollar cum_fafhfreq cum_fahfreq ///
            i.init_time_cat i.num_all_cat i.refuseTotal) iter(100)



            *generate an indicator for model convergence;
            gen didconverge = e(converged)

            *obtain predicated FAH expenditures;
            predict pred_fahdollar, yexpected

            *generate the variable θ̂imp which include the imputed expenditures
            gen fahdollar_updat_temp = .
            replace fahdollar_updat_temp = fahdollar if (status2==1)
            replace fahdollar_updat_temp = 0 if (status2==2)
            replace fahdollar_updat_temp = pred_fahdollar if (status2>=3)

            *save the dataset which includes θ̂imp ;
            save "K:\Research\...\fahall_weights_temp.dta", replace

            *obtain the mean and sum estimates and save them to a separate dataset - loopfah`i'.dta;
```

```
*orig_mean is θ̂_obs;
*updt_mean is θ̂_imp;
collapse (mean) converged = didconverge orig_mean =  fahdollar  ///
updt_mean =  fahdollar_updat_temp (sum) orig_sum = fahdollar   ///
updt_sum=fahdollar_updat_temp [pw=V`i']

save "K:\Research\...\loopfah`i'.dta", replace

*append loopfah`i'.dta to overallfah.dta and save the dataset by replacing the old dataset;
use  "K:\Research\...\overallfah.dta", clear
append using "K:\Research\...\loopfah`i'.dta"
save "K:\Research\...\overallfah.dta", replace

*remove the dataset loopfah`i'.dta;
erase "K:\Research\...\loopfah`i'.dta"


* obtain the mean and sum estimates by target groups and
* save them to a separate dataset - loopfah`i'.dta;
use "K:\Research\...\fahall_weights_temp.dta", clear

collapse (mean) converged = didconverge orig_mean =  fahdollar  ///
updt_mean =  fahdollar_updat_temp (sum) orig_sum = fahdollar ///
updt_sum=fahdollar_updat_temp [pw=V`i'], by(targetgrp3cat)


save "K:\Research\...\loopfah`i'.dta", replace

*append loopfah`i'.dta to overallfah_tgtgrp.dta and save the dataset;
use  "K:\Research\...\overallfah_tgtgrp.dta", clear
append using "K:\Research\...\loopfah`i'.dta"
save "K:\Research\...\overallfah_tgtgrp.dta", replace

*remove the dataset loopfah`i'.dta;
erase "K:\Research\...\loopfah`i'.dta"

      }

/*********************************/
```

```
/*   Final overall estimates     */
/*********************************/

*use the dataset which includes saved estimates for the 500 bootstrap sample;
use "K:\Research\...\overallfah.dta", clear
*remove the first row in the dataset, which is not part of the estimates;
drop if converged==.
*remove the estimates from the unconverged models
drop if converged==0
*compute differences between θ̂obs and θ̂imp
gen mean_diff = orig_mean-updt_mean
*compute differences between observed sum and sum measure including imputed expenditures;
gen sum_diff = orig_sum - updt_sum

*obtain final estimates;
summarize orig_mean updt_mean orig_sum updt_sum mean_diff
*tabstat is used to remove the scientific notations;
tabstat  orig_sum updt_sum sum_diff, stat(mean sd) format(%20.2f)

/*************************************/
/* Final estimates by target groups   */
/*************************************/

*use the dataset which includes saved target-group estimates for the 500 bootstrap sample;
use "K:\Research\...\overallfah_tgtgrp.dta", clear
*remove the first row in the dataset, which is not part of the estimates;
drop if converged==.
*remove the estimates from the unconverged models
drop if converged==0
*compute differences between θ̂obs and θ̂imp
gen mean_diff = orig_mean-updt_mean
*compute differences between observed sum and sum measure including imputed expenditures;
gen sum_diff = orig_sum - updt_sum

*obtain final estimates by target group;
sort targetgrp3cat
by targetgrp3cat: summarize orig_mean updt_mean orig_sum updt_sum mean_diff sum_diff
tabstat  orig_sum updt_sum sum_diff , stat(mean sd) by(targetgrp3cat) format(%20.2f)
```


For Review Only