Comparing demographic and health characteristics of new and existing SNAP recipients: application of a machine learning algorithm

Rita Hamad,[1] *Zachary S Templeton,*[2] *Lena Schoemaker,*[2] *Michelle Zhao,*[2] *and Jay Bhattacharya*[2]

[1]Philip R Lee Institute for Health Policy Studies, Department of Family and Community Medicine, University of California San Francisco, San Francisco, CA; and [2]Center for Primary Care and Outcomes Research/Center for Health Policy, Department of Medicine, Stanford University, Stanford, CA

ABSTRACT

Background: The Supplemental Nutrition Assistance Program (SNAP) expanded significantly after the Great Recession of 2008–2009, but no studies have characterized this new group of recipients. Few data sets provide details on whether an individual is a new or established recipient of SNAP.

Objective: We sought to identify new and existing SNAP recipients, and to examine differences in sociodemographic characteristics, health, nutritional status, and food purchasing behavior between new and existing recipients of SNAP after the recession.

Methods: We created a probabilistic algorithm to identify new and existing SNAP recipients using the 1999–2013 waves of the Panel Study of Income Dynamics. We applied this algorithm to the National Household Food Acquisition and Purchase Survey (FoodAPS), fielded during 2012–2013, to predict which individuals were likely to be new SNAP recipients. We then compared health and nutrition characteristics between new, existing, and never recipients of SNAP in FoodAPS.

Results: New adult SNAP recipients had higher socioeconomic status, better self-reported health, and greater food security relative to existing recipients, and were more likely to smoke relative to never recipients. New child SNAP recipients were less likely to eat all meals and had lower BMI relative to existing recipients. New SNAP households exhibited differences in food access and expenditures, although dietary quality was similar to that of existing SNAP households.

Conclusion: We developed a novel algorithm for predicting new and existing SNAP recipiency that can be applied to other data sets, and subsequently demonstrated differences in health characteristics between new and existing recipients. The expansion of SNAP since the Great Recession enrolled a population that differed from the existing SNAP population and that may benefit from different types of nutritional and health services than those traditionally offered. *Am J Clin Nutr* 2019;109:1164–1172.

Keywords: Supplemental Nutrition Assistance Program, nutrition, health, machine learning, diet quality

Introduction

The Supplemental Nutrition Assistance Program (SNAP) provides an important buffer for low-income families against nutritional shortfalls. SNAP has been shown to improve household food security and infant birth weights, reduce expenditures on health care for children, and lessen the prevalence of poverty (1–6). Since 2008, largely as a response to the Great Recession, there have been considerable expansions in SNAP and other US safety net programs (7). A substantial portion of the growth in SNAP was due to new recipients rather than increased benefits among existing recipients. The number of recipients grew from 28 million individuals in 2008 to a peak of 48 million in 2013 (8).

It is unclear whether these new recipients were individuals of higher socioeconomic status (SES) who were newly affected by the economic challenges presented by the recession, or whether they represented more vulnerable, previously eligible individuals who newly enrolled owing to increased financial duress. Moreover, new recipients may have been less familiar with community resources, more likely to be concerned about the stigma of receiving welfare benefits, and less aware of the restrictions on how the funds can be spent (9, 10).

Supported by a grant from the USDA through the University of Kentucky Center for Poverty Research, and by National Heart, Lung, and Blood Institute grant K08 HL132106 (to RH). The collection of the Panel Study of Income Dynamics was supported in part by NIH grants R01 HD069609 and R01 AG040213 and National Science Foundation grants SES 1157698 and 1623684.

Supplemental Figures 1 and 2, Supplemental Tables 1–5, and Supplemental Methods are available from the "Supplementary data" link in the online posting of the article and from the same link in the online table of contents at https://academic.oup.com/ajcn/.

Address correspondence to RH (e-mail: rita.hamad@ucsf.edu).

Abbreviations used: FAFH, Food Away from Home; FAH, Food at Home; FoodAPS, National Household Food Acquisition and Purchase Survey; HEI, Healthy Eating Index; PSID, Panel Study of Income Dynamics; SES, socioeconomic status; SNAP, Supplemental Nutrition Assistance Program.

Received June 1, 2018. Accepted for publication November 16, 2018.

First published online March 4, 2019; doi: https://doi.org/10.1093/ajcn/nqy355.

Previous studies have compared the demographic and health characteristics of SNAP recipients and nonrecipients (11–14). However, no studies to our knowledge have compared the characteristics of new and existing SNAP recipients, in particular since the Great Recession when the population of new SNAP recipients grew substantially. It is therefore important to study differences between new SNAP recipients and existing recipients, enabling public health practitioners to better tailor health and nutritional interventions if this new population represents a distinct group with different needs.

The primary hurdle to answering this research question is that few (if any) data sets have information on whether an individual is a new or an existing SNAP recipient, in addition to detailed data on health and nutritional characteristics. We therefore developed a probabilistic algorithm to identify individuals who were likely to be new recipients of SNAP since the recession, applying machine learning methods to the Panel Study of Income Dynamics (PSID) (15), a longitudinal, nationally representative data set with detailed information on social welfare receipt. Yet the PSID has limited information on the health of study participants. Consequently, we applied the probabilistic algorithm to unique data from the National Household Food Acquisition and Purchase Survey (FoodAPS), a cross-sectional, nationally representative data set with detailed information on food purchases and health (16)—but not history of SNAP receipt—to predict whether FoodAPS participants were likely to be new or existing SNAP recipients. We then examined the ways in which new recipients of SNAP benefits differed relative to existing recipients. The application of this algorithm demonstrates the utility of leveraging multiple data sets to comprehensively examine the social and economic context of health, and enables us to produce some of the first evidence comparing sociodemographic and health characteristics of new and existing SNAP recipient families.

Methods

Data

This study involved the linkage of 2 rich data sets that include data on SNAP recipients.

PSID.

The PSID is a longitudinal survey that has followed a nationally representative sample of households since 1968, collecting data on respondents' SES and social welfare receipt, but with only limited information on health and nutrition. Critically, the PSID contains annual information on families' receipt of SNAP benefits. We restricted our sample to the 1999–2013 waves of the study, capturing health and sociodemographic data for the period 1999–2012. Earlier time periods may differ in terms of determinants of SNAP recipiency. We included individuals 18 y or older who had nonmissing information on the demographic variables of interest ($n = 21{,}806$). Because SNAP recipiency is determined by adult characteristics, and because most of the variables of interest were at the household level (e.g., income), observations from child participants were not included in our probabilistic algorithm.

SNAP status in the PSID was recorded at the household level, so we assumed that this applied to all individuals in the household. A given individual in the PSID was classified as being a new SNAP recipient in a given year if it was the first year in which that household received SNAP during our study period. In subsequent years, individuals were classified as existing SNAP recipients, even if they did not receive SNAP benefits in that year. This modeling decision reflected our research question of interest, namely to characterize individuals who were likely to be new recipients of SNAP who had never before received benefits.

Of note, <5% of values were missing for variables included in our analysis. Complete case analysis is not thought to introduce bias at such low levels of absence (17–20), so we did not impute missing values.

FoodAPS.

FoodAPS, commissioned by the USDA, is a cross-sectional study of US households' food acquisitions ($n = 4826$). Our study included households that had nonmissing information on the input variables for the probabilistic algorithm ($n = 4775$ households, including 4548 children and 9607 adults) (**Supplemental Figure 1**). Fewer than 5% of values were missing for variables included as predictors in our models. FoodAPS included both SNAP recipient and nonrecipient households, although it did not differentiate between new and existing SNAP recipients. The study was fielded between April 2012 and January 2013 and collected information on nutrition, expenditures, and distance traveled for each food purchase made by household members over a 7-d period. Interviews were conducted with each household's primary respondent, who was the "main food shopper or meal planner" (21). The primary respondent reported data on household members' health status, tobacco use, BMI, meals eaten, and household diet quality.

Variables.

Predictors. FoodAPS included socioeconomic and demographic variables that were similar to variables collected in the PSID. Variables common to both data sets were used to develop our probabilistic algorithm for determining whether an individual was a new or an existing SNAP recipient. These included sex, age, educational attainment (less than high school, high school, at least some college), household income (categorized into quartiles), employment status, housing status, household size, household vehicle ownership, state of residence, types of individuals in the household (spouse, child, grandchild, other relative, or nonrelative), and language of survey interview (English compared with other), as well as race/ethnicity, age, and marital status of the head of household.

Health characteristics. We examined several household-level nutritional characteristics that are likely to be associated with SNAP receipt. FoodAPS participants were queried about purchases of food prepared at home ["Food at Home" (FAH)] and away from the home ["Food Away from Home" (FAFH)] across the entire survey week. Because the questionnaire did not include information on individual-level consumption of these foods, we aggregated these purchases at the household level to determine nutritional content, types of foods, and locations where food was acquired. From these we calculated daily intake overall and in specific categories of macronutrients. Household food security was calculated from 10 questions derived from the USDA's 30-d Food Security Scale (22). We assessed dietary quality using the Healthy Eating Index (HEI) (23). In addition to nutritional content, we examined characteristics related to food access through the construction of variables such as total distance to FAH purchases and the number of nearby fast-food restaurants. Additional characteristics included the type of primary store—the store in which the household does most of its food shopping—and the distance between home and the primary store.

We also included several individual-level health characteristics available for adult and child household members, including health status, tobacco use (adults only), BMI, and meals eaten.

Data analysis

Probabilistic algorithm.

Because FoodAPS did not include a variable indicating whether respondents were new or existing SNAP recipients, we developed a probabilistic machine learning algorithm to distinguish these groups. Any variable shared by the PSID and FoodAPS—except the health characteristics—could potentially serve as a predictor variable. To avoid overfitting our model by including predictors that are unassociated with SNAP recipiency, and because of computational limitations, we selected variables that would plausibly be associated based on prior empirical and theoretical work. We also aimed to avoid strong functional form assumptions about the way that our predictors entered our prediction function. Given these parameters, we employed the machine learning method of lasso regression (24) and the Glmnet package (25) using R version 3.4.3 (R Foundation for Statistical Computing). This algorithm was developed using the PSID and was then used to identify people who were most likely to be new SNAP recipients in FoodAPS.

Lasso regression is a machine learning method that encourages sparsity in regression models. Predictors in the lasso are selected by minimizing the residual sum of squares plus a penalty equal to the sum of the absolute values of the regression coefficients. Lasso selects a subset of predictors from a large set of covariates by shrinking coefficient estimates to zero for those variables that are poor predictors of the outcome variable. We used 10-fold cross-validation to fit a multinomial logit model with 3 classes: new SNAP recipient, existing SNAP recipient, and never received SNAP ("never recipient"). Lasso selected predictors from the aforementioned predictor variables, as well as all possible pairwise interactions. This method is an improvement on propensity scores, which can also be used to produce predicted probabilities of a given treatment. However, propensity score methods do not perform variable selection; the inclusion of unnecessary covariates can increase bias while reducing statistical efficiency (26, 27), whereas lasso eliminates variables that are poor predictors.

Because the predictor variables that a single lasso iteration selects may be subject to sampling error, we estimated our lasso model on 50 bootstrap samples from the PSID. Each bootstrap run generated 50 unique sets of coefficients. We applied the model generated in each bootstrap run to generate predictions of new, existing, and never SNAP recipiency in the FoodAPS data, producing 50 sets of predicted SNAP status for each household. Household probabilities were calculated as the mean of the parents' probabilities, or the single parent's probability for unmarried individuals. These 3 household probabilities (new, existing, and never SNAP recipiency) were used as the primary predictors in our FoodAPS models. This algorithm has been made freely available on the public website of the senior author to enable other researchers to apply the algorithm to other data sets that do not provide direct information about SNAP recipiency.

Differences between new and existing SNAP recipients.

We next examined the association between SNAP recipiency and the health characteristics of interest. We employed ordinary least squares linear models using Stata 14.2 (StataCorp LP) and regressed each of the aforementioned individual-level health characteristics on the household probabilities of being an existing SNAP recipient or a never recipient (reference group: new recipient). Regressions were conducted separately for children and adults. Covariates included race/ethnicity, age, age^2, sex, education (for adult respondents), monthly household income, number of household adults and children, and month of survey interview. The coefficient on the 2 primary predictor variables can be interpreted as the change in the health characteristic of interest for every percentage increase in the likelihood of being an existing recipient or never recipient, relative to new recipients.

To estimate SEs, we ran 2500 iterations of each model (50 bootstrapped probabilities generated from PSID $\times$ 50 sets of replicate household weights provided by FoodAPS). Robust SEs were clustered by household to account for correlated observations and heteroskedasticity. For each health characteristic, our final estimate was the mean of the coefficient estimates from all 2500 iterations. Ninety-five percent CIs reflect the 2.5th and 97.5th percentiles of coefficient estimates.

These analyses were conducted for individual health characteristics, and we conducted similar analyses to examine food-purchasing behaviors at the household level. Here, the 2 primary predictor variables were the household-level probabilities of being an existing SNAP recipient and being a never recipient. Covariates included race/ethnicity, age, age^2, sex, and education of the head of household; monthly household income; number of household adults and children; and month of survey interview. As in the individual-level analyses, we ran linear models for each household health variable 2500 times to generate coefficient estimates and CIs.

Because the magnitudes of coefficient estimates and CIs varied based on the scale of the health variable, we divided these estimates by the weighted mean of each health characteristic. This allowed for the presentation of coefficient estimates of several health variables in a single graph and did not alter levels of statistical significance.

Ethics

This study was approved by the Stanford University Institutional Review Board (protocol #38109).

Results

Sample characteristics

In the PSID, the mean age was 42.3 y, with 21,806 individuals followed for a mean of 8.8 y. Existing recipients were more likely than new recipients to be female, black, unmarried, renting, living in public housing, unemployed, and without a car (**Table 1**).

In the FoodAPS sample, the mean age among adults was 43.8 y, and the mean age among children was 8.8 y (**Table 2**). Compared with PSID participants, FoodAPS participants had lower income, were less likely to be black, and more likely to be Hispanic.

Probabilistic algorithm

Lasso selected among 4095 variables as described above. Tenfold cross-validation resulted in a median size penalty (λ) of 2.8×10^{-4} (median minimum mean squared error = 0.531) across 50 bootstrap samples (**Supplemental Table 1**).

To validate our algorithm, we tested its accuracy in classifying SNAP and non-SNAP households in FoodAPS. To measure the tradeoff between type I and type II error, we plotted a receiver operator characteristic curve, which demonstrated an area under the curve of 0.78 (**Supplemental Figure 2**). In this case, type I error refers to the rate at which our model incorrectly identifies a household as receiving SNAP, and type II error refers to the

TABLE 1 PSID sample characteristics by SNAP status[1]

	Never (n = 147,220)		New (n = 3082)		Existing (n = 41,048)		Difference
	Mean/%[3]	SD	Mean/%	SD	Mean/%	SD	(new − existing)[2]
A: individual characteristics							
Age, y	43.9	16.5	37.4	15.0	37.0	14.9	0.4
Male	48.2	—	43.3	—	37.3	—	6.0*
Education							
Less than high school	11.0	—	23.4	—	30.1	—	− 6.7*
High school	35.0	—	45.7	—	42.8	—	2.8*
At least some college	53.9	—	31.0	—	27.1	—	3.9*
Employment status							
Employed	72.5	—	58.1	—	52.6	—	5.6*
Looking for work	3.9	—	16.0	—	16.2	—	− 0.2
Not looking for work	23.5	—	25.8	—	31.2	—	− 5.4*
B: household characteristics							
Age of head, y	47.5	15.1	42.5	14.2	42.0	13.9	0.5*
Race of head							
White	67.3	—	36.4	—	25.3	—	11.1*
Black	23.4	—	51.4	—	61.2	—	− 9.8*
Hispanic	5.9	—	9.3	—	10.9	—	− 1.6*
Other	3.5	—	3.0	—	2.6	—	0.3
Married (head)	75.3	—	54.8	—	44.3	—	10.5*
Number in family	3.0	1.4	3.4	1.7	3.6	1.8	− 0.2*
Family income, $	85,213	115,358	35,420	32,950	34,123	34,584	1297*
Owns or is buying home	74.1	—	48.8	—	34.8	—	14.0*
Lives in public housing	1.7	—	6.1	—	11.9	—	− 5.8*
Car ownership							
0 cars	5.6	—	16.4	—	27.9	—	− 11.5*
1 car	22.5	—	35.2	—	37.2	—	− 2.0*
2 cars	43.2	—	30.7	—	24.0	—	6.7*
3 cars	18.9	—	13.2	—	7.8	—	5.4*
≥4 cars	9.9	—	4.5	—	3.2	—	1.3*
Types of individuals in household							
Spouse	77.1	—	59.6	—	47.1	—	12.6*
Child	55.6	—	70.0	—	73.6	—	− 3.7*
Grandchild	3.7	—	13.2	—	12.6	—	0.6
Nonrelative	0.9	—	2.0	—	1.8	—	0.4
Other relative	5.7	—	11.6	—	11.2	—	0.2
English interview	97.0	—	94.2	—	92.9	—	1.3*

[1] n = 21,806 individuals, 191,350 person-years. Sample includes PSID adults who were followed in the 1999–2013 waves of the study and had nonmissing values for the above variables. *$P < 0.05$. PSID, Panel Study of Income Dynamics; SNAP, Supplemental Nutrition Assistance Program.

[2] The final column shows the difference in means between new and existing recipients (statistical significance determined through 2-sample t tests with equal variances).

[3] Means and SDs were calculated without sample weights. Weighted statistics are available upon request.

TABLE 2 FoodAPS sample characteristics[1]

	Mean/%[2]	SD
A: individual characteristics		
Adults		
Age, y	43.8	17.0
Male	45.8	—
Race		
White	55.8	—
Black	14.3	—
Hispanic	22.1	—
Other	7.9	—
Education		
Less than high school	16.2	—
High school	35.4	—
At least some college	48.4	—
Employment status		
Employed	54.1	—
Looking for work	8.5	—
Not looking for work	37.4	—
Children		
Age, y	8.8	5.4
Male	50.9	—
Race		
White	44.2	—
Black	17.1	—
Hispanic	31.1	—
Other	7.6	—
B: household characteristics		
Age of head, y	46.8	16.4
Race of head		
White	58.8	—
Black	15.3	—
Hispanic	18.9	—
Other	7.0	—
Married (head)	42.1	—
Number in family	2.8	1.7
Family income, $	44,622	50,754
Owns or is buying home	47.8	—
Lives in public housing	7.2	—
Car ownership		
0 cars	15.8	—
1 car	35.5	—
2 cars	31.8	—
3 cars	11.3	—
≥4 cars	5.7	—
Types of individuals in household		
Spouse	52.1	—
Child	49.4	—
Grandchild	5.6	—
Nonrelative	7.3	—
Other relative	13.8	—
English interview	91.5	—

[1] n = 9607 (adults), 4548 (children), and 4775 (households). Sample includes FoodAPS individuals with nonmissing values for the demographic variables listed. Education and employment status were reported only for adults because FoodAPS defined those variables for individuals 16 y or older. FoodAPS, National Household Food Acquisition and Purchase Survey.

[2] Means and SDs were calculated without sample weights. Weighted statistics are available upon request.

rate at which our model fails to correctly identify a non-SNAP household.

Differences between new and existing SNAP recipients

Individual health characteristics—adults.

Among FoodAPS adults, existing SNAP recipients were less likely to report excellent or very good health status than were new SNAP recipients ($\beta = -0.46$; 95% CI: −0.65, −0.27) (**Figure 1**A), but were similar with respect to BMI and tobacco use. In contrast, never recipients were less likely to report tobacco use than were new SNAP recipients ($\beta = -1.33$; 95% CI: −1.82, −0.66), but were similar with respect to BMI and health status.

Existing SNAP recipients were less likely to eat all meals than were new SNAP recipients (Figure 1A; **Supplemental Table 2**), with no statistically significant differences between new and never recipients.

Individual health characteristics—children.

Patterns among FoodAPS children differed from those of adults. Existing recipients were more likely to have greater BMI than were new recipients ($\beta = 0.14$; 95% CI: 0.08, 0.28) (Figure 1B), but were similar with respect to health status. Never recipients had similar BMI and health status to new recipients. Existing and never recipients were more likely to eat all meals than were new SNAP recipients (Figure 1B; **Supplemental Table 3**).

Household health characteristics.

Both existing and never SNAP households purchased more energy, carbohydrates, sugar, total fat, and alcohol per household member than new SNAP households (**Figure 2**). Existing SNAP households purchased a greater proportion of energy from FAH events than new SNAP households ($\beta = 0.10$; 95% CI: 0.04, 0.24). Both existing and never SNAP households had smaller proportions of FAFH events that occurred at fast-food restaurants than new SNAP households.

On measures of diet quality, existing households had similar total HEI scores to new households (Figure 2; $\beta = 0.05$; 95% CI: −0.03, 0.09), although results did differ—both improved and worsened—for some components of the HEI score (**Supplemental Table 4**). This result was consistent with households' self-assessment of diet quality, which indicated no statistically significant differences between existing and new households ($\beta = 0.05$; 95% CI: −0.11, 0.18). In contrast, never households had improved scores on 7 of the 13 HEI components relative to new households (Supplemental Table 4). Consistent with this result, never households had better diet quality than new SNAP households, as determined through total HEI scores (Figure 2; $\beta = 0.09$; 95% CI: 0.02, 0.13) and households' self-assessment ($\beta = 0.24$; 95% CI: 0.02, 0.40).

Existing SNAP households had worse food security than new SNAP households (**Figure 3**; $\beta = -0.80$; 95% CI: −1.12, −0.64), whereas never recipient households showed no statistically significant difference in food security compared with new households ($\beta = 0.00$; 95% CI: −0.28, 0.15). Existing SNAP and never recipient households traveled a greater total

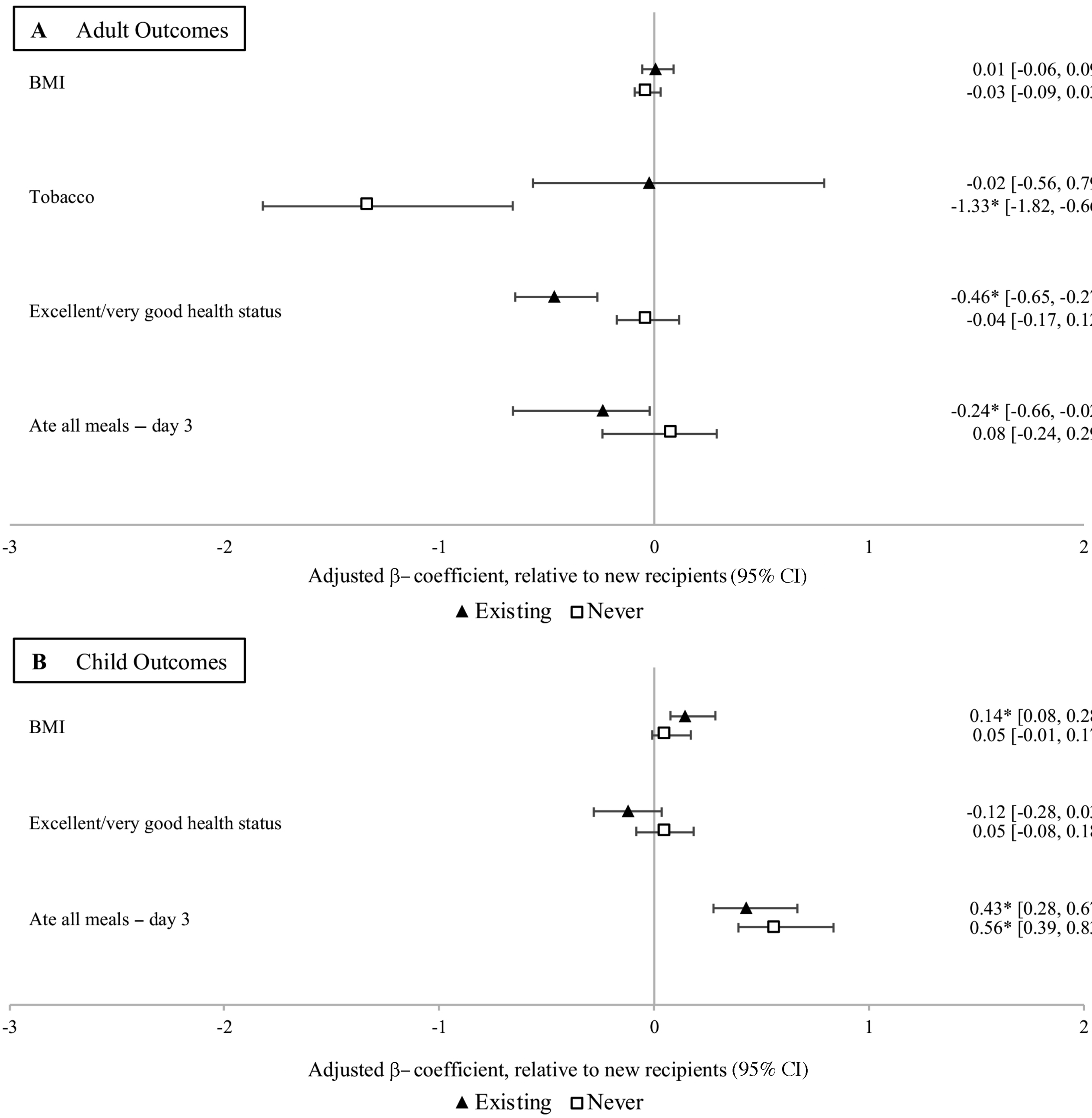

FIGURE 1 Association between SNAP and recipiency and individual-level health characteristics (reference group: new SNAP recipients). $^*P < 0.05$. $n = 9607$ (adults) and 4548 (children). Sample includes National Household Food Acquisition and Purchase Survey (A) adults and (B) children with nonmissing values for variables of interest. Models involved linear regressions with 2 primary predictor variables: the household probability of being an existing SNAP recipient and the household probability of being a never recipient (reference: new SNAP recipients). Models adjusted for race/ethnicity, age, age^2, sex, education (adults only), monthly household income, number of household adults and children, and month of interview. To allow for presentation in a single graph, we divided the β coefficients and associated CIs from these models by the weighted mean of each outcome. "Ate all meals—day 3" was selected as a representative variable. Other indicator variables for the remaining days of the survey week were similar and can be found in Supplemental Tables 2 and 3. SNAP, Supplemental Nutrition Assistance Program.

distance to FAH events than new households. In addition, existing SNAP and never recipient households had greater distances between their homes and primary food stores than new SNAP households. Existing SNAP and never recipient households were less likely to report their primary store to be a superstore and more likely to report a primary store other than a superstore or supermarket, relative to new households (**Supplemental Table 5**). FoodAPS classifies superstores as "very large supermarkets [or] 'big box' stores … [that] include … mass merchandisers under a single roof and membership retail/wholesale hybrids" (28). In contrast, supermarkets are smaller and include most grocery stores. Existing households had fewer SNAP retailers within 1 mile than new households ($\beta = -1.52$; 95% CI: -1.98, -0.70).

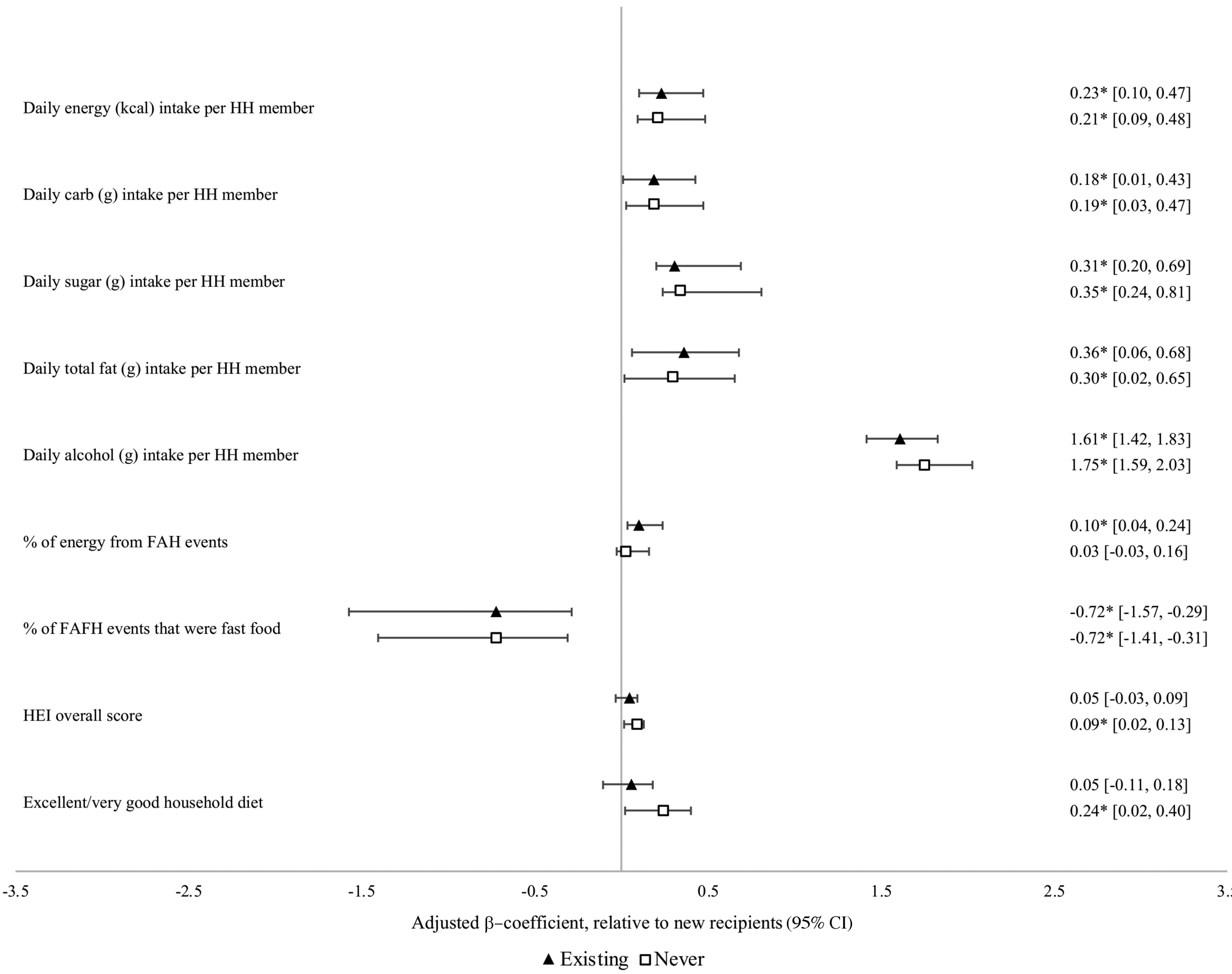

FIGURE 2 Association between SNAP and recipiency and household nutrition (reference group: new SNAP recipients). $^{*}P < 0.05$. $n = 4775$. Sample includes all National Household Food Acquisition and Purchase Survey households with nonmissing values for variables of interest. Models involved linear regressions with 2 primary predictor variables: the household probability of being an existing SNAP recipient and the household probability of being a never recipient (reference: new SNAP recipients). Models adjusted for race/ethnicity, age, age^2, sex, and education of heads of household; monthly household income; number of household adults and children; and month of interview. To allow for presentation in a single graph, we divided the β coefficients and associated CIs from these models by the weighted mean of each outcome. Please refer to the **Supplemental Methods** for more information on the construction of HEI scores and Supplemental Table 4 for results on HEI component scores. FAFH, Food Away from Home; FAH, Food at Home; HEI, Healthy Eating Index; HH, household; SNAP, Supplemental Nutrition Assistance Program.

Existing SNAP households spent less than new SNAP households on FAH items (Figure 3; $\beta = -0.60$; 95% CI: -0.71, -0.41) but were similar with respect to FAFH expenditures ($\beta = 0.05$; 95% CI: -0.21, 0.28). Existing households were more likely than new households to go out for dinner and less likely to prepare food for dinner at home. Never households spent less on FAH items ($\beta = -0.53$; 95% CI: -0.64, -0.28) and more on FAFH items ($\beta = 0.41$; 95% CI: 0.21, 0.61) compared with new households. As with existing households, never households were more likely than new households to go out for dinner and less likely to prepare food for dinner at home.

Discussion

Our study involved the development and application of a probabilistic algorithm using 2 large linked data sets, finding that new SNAP recipients are a distinct group with key differences relative to both existing SNAP recipients and never recipients. Demographic characteristics of new recipients appear consistent with their SES being higher relative to existing recipients, but not as high as that of never recipients.

Although new and existing SNAP households were similar with respect to some health and nutrition characteristics (e.g., poor diets as assessed through HEI scores and self-reporting), new recipients had better health on several measures such as food security, eating meals in the last week, and overall health status. These results suggest that the expansion of the SNAP program over the last decade has enrolled a population of higher SES and better health and nutrition than the existing SNAP population. This supports a hypothesis that the expansion enrolled individuals who had perhaps only recently been impoverished as a result of the Great Recession. These individuals may have

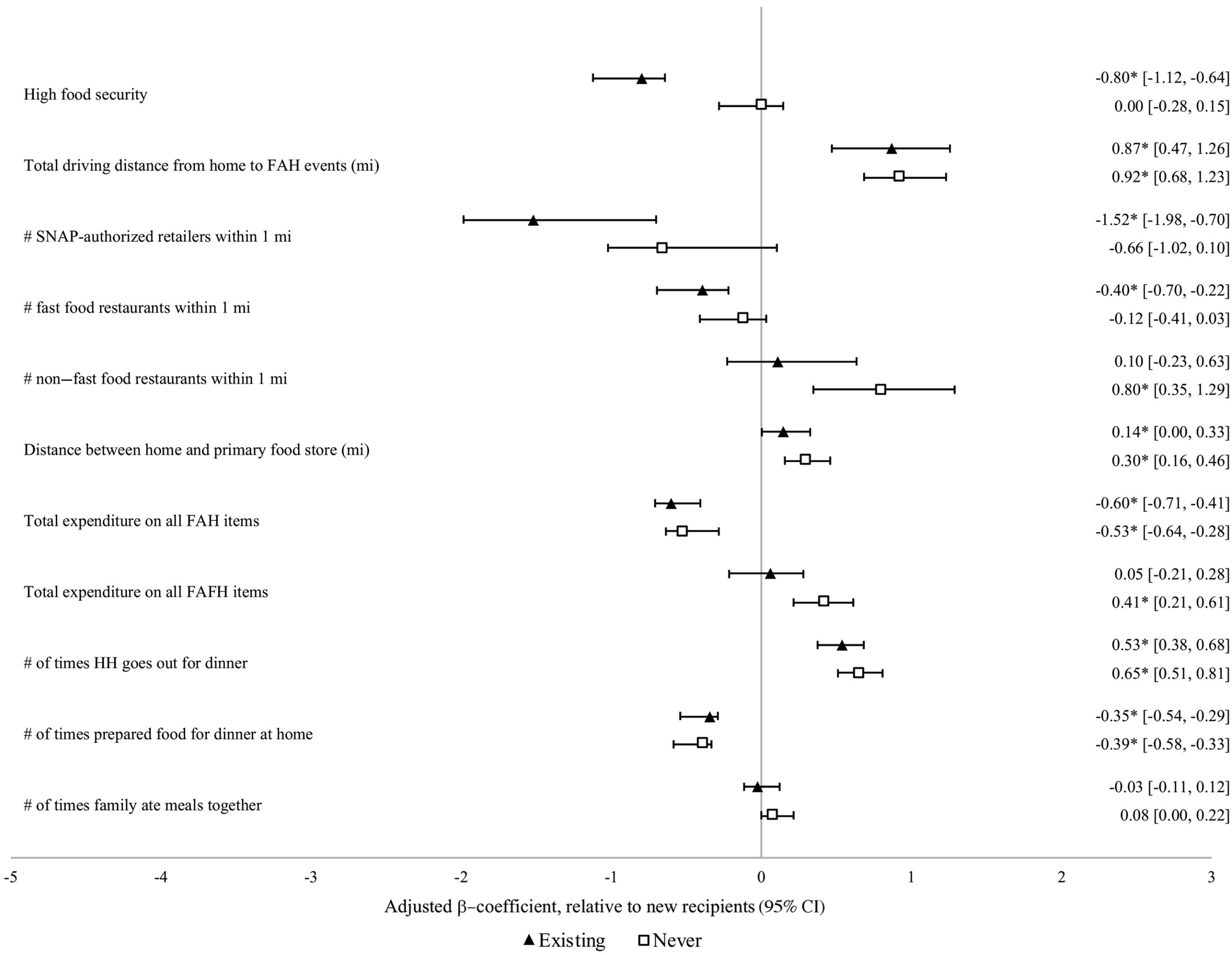

FIGURE 3 Association between SNAP and recipiency and household food access (reference group: new SNAP recipients). $^*P < 0.05$. $n = 4775$. Sample includes all FoodAPS households with nonmissing values for variables of interest. Models involved linear regressions with 2 primary predictor variables: the household probability of being an existing SNAP recipient and the household probability of being a never recipient (reference: new SNAP recipients). Models adjusted for race/ethnicity, age, age^2, sex, and education of heads of household; monthly household income; number of household adults and children; and month of interview. The food security indicator was based on the USDA's 30-d Adult Food Security Scale (22). To allow for presentation in a single graph, we divided the β coefficients and associated CIs from these models by the weighted mean of each outcome. The last 3 variables refer to the 7-d period before the FoodAPS interview. FAFH, Food Away from Home; FAH, Food at Home; FoodAPS, National Household Food Acquisition and Purchase Survey; HEI, Healthy Eating Index; HH, household; mi, mile; SNAP, Supplemental Nutrition Assistance Program.

fewer—or different—obstacles to achieving good health, and future interventions targeting SNAP recipients may want to separately examine impacts among these heterogeneous subgroups (29, 30).

A novel contribution of our study was the development of a probabilistic algorithm to distinguish between new and existing SNAP recipients. This algorithm is freely available on the senior author's website, and we hope that it will be employed by future investigators to add value to existing data sets that do not contain this key variable of interest. This will allow for additional work to examine the effects of SNAP among different subgroups, or to further characterize this population based on other characteristics not available in FoodAPS, e.g., geography. Another contribution of our study is the examination of multiple types of demographic, health, and nutritional characteristics that differentiate new and existing SNAP recipients; this will allow practitioners to develop interventions that target potential risk factors, e.g., food purchasing behaviors or transportation problems.

Our study has several limitations. First, both the PSID and FoodAPS contain sociodemographic variables that are self-reported, which likely suffer from measurement error. In particular, prior work has demonstrated underreporting of SNAP receipt in national surveys (31). If underreporting is conditional on new compared with existing SNAP recipiency, this may introduce bias into the algorithm. Future studies should attempt to externally validate the algorithm using linked survey and administrative data, or in other surveys that include information on new and existing SNAP recipiency. Nevertheless, our study is illustrative of the strengths of linking multiple data sets with complementary data elements to understand the social and economic contexts of health. Second, our analysis is correlational, and therefore we cannot determine the causal relation between SNAP recipiency and the health characteristics we examined. However, this was not the goal of our study; rather, we aimed to descriptively assess demographic, nutritional, and health differences between new and existing recipients. Furthermore, the PSID only includes information on SNAP recipiency at the household level, which we assumed applies to all individuals in a household. This likely caused some misclassification and attenuation in the accuracy of our probabilistic algorithm. In addition, FoodAPS was a cross-sectional survey; future studies could expand on our analysis by examining trajectories of SNAP recipiency and resulting health outcomes in a longitudinal fashion, which would strengthen causal inference. Finally, our categorization of new, existing, and never recipients did not distinguish between those who are chronically on SNAP and those who intermittently receive the benefit; both were categorized as "existing" recipients, because of our specific interest in evaluating the characteristics of new SNAP recipients. Future work could characterize the different subgroups of existing SNAP recipients.

Because new SNAP recipients since the Great Recession are distinct from existing recipients in terms of both demographics and nutritional and health characteristics, existing programs targeted to traditional SNAP recipients may not be appropriate for this group. Given that new recipients are generally better off than existing recipients, it may be more impactful from a public health perspective to instead intervene among those existing recipients who may have more long-standing challenging socioeconomic circumstances.

The authors' responsibilities were as follows—RH and JB: designed the research; LS and ZST: collected and cleaned the data; RH, ZST, MZ, and JB: analyzed the data and performed the statistical analyses; RH and ZST: wrote the manuscript; RH: had primary responsibility for final content; and all authors: read and approved the final manuscript. None of the authors reported a conflict of interest related to the study.

References

1. Swann CA. Household history, SNAP participation, and food insecurity. Food Policy 2017;73:1–9.
2. Gundersen C, Kreider B, Pepper JV. Partial identification methods for evaluating food assistance programs: a case study of the causal impact of SNAP on food insecurity. Am J Agric Econ 2017;99: 875–93.
3. Hoynes HW, Schanzenbach DW. Consumption responses to in-kind transfers: evidence from the introduction of the food stamp program. Am Econ J Appl Econ 2009;1:109–39.
4. Bronchetti ET, Christensen G, Hoynes HW. Local food prices, SNAP purchasing power, and child health [Internet]. University of California at Berkeley, Working Paper; 2017 [cited 21 May, 2018]. Available from: https://www.ocf.berkeley.edu/∼garret/NHIS_JMP.pdf.
5. Tiehen L, Jolliffe D, Gundersen C. Alleviating poverty in the United States: the critical role of SNAP benefits [Internet]. USDA Economic Research Service Economic Research Report 132; 2012. Available from: https://www.ers.usda.gov/webdocs/publications/44963/17742_err132_1_.pdf?v=41995.
6. Almond D, Hoynes HW, Schanzenbach DW. Inside the War on Poverty: the impact of food stamps on birth outcomes. Rev Econ Stat 2011;93:387–403.
7. Moffitt RA. The Great Recession and the social safety net. Ann Am Acad Pol Soc Sci 2013;650:143–66.
8. US Department of Agriculture. Supplemental Nutrition Assistance Program participation and costs [Internet]. 2018. Available from: https://fns-prod.azureedge.net/sites/default/files/pd/SNAPsummary.pdf.
9. Sherman J. Surviving the Great Recession. Soc Probl 2013;60:409–32.
10. Ghavami S, Peterman J, Anliker J. Food access, social safety nets, and social stigma among low-income families in Massachusetts. FASEB J 2013;27(1_Supplement):lb371.
11. Leung CW, Ding EL, Catalano PJ, Villamor E, Rimm EB, Willett WC. Dietary intake and dietary quality of low-income adults in the Supplemental Nutrition Assistance Program. Am J Clin Nutr 2012;96:977–88.
12. Grummon AH, Taillie LS. Nutritional profile of Supplemental Nutrition Assistance Program household food and beverage purchases. Am J Clin Nutr 2017;105:1433–42.
13. Lauffer S. Characteristics of Supplemental Nutrition Assistance Program households: fiscal year 2016 [Internet]. Alexandria, VA: USDA Food and Nutrition Service; 2017. Available from: http://www.fns.usda.gov/ops/research-and-analysis.
14. Bitler M. The health and nutrition effects of SNAP. In: Bartfeld J, Gundersen C, Smeeding T, Ziliak J , editors. SNAP matters: how food stamps affect health and well-being. Stanford, CA: Stanford University Press; 2015:134–160.
15. University of Michigan. Panel Study of Income Dynamics, public use dataset. Ann Arbor, MI: Survey Research Center, Institute for Social Research; 2016.
16. Kirlin JA, Denbaly M. FoodAPS National Household Food Acquisition and Purchase Survey [Internet]. US Department of Agriculture, Economic Research Service; 2013[cited 28 Mar, 2014]. Available from: http://www.ers.usda.gov/data-products/foodaps-national-household-food-acquisition-and-purchase-survey.aspx.
17. Bennett DA. How can I deal with missing data in my study? Aust N Z J Public Health 2001;25:464–9.
18. Dong Y, Peng C-YJ. Principled missing data methods for researchers. Springerplus 2013;2:222.
19. Langkamp DL, Lehman A, Lemeshow S. Techniques for handling missing data in secondary analyses of large surveys. Acad Pediatr 2010;10:205–10.
20. Allison PD. Missing data. In: Millsap R, Maydeu-Olivares A , editors. Handbook of quantitative methods in psychology. Thousand Oaks, CA: SAGE Publications; 2009. p. 72–89.
21. US Department of Agriculture. National Household Food Acquisition and Purchase Survey (FoodAPS): user's guide to survey design, data collection, and overview of datasets [Internet]. USDA Economic Research Service; 2016. Available from: https://www.ers.usda.gov/media/8804/0_foodaps-user-guide-puf.pdf.
22. Bickel G, Nord M, Price C, Hamilton W, Cook J. Guide to measuring household food security, revised 2000 [Internet]. Alexandria, VA: USDA Food and Nutrition Service; 2000. Available from: http://www.fns.usda.gov/oane.
23. Reedy J, Lerman JL, Krebs-Smith SM, Kirkpatrick SI, Pannucci TE, Wilson MM, Subar AF, Kahle LL, Tooze JA. Evaluation of the Healthy Eating Index-2015. J Acad Nutr Diet 2018;118: 1622–33.
24. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Series B Stat Methodol 1996;58(1):267–88.
25. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33: 1–22.
26. Greenland S. Invited commentary: variable selection versus shrinkage in the control of multiple confounders. Am J Epidemiol 2007;167:523–9.
27. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. Epidemiology 2009;20:488–95.
28. US Department of Agriculture. National Household Food Acquisition and Purchase Survey (FoodAPS): codebook: household-level data file. [Internet]. 2016. [cited 5 Dec, 2018]. Available from: https://www.ers.usda.gov/data-products/foodaps-national-household-food-acquisition-and-purchase-survey..
29. Leung CW, Hoffnagle EE, Lindsay AC, Lofink HE, Hoffman VA, Turrell S, Willett WC, Blumenthal SJ. A qualitative study of diverse experts' views about barriers and strategies to improve the diets and health of Supplemental Nutrition Assistance Program (SNAP) beneficiaries. J Acad Nutr Diet 2013;113:70–6.
30. Brambila-Macias J, Shankar B, Capacci S, Mazzocchi M, Perez-Cueto FJA, Verbeke W, Traill WB. Policy interventions to promote healthy eating: a review of what works, what does not, and what is promising. Food Nutr Bull 2011;32:365–75.
31. Meyer BD, Mok WKC, Sullivan JX. Household surveys in crisis. J Econ Perspect 2015;29:199–226.