Southern Economic Journal 2019, 86(1), 202–228
DOI: 10.1002/soej.12364

Symposium: Food Access, Program Participation, and Health: Research Using FoodAPS

Estimating the Associations between SNAP and Food Insecurity, Obesity, and Food Purchases with Imperfect Administrative Measures of Participation

Charles Courtemanche,* Augustine Denteh,† and Rusty Tchernis‡

Administrative data are considered the "gold standard" when measuring program participation, but little evidence exists on their potential problems or implications for econometric estimates. We explore these issues using the FoodAPS, a unique data set containing two different administrative measures of Supplemental Nutrition Assistance Program (SNAP) participation and a survey-based measure. We document substantial ambiguity in the two administrative measures and show that they disagree with each other almost as often as they disagree with self-reported participation. Estimated participation and misreporting rates can be meaningfully sensitive to choices made to resolve this ambiguity and disagreement. We explore sensitivity in regression estimates of the associations between SNAP and food insecurity, obesity, and the healthy eating index. The signs are unchanged across the three measures, and the estimates are mostly not statistically different from each other. However, there are some meaningful differences in the magnitudes and levels of statistical significance of the estimates.

JEL Classification: C81, H51, I12, I18

1. Introduction

A growing literature documents the problems with relying on survey measures of program participation, which suffer from significant reporting error, when conducting impact evaluations (Meyer, Mok, and Sullivan 2015; Mittag 2016; Nguimkeu, Denteh, and Tchernis 2019). Administrative data are ordinarily assumed to be the "gold standard" to overcoming these econometric challenges, but relatively little evidence exists on the potential problems with administrative records or econometric strategies to address them. We investigate these issues using data from the FoodAPS, which combines a panel of household purchases with a survey and linked administrative data on Supplemental Nutrition Assistance Program (SNAP) participation from both state enrollment records and electronic benefit transfer (EBT) card expenditures. The data, therefore, provide the unique opportunity to evaluate the reliability of administrative records by comparing the two different administrative measures to each

* Department of Economics, Gatton College of Business and Economics, University of Kentucky, Lexington, KY, 40506-0034, USA; E-mail: courtemanche@uky.edu; corresponding author.
† Department of Economics, 6823 St. Charles Avenue, Tulane University, New Orleans, LA, 70118.
‡ Economics Department, Andrew Young School of Policy Studies, Georgia State University, P.O. Box 3992, Atlanta, GA, 30302-3992, USA.

other as well as to self-reported participation. Moreover, the data also allow us to examine the sensitivity of participation and misreporting rates and estimated associations between SNAP and food insecurity, obesity, and diet healthfulness to different approaches to cleaning and combining the administrative participation variables.

SNAP is the largest means-tested nutrition assistance program in the United States, serving millions of low-income individuals and households. It is administered by the U.S. Department of Agriculture (USDA) with the objectives of increasing food security, reducing hunger, and improving health and well-being of low-income individuals and households by expanding access to food, nutritious diets, and nutrition education (Mabli et al. 2013). The number of Americans receiving SNAP benefits tripled from about 17 million to 46 million between 2000 and 2014, while total spending on SNAP has more than quadrupled from about \$17 billion to almost \$75 billion.[1]

Proponents assert that SNAP participation reduces food insecurity, lifts millions from poverty, and provides a fiscal boost to the economy during downturns (U.S. Department of Agriculture 2012). However, the empirical literature on the impacts of SNAP has produced mixed results. Several studies have documented the expected negative relationship between SNAP and food insecurity (Van Hook and Ballistreri 2006; Nord and Prell 2011; Schmidt, Shore-Shephard, and Watson 2016), but others have found statistically insignificant or even positive associations (Gundersen and Oliveira 2001; Huffman and Jensen 2003; Hofferth 2004; Wilde and Nord 2005; Hoynes and Schanzenbach 2016). SNAP is also often found to be positively correlated with obesity, but some studies find insignificant or negative effects (Meyerhoefer and Pylypchuk 2008; Gundersen 2015; Almada, McCarthy, and Tchernis 2016; Denteh, 2017; Almada and Tchernis 2018; Nguimkeu, Denteh, and Tchernis 2019).

These mixed results reflect two main methodological challenges in evaluating the causal effects of SNAP. The first is nonrandom selection. SNAP participation is endogenous, so there is a strong likelihood that specific unobservable characteristics are correlated with both SNAP participation and nutrition-related outcomes. Such factors might include current or expected future health, human capital, financial stability, and attitudes toward work (Currie 2003; Kreider et al. 2012).

The second identification problem, and the focus of our article, is measurement error in SNAP participation, which occurs when SNAP participants are coded as receiving no benefits when they truly did (false negatives) or vice versa (false positives). Misreporting of SNAP participation in national surveys has been documented with false negatives being much more prevalent than false positives.[2] For instance, the estimated false negative rates for SNAP in various surveys range from 20% to almost 50% (Mittag 2013; Meyer, Mittag, and George 2018). There is a growing literature suggesting that the estimated effect of a misclassified binary explanatory variable (such as SNAP participation) may be substantially biased and may even yield "wrong signs" (Kreider 2010; Kreider et al. 2012; Nguimkeu, Denteh, and Tchernis 2019). Within a one-sided model of endogenous misreporting, Nguimkeu, Denteh, and Tchernis (2019) provide sign-switching results for the ordinary least squares (OLS) estimator even when participation is exogenous. In this case, they show that the OLS estimator yields the wrong sign if misreporting is endogenous, with the size of the sign-switching region increasing with the rate of false negatives and decreasing with the true participation rate.[3] Meyer and Mittag (2018) show that the likelihood of misreporting is systematically related to observable characteristics such as income, employment, and geography, suggesting that it

[1] Statistics are from http://www.fns.usda.gov/pd/supplemental-nutrition-assistance-program-snap.

[2] See Bound, Brown, and Mathiowetz (2001) for a comprehensive review of measurement error in survey data.

[3] Similar severe consequences of reporting errors also occur within an instrumental variables framework (Almada, McCarthy, and Tchernis 2016).

is likely related to unobservable characteristics as well. Most researchers using survey data to study SNAP do not account for the possibility of nonclassical measurement error and the few that do so make assumptions akin to random misreporting.

A fundamental difficulty in dealing with misreporting is that true participation status is unobserved in almost all surveys, and validation data sets that link survey responses to administrative records are scarce. Even when administrative data are available, their usefulness depends crucially on the quality of the linkage. While administrative data are usually considered the "gold standard," they can still be missing, incorrectly entered, or outdated. Some measurement error may therefore remain. By linking survey responses to administrative data on SNAP participation from two different sources, FoodAPS provides a unique opportunity to investigate issues related to measurement error in both self-reported and administrative measures.

Specifically, we use data from the FoodAPS to offer some novel insights related to the reliability of linked administrative SNAP measures. First, we document several sources of ambiguity in both of the administrative measures and show that they are only slightly more strongly correlated with each other than with self-reported participation. Estimated SNAP participation and misreporting rates vary with the coding rules used to resolve this ambiguity and disagreement. We then examine the relationships between SNAP and food insecurity, obesity, and the healthy eating index (HEI). The signs of regression estimates are not sensitive to different coding rules, and the various estimates are largely not statistically different from each other. However, the magnitudes of the estimates and levels of statistical significance exhibit meaningful variability in certain cases. In sum, these results serve as a cautionary tale about uncritically relying on linked administrative records when conducting program evaluation research, but also an illustration of how researchers can still use multiple imperfect measures to reach robust conclusions.

2. Data

The FoodAPS survey is the first nationally representative survey of U.S. households to collect comprehensive data about household food purchases as well as health and nutrition outcomes. FoodAPS is sponsored by the Economic Research Service (ERS) and the Food and Nutrition Service (FNS) of the USDA to support critical research that informs policymaking on health and obesity, food insecurity, and nutrition assistance policy.

The FoodAPS surveyed 4826 households through a multistage sampling design with a target population roughly equally divided into SNAP households, nonparticipating low income households with income less than the poverty guideline, nonparticipating households with income between 100 and 185% of the poverty guideline, and nonparticipating households with income at least equal to 185% of the poverty guideline.[4] Survey questions relate to demographic characteristics, income, program participation, food insecurity, health, weight, and height. Also, FoodAPS contains detailed information about individual food purchases and acquisitions (merged with nutrition information), along with variables related to local food availability and prices. A unique feature of FoodAPS that makes it well-suited for our study is the linked administrative records on SNAP participation for consenting respondents. This presents an opportunity to study SNAP misreporting more thoroughly than past research.

[4] The FoodAPS field operations were conducted from April 2012 through January 2013, during which each participating household provided information on all food acquisitions of all household members during a 7-day interview period.

Participants were interviewed before they were given a survey to record their food purchases for one week. Self-reported SNAP participation comes from this initial interview. The primary respondent (PR) was asked about SNAP receipt, including information on the date of last receipt and the amount of benefits received. The PR was the designated "main food shopper" for the household. The specific question asking about SNAP participation states, "(Do you/Does anyone in your household) receive benefits from the SNAP program? This program used to be called food stamps. It puts money on a SNAP EBT card that you can use to buy food." This question (named SNAPNOWREPORT on the FoodAPS data files) does not specify a reference period, and only respondents who answered "yes" were further asked to provide dates of the last receipt as well as benefit amounts received. Respondents who answered "no" were then asked, "Have (you/anyone in your household) ever received benefits from the SNAP program?" Households who responded in the affirmative to this follow-up question were further asked, "Did (you/anyone in your household) receive SNAP benefits in the last 12 months?" Respondents who answered "yes" to both follow-up questions were also asked to provide a date of the last receipt.[5] For our indicator of reported SNAP participation (hereinafter "REPORT"), we consider all respondents who answered "yes" to be self-reported participants (including those who answered "no" to the first participation question but "yes" to both follow-up questions), and consider the time frame to reflect either current or recent participation.[6] In other words, REPORT captures participation within the previous 12 months, unless otherwise stated. In our view, a flexible time frame is reasonable, as our outcomes (particularly body mass index [BMI], which is a stock variable) may not respond immediately to changes in benefit receipt, while people who have recently become nonparticipants may still spend down previously accrued benefits during the reference period.[7]

The FoodAPS contains two distinct administrative measures of SNAP participation. The first is from state caseload files covering March 2012 to November 2012 ("ADMIN"). The second is from the EBT Anti-Fraud Locator EBT Retailer Transactions database ("ALERT").[8] The ALERT transaction data contain one recorded swipe of an EBT Card per user from April through December 2012. FoodAPS is the only nationally representative survey that links reported SNAP participation to two administrative sources, thus making it particularly suitable for our purposes.

While such administrative records sound appealing, they have several limitations that likely lead to measurement error. ADMIN and ALERT participation measures contain various levels of missing data and do not always agree with each other. The quality and availability of the administrative data vary considerably across states. Households can fall into one of four state groups: (i) Group 1: one-to-one match was possible between ADMIN and ALERT data, because they both

[5] Sixty-six out of the 1461 people who answered "yes" to the first participation question subsequently reported date of the last receipt outside of the previous 31 days. Also, 8 out of 171 people who answered "no" to the first participation question but "yes" to both follow-up participation questions reported date of the last receipt within the previous 31 days. These reported dates of the last receipt reflect the ambiguity about whether the respondents perceived the initial participation question as relating to current or recent receipt of SNAP. Our conclusions remain similar if we code these individuals as nonparticipants.

[6] We thank John Kirlin for suggesting this modification to the original SNAPNOWREPORT via E-mail correspondence.

[7] Nonetheless, our main conclusions are unchanged if we count receipt as measuring participation only if the date of last receipt can be determined to be within 30 days of the initial interview for self-reported as well as both administrative measures. Our online appendix containing these additional results is available at https://sites.google.com/site/cjcourtemanche/research/online-appendices.

[8] The EBT ALERT database is ALERT system of the FNS of the USDA designed to help detect signs of abuse, fraud, and waste in the SNAP program. Each record of the EBT ALERT data represents one swipe of the EBT card and includes such variables as information on the state, store ID, EBT account number, date/time of the event, and purchase amount.

contain the same case identifiers (CASEIDs) (13 states), (ii) Group 2: either the CASEIDs in the ALERT data are scrambled or they are different in the ALERT and caseload data (eight states), (iii) Group 3: CASEIDs are different in the caseload and ALERT data, and the former does not include benefit disbursement dates (two states), and (iv) Group 4: the state did not provide SNAP enrollment data, and thus, all matches were to ALERT data (five states). State Group 1 presumably has the highest quality of linkage, since households were matched to both administrative sources. Specifically, if a household *first* matched probabilistically to caseload data, then a deterministic (one-to-one) match was possible to the ALERT data using CASEIDs. Thus, it is reasonable to presume that the quality of the administrative linkage would be highest in the 13 states in State Group 1. In State Groups 2 and 3, households were probabilistically matched separately to caseload and ALERT data as a one-to-one match was not possible, because the CASEIDs were different in both data sets. The difference between the two groups is that the caseload data for State Group 3 households do not include benefit disbursement dates. For State Group 4, only an ALERT match was possible.

Another source of measurement error is that matching from the FoodAPS to administrative SNAP records was probabilistic. All the matches to ADMIN data were based on first name, last name, phone number, and house address (including apartment number), and links were considered "certain matches" if the associated matching score exceeded a predetermined threshold.[9] The linkage to the ALERT data was similarly probabilistic, except in the state Group 1 described above. The quirks of probabilistic matching would suggest unknown degrees of error in the administrative measures of participation in all states.

Additionally, the ADMIN and ALERT data may contradict each other because of discrepancies in timing. In the ADMIN data, participation is in most cases defined based on current enrollment status during the interview week. However, in the two states in Group 3 mentioned above, exact dates are not available; thus, their current participation status was conditional on the results of the EBT ALERT linkage. For instance, in a few cases, an individual is considered a current participant if they matched at any point during the 9-month data availability window and also matched to the EBT ALERT, with the date of the last receipt (per ALERT) within 36 days of the end of the survey week.[10] Some former and future participants will therefore incorrectly be coded as current participants. The same logic applies to the EBT ALERT data. In the ALERT data, an individual is coded as a participant if she had an EBT card transaction during the survey week and matched to the EBT ALERT data. SNAP participants who did not use the EBT card that week—for instance, because they stocked up on groceries the previous week, or because their monthly benefits already ran out (food stamp cycling)—were coded as nonparticipants if they were also current nonparticipants per ADMIN.[11]

[9] The probabilistic matching was implemented using the LinkageWIZ record linkage software and resulted in a Cartesian join of each surveyed household with all SNAP enrollment record (or EBT ALERT). The contractors determined a prespecified score above which to classify a match as "certain." FoodAPS does not contain the raw matching scores.

[10] FoodAPS's combined measure of current SNAP participation based on the two administrative linkages is SNAPNOWADMIN. FoodAPS's SNAPNOWHH builds on SNAPNOWADMIN by imputing missing data (e.g., for nonconsenting households) using the self-report, thus effectively creating a participation measure based on all three measures—ADMIN, ALERT, and REPORT. We return to the task of constructing a preferred participation measure based on all three variables in section 3.

[11] In the remainder, majority of cases in the two states whose current SNAP participation cannot be determined based on EBT ALERT matching (conditional on ADMIN) or ADMIN (conditional on ALERT) due to missing information or nonmatches, their current SNAP participation is coded as "no match" in SNAPNOWADMIN.

Another source of discrepancy regarding timing is that, while the ADMIN variable considers matches to represent current participation if the date of the last receipt is within 32 days of the end of the survey week, the ALERT variable uses 36 days of the end of the survey week. This may be related to the fact that the ALERT data do not have variables indicating the exact timing of deposits into the SNAP accounts. The ALERT issuance dates are approximate, because issuance dates are determined by noting increments in the last SNAP balance between swipes. Thus, households classified as current recipients per ALERT may show up as current nonrecipients per the ADMIN variable due to the shorter window used by the latter.[12]

Finally, another issue with the ALERT data is that matches were not always attempted. For instance, no match was attempted if the respondent did not self-report SNAP receipt or any EBT-type payments. While most such individuals are likely true nonparticipants, this might not be the case for all of them. The high prevalence of false negatives reported in the literature tells us that some people incorrectly report not receiving SNAP, and it seems plausible that some of these same people would also not voluntarily disclose using an EBT card for any of their purchases. A match to EBT records was also not attempted if the individual reported SNAP participation, but did not make a purchase at an SNAP-eligible store during the survey week. While some of these individuals may be genuine false positives, others might have simply not gone to the grocery store that week.

These issues create substantial ambiguity about the "correct" ways to code the administrative variables that we will explore in more detail in the following section. For now, we define the baseline versions of these two measures to correspond with REPORT as follows.[13] We set "ADMIN Baseline" = 1 if there is a successful match to caseload records, even if the date of the match is outside of the previous month or missing. The rationale for the flexible timing mirrors that discussed above for the self-reported measure. We set "ADMIN Baseline" = 0 for individuals who did not match to the caseload records, and leave the variable missing for those in states that did not provide caseload records. For "ALERT Baseline," we assign a value of 1 if there is a confirmed match (again, regardless of whether the match occurs during the survey month) and 0 if a match was attempted but unsuccessful. If no match was attempted, we set "ALERT Baseline" to missing.

Turning to a discussion of the other variables used in our analyses, our first two dependent variables relate to food insecurity. These come from the ten-question household food security questionnaire included in FoodAPS based on USDA's 30-day Food Security Scale.[14] The specific outcomes are a dummy for whether the household has low food security (defined as having affirmative responses to three to five questions) and a dummy for whether the household has very low food security (six or more affirmative responses).

The next several dependent variables relate to body weight. The FoodAPS contains self-reported height and weight for the household responder. We use this information to create three outcomes: BMI and indicators for obese (BMI $\geq$ 30) and severely obese (BMI $\geq$ 35).[15] Dichotomous variables are often used in addition to continuous BMI in the obesity literature, since health is not monotonically decreasing in weight. Weight gain generally improves health at low levels of BMI, and

[12] This issue ceases to be relevant if we ignore the dates of the last receipt in classifying matches to the administrative sources, as we do in our baseline classifications discussed below.

[13] Unless otherwise stated in the rest of the paper, "ADMIN" and "ALERT" refer to "ADMIN Baseline" and "ALERT Baseline," respectively.

[14] Please see the Appendix for the list of question on the 10-question household food security question.

[15] Body mass index is defined as weight in kilograms divided by height in meters squared.

the large increase in mortality risk from excess weight does not begin until around the severe obesity threshold (Courtemanche et al. 2016). The health implications of any impacts of SNAP would depend on the portion of the BMI distribution in which the effects are strongest (i.e., the health implications of SNAP's effects would potentially be more substantial if they are stronger on severe obesity).

The final dependent variable relates to food purchases. Following prior studies such as Volpe, Okrent, and Leibtag (2013), we use a summary measure of the healthfulness of food purchases called the HEI. The HEI-2010, designed by the USDA, aims to capture the degree of adherence to dietary guidelines. We use the total HEI-2010 scores for all items for all the entire survey week for each household.[16] The HEI score is made up of 12 components that sum up to a maximum score of 100. This HEI variable is computed by FoodAPS staff and available as a linkable auxiliary data set.

The FoodAPS also contains a number of variables that we use as controls. These include dummy variables for gender, educational attainment (dummy variables for having less than high school diploma, high school diploma but no college education, and some college education, with college degree or higher being the omitted base category), race/ethnicity (non-Hispanic black and non-Hispanic white, with other being the base category), marital status (married and formerly married, with never married as the base category), whether any individuals under five years old or at least 65 years old are present in the household, whether the respondent worked last week, whether the household lives in rural census tract, and whether the household's primary food store is SNAP-authorized. Continuous controls include respondent's age, household size, number of children, household monthly gross total income, and straight-line distance from household's residence to its primary food store (in miles).

Our "main sample" is subject to four restrictions. First, we include only households in which the PR is at least 18 years old. Second, we drop households with missing values of any variables. Next, we follow Mykerezi and Mills (2010) and Almada, McCarthy, and Tchernis (2016) and drop those with incomes over 250% FPL. The final step is to exclude 122 households who did not provide consent for administrative verification. The resulting sample contains 2108 households. The sample sizes in some of the sensitivity checks will vary, though, as we will experiment with different ways to handle ambiguous cases in the administrative SNAP variables.

Table 1 presents means and standard deviations for our main sample. The SNAP participation rate is 32% using the self-report compared with 29% with ADMIN and 30% with ALERT. The correlations between the three measures are 0.782 for REPORT and ADMIN, 0.792 for REPORT and ALERT, and 0.847 for ADMIN and ALERT. In other words, the two administrative measures exhibit almost as much disagreement with each other as either of them does with the self-reported measure. FoodAPS's PRs have an average BMI of 28.81, while 38% are obese and 16% are severely obese. About 20% of FoodAPS households are food insecure (low food security), while 13% experience very low food security. In terms of compliance with the U.S. Dietary Guidelines for Americans, FoodAPS households have an average HEI score of 50.56 out of a maximum score of 100; higher HEI scores indicate greater conformity with recommended dietary guidelines.

The PRs are on average, about 49 years old with a household size of about 2.56. Also, almost 71% of the PRs are female, 31% are married, 16% are black, 71% are white, and about 38% report

[16] Further information on HEI scores can be found at http://epi.grants.cancer.gov/hei.

Table 1. Summary Statistics

Variable	Mean (Standard Error)
SNAP participation	
Self-reported (REPORT)	0.32 (0.02)
Administrative from caseload data (ADMIN)	0.29 (0.02)
Administrative from EBT transactions (ALERT)	0.30 (0.02)
Dependent variables	
Low food security	0.20 (0.02)
Very low food security	0.13 (0.01)
Total 2010 HEI score	50.56 (0.58)
Body mass index	28.81 (0.25)
Obese	0.38 (0.02)
Severely obese	0.16 (0.01)
Control variables	
Age (years)	49.62 (0.98)
Female	0.71 (0.02)
Black	0.16 (0.03)
White	0.71 (0.04)
Other race (non-black, non-white)	0.13 (0.02)
Married	0.31 (0.02)
Formerly married	0.43 (0.02)
Household size	2.56 (0.10)
Number of children	0.93 (0.07)
Rural tract	0.33 (0.06)
Less than high school education	0.19 (0.02)
High school graduate	0.34 (0.02)
Some college education	0.21 (0.01)
College degree or higher	0.26 (0.02)
Worked last week	0.38 (0.03)
Gross monthly family income ($1000s)	1.86 (0.06)
Child less than five years present in HH	0.61 (0.02)
Elderly at least 65 years present in HH	0.28 (0.03)
Never married	0.26 (0.02)
Miles from residence to primary food store	3.15 (0.34)
Primary food store is SNAP-authorized	0.98 (0.00)

Notes: Statistics are from main analysis sample of 2108 observations. Observations are weighted to account for the complex sampling design of FoodAPS.

having worked last week. FoodAPS PRs have a gross monthly income of about $1860 and live in households with 26% holding college degrees or higher, 21% with some college education, 34% with a high school diploma, and 19% with less than high school diploma. Finally, 33% of FoodAPS household live in a rural census tract, 61% have children at most five years of age, and 28% have elderly at least 65 years present.

3. Sensitivity of Participation and Misreporting Rates

This section examines the sensitivity of SNAP participation and misreporting rates along two dimensions. The first type of sensitivity concerns different classification choices for the potentially ambiguous cases when continuing to use ADMIN and ALERT separately. The second is with respect to different approaches to combining ADMIN and ALERT into a single, "true" participation measure.

Different Classification Choices for ADMIN and ALERT Separately

The discussion of the SNAP variables in section 2 revealed several challenges when coding ADMIN and ALERT variables. Tables 2 and 3 categorize the potential values of these variables to elucidate the specific sources of ambiguity. The tables also report the number of households in each category, how they are classified in the "baseline" classification used in section 2, and other defensible ways in which they could be classified by researchers. The latter is given in the column names "Alternate 1" through "Alternate 4," where in each column, the specific categorization that differs from the baseline choice is in bold.

Focusing first on the ADMIN variable in Table 2, we see that there are five different broad categories a household can fall into. First is the straightforward case where the state did not make caseload records available, and therefore, the ADMIN variable is clearly missing. Category 2 consists of the clearest participants: those who matched to caseload records within the 32 days before and including the survey week.

The third category contains households who did not match to caseload records despite a match being attempted. Most such households are likely true nonparticipants, and we label them as such in the baseline classification system. However, someone could be a true participant but did not match due to, for instance, name misspellings, changes in household identifying information such as addresses and phone numbers, the small number of matching variables, or simply due to the nature of the probabilistic matching. Some matches to the caseload data not deemed to be automatically "certain" were manually reviewed, helping to alleviate these concerns. Nonetheless, a sensitivity analysis is warranted. In Alternate Classification 1, we classify households in Category 3 as participants if they are statistically likely to be in the caseload records based on their values of the control variables listed in section 2. Specifically, using the baseline classifications, we run a probit regression of ADMIN on the control variables, assigning anyone with a predicted value of greater than 0.5 (i.e., greater than 50% predicted probability of participation) as a participant in the alternative classification. The idea is that, based on the available information, these households are the most likely to be erroneously unmatched.

Category 4 contains households that matched to caseload records but with a date outside the 32-day window. There are two layers to the classification challenge here. First, assuming that the dates are correct, does it make sense to narrowly define a household as a participant if and only if participation is in the current month? As discussed in section 2, the intention of our baseline classifications is to measure either current or recent participation, since recent participation could still plausibly influence the outcomes. Therefore, the most natural baseline classification of these households is as participants. However, since some researchers may choose to measure participation more strictly as just participation in the current month, Alternate Classification 2 labels households in Category 4 as nonparticipants. The second layer of the classification challenge is whether the dates revealed by the caseload records are actually correct. A sizeable number of households show matches in both the months immediately before and after the survey month, but not in the survey month.[17] This, coupled with the fact that the probabilistic matching could introduce errors in and of itself, creates sufficient ambiguity in the dates from the caseload records to warrant another sensitivity analyses, Alternate Classification 3, that treat households in Category 4 as missing participation status.

Category 5 consists of those who matched to caseload records, but the dates of SNAP receipt are not available. Again, since our goal with the baseline classifications is to capture current or recent participation, the lack of an exact date is not especially problematic, so we consider these

[17] We are grateful to Bruce Meyer and Nikolas Mittag for this observation.

Table 2. Possible Classifications for Administrative Participation Measure from Caseload Data (ADMIN)

Category	Description	N	Baseline Classification	Alternate 1	Alternate 2	Alternate 3	Alternate 4
1	ADMIN data not available from state	448	Missing	Missing	Missing	Missing	Missing
2	Match confirms participation *within 32 days* of the survey week	763	Participant	Participant	Participant	Participant	Participant
3	Match to caseload data attempted but did not meet threshold for certainty	1268	Nonparticipant	**Participant if probability > 0.5**	Nonparticipant	Nonparticipant	Nonparticipant
4	Match confirms participation outside of survey week	134	Participant	Participant	**Nonparticipant**	**Missing**	Participant
5	Match confirms participation but dates not available	175	Participant	Participant	Participant	Participant	**Missing**

Notes: Based on the main sample augmented with observations with missing ADMIN or ALERT but not any other variable.

Table 3. Possible Classifications for Administrative Participation Measure from EBT Transactions (ALERT)

Category	Description	N	Baseline Classification	Alternate 1	Alternate 2	Alternate 3	Alternate 4
1	No acquisitions available for matching and no match to ADMIN to provide CASEID	574	Missing	Missing	Missing	Missing	**Nonparticipant**
2	Match confirms participation *within 36 days* of the survey week	961	Participant	Participant	Participant	Participant	Participant
3	Match to ALERT data attempted but did not meet threshold for certainty	1174	Nonparticipant	**Participant if probability > 0.5**	Nonparticipant	Nonparticipant	Nonparticipant
4	Match confirms participation outside of survey week	79	Participant	Participant	**Nonparticipant**	**Missing**	Participant

Notes: Based on the main sample augmented with observations with missing ADMIN or ALERT but not any other variable.

households to be participants. However, researchers who wish to define households as participants if and only if they participated in the month leading up to the survey would have no way to do so without exact match dates. Therefore, Alternate Classification 4 treats households in Category 5 as having missing participation information.

Turning to the ALERT variable in Table 3, households can fall into four different categories. Category 1 contains those for whom no match was attempted, or who did not match to the ADMIN data to provide a CASEID that would permit a deterministic match to the EBT ALERT database. We code these individuals as missing in our baseline classification. However, since most individuals for whom no match was attempted are likely true nonparticipants (as discussed in section 2), we code households in Category 1 as nonparticipants in Alternate Classification 4.

ALERT Category 2 indicates a match to the EBT ALERT database was successful with date of last receipt within the 36-day window. Some Category 2 households may have been determined manually when a single FoodAPS transaction matched to multiple ALERT transactions. Thus, the final account number assigned to the FoodAPS transaction may result in an erroneous Category 2 determination. However, in our judgment, such misclassification is unlikely to happen in more than a few cases, so we do not consider alternative classifications of ALERT Category 2.

Category 3 contains those for whom a match was attempted but not successful, suggesting nonparticipation. Since the ALERT matches were probabilistic based on STOREID, amount, and date, it is conceivable that some of the Category 3 households may have failed to match due to reasons such as mistakes in the reported amounts and dates or even the number of matching variables. Similar to the ADMIN data, Alternate Classification 1 treats ALERT Category 2 members as participants if they have a predicted probability of participation of over 50% based on a probit regression of ALERT on the control variables.

ALERT Category 4 households are similar to ADMIN Category 4 above; they matched to the EBT ALERT database, but the associated date of the last receipt is outside of the 36-day window. For similar reasons as mentioned above, we initially consider these households to be true participants, but Alternate Classification 2 considers them to be nonparticipants, while Alternate Classification 3 treats them as missing.

Table 4 presents estimated participation and error rates for the various ADMIN and ALERT classification choices discussed above. Panel A uses the baseline classifications and the main sample that drops observations with missing values of either ADMIN or ALERT (under their baseline classifications) or any of the control variables. This enables an "apples-to-apples" comparison of the differences caused by ADMIN versus ALERT within the same sample. Panel B allows the sample size to vary depending on the treatment of missing data. The row labeled "ADMIN Baseline" in Panel B adds back in the observations with a valid value of that variable but missing "ALERT Baseline," and vice versa. The rows for the alternate classifications can either contain more or fewer observations depending on the relative stringency of the criteria for handling ambiguous cases. For instance, the sample is much larger for "ALERT Alternate 4" than "ALERT Baseline," because the former treats the large number of households for whom no match was attempted (Category 1) as nonparticipants, whereas the latter considers them missing.

Panel A shows that the participation and misreporting rates in the main sample are broadly similar using the baseline constructions of ADMIN and ALERT. The estimated SNAP participation rate is 29% using ADMIN compared with 30% using ALERT. The false negative rates are 11.65% using ADMIN and 11.46% using ALERT, while the false positive rates are 8.39% using ADMIN and 7.83% using ALERT. Interestingly, for both participation measures, the prevalence of false

Table 4. Estimated Participation and Misreporting Rates under Different Approaches to Using ADMIN and ALERT Separately

Determination of Final ADMIN and ALERT Status	Sample Size	Participation Rate (%)	False Negative Rate (%)	False Positive Rate (%)
Panel A: main sample				
ADMIN baseline	2108	29.00	11.65	8.39
ALERT baseline	2108	30.00	11.46	7.83
Panel B: varying samples				
ADMIN baseline	2340	28.59	12.28	8.08
ADMIN alternate 1	2340	41.44	32.31	4.78
ADMIN alternate 2	2340	23.75	11.70	6.75
ADMIN alternate 3	2206	24.96	6.83	8.08
ADMIN alternate 4	2165	25.55	13.23	8.08
ALERT baseline	2214	33.51	10.49	8.64
ALERT alternate 1	2214	45.07	28.01	6.00
ALERT alternate 2	2214	29.73	10.89	7.04
ALERT alternate 3	2135	30.89	6.55	8.64
ALERT alternate 4	2788	24.14	10.49	12.17

Notes: Observations are weighted to account for the complex sampling design of FoodAPS.

negatives is substantially lower than previously reported by studies using more traditional survey data sets (Mittag, 2013; Meyer, Mittag, and George 2018). One possible explanation is that FoodAPS households were asked to consent to having their responses verified. Even though all but 122 households gave consent, it is reasonable to presume that merely informing respondents about data verification and asking for consent may elicit more truthful responses and partly account for the lower estimated false negatives. While we might expect consenting to data verification to lead to more truthful responses among nonparticipants as well, this does not appear to be the case, as estimated false positives in the FoodAPS are much higher than typically found. Conceivably, individuals who were unsure whether or not their household received SNAP—perhaps because they actually received some other program—might have been more inclined to report affirmatively because of the looming verification.

Panel B documents considerable variation in participation and misreporting rates depending on the particular classification decisions for ADMIN and ALERT.[18] The estimated participation rates for ADMIN vary from 23.75% (Alternate 2) to 41.44% for Alternate 4, for a spread of 17.69 percentage points, or 74% of the lower end of the range. The ALERT classification choices lead to even more variability, ranging from about 24.14% (Alternate 4) to 45.07% (Alternate 1), for a spread of 20.93 percentage points, or 87%. Not surprisingly, the highest participation rates are for the classification rule (Alternate 1 for both ADMIN and ALERT) where households with a predicted probability of participation over 50% are flipped from nonparticipants to participants, mechanically increasing the participation rate. Excluding this classification, the ranges tighten to 23.75–28.59% for ADMIN and 24.14–33.51% for ALERT.

The sensitivity in false negative rates is even more striking. For ADMIN, the estimated false negative rates vary from 6.83% (Alternate 3) to 32.31% (Alternate 1), meaning that judgment calls

[18] Note that for estimating reporting errors when using Alternate Classification 2, which recodes households as nonparticipants if the match occurred outside of a 32-day time frame for ADMIN and 36 days for ALERT, we temporarily recode the self-reported measure accordingly. In other words, when computing the misreporting rates using ADMIN (ALERT) Alternate 2, the comparisons are with self-reported participation in the past 32 (36) days.

about classifications could potentially cause this rate to vary by nearly 400%. The false negative rates using ALERT range from 6.55% (Alternate 3) to 28.01% (Alternate 1), again a spread of over 300%. The high ends of these ranges again come from the classification that flips the administrative measures to participation for individuals with high predicted probabilities. Excluding this classification (Alternate 1), the ranges narrow to 6.83–13.23% for ADMIN and 6.55–10.89% for ALERT. While these ranges are still sizeable in percentage terms, they are all well below the rate of false negatives found by prior studies.

Comparatively, there is less variation in the false positive rates. For ADMIN, the range is from a low of 4.78% (Alternate 1) to 8.08% in each of three cases. For ALERT, the estimated false positives range from 6% (Alternate 1) to 12.17% (Alternate 4). It is not surprising that Alternate 1 provides the low end of both ranges, since that is consistent with the flipping of some nonmatches to participants. However, the false positive rates are much less sensitive than the false negative rates to the use of this classification. This implies that most individuals who did not match but had a high probability of participation based on the covariates did not self-report participation. Despite the modest variation in these estimates, the finding that the false positive rate is higher in the FoodAPS than other data sets is nonetheless robust to all classifications.

Different Classification Choices when Combining ADMIN and ALERT

This section introduces several approaches or ad hoc rules to consolidate the two administrative participation measures into a single "true" participation variable and then evaluates how these rules influence the estimated rates of SNAP participation and misreporting. For the rest of this section, ADMIN and ALERT refer to the baseline classification choices as described in Tables 2 and 3, respectively. We develop five decision rules to combine the administrative participation variables as follows:

1. Always use ADMIN unless missing. For households missing ADMIN data, their participation status is set to ALERT.
2. Always use ALERT unless missing. For households missing ALERT data, their participation status is set to ADMIN.
3. Drop if Disagreement: This rule sets the "true" participation variable to equal to both ADMIN and ALERT, *only* if they agree (i.e., if ADMIN = ALERT = i, i = 0, 1). When they disagree or if either of them is missing, the "true" variable is set to missing. This conservative approach will minimize errant classification but at a substantial cost to sample size.
4. More weight to matches: This rule is similar to Rule 3 as it uses both ADMIN and ALERT if they agree. However, when they disagree, we set the "true" status to participation ("1"), unless either is missing in which case the "true" status is set to the value of the nonmissing variable. In other words, this rule treats households as "true" participants if at least ADMIN or ALERT confirms participation. Otherwise, the household is considered a nonparticipant unless both are missing. This is the same as the rule used by the USDA's ERS in their measure of current participation based solely on the results of the two administrative matches (the variable SNAPNOWADMIN in the data set).
5. More weight to nonmatches: This rule is the reverse of Rule 4. When ADMIN and ALERT disagree, we set the "true" status to nonparticipation ("0"), unless either is missing in which case the "true" status is set to the value of the nonmissing variable. In other words, this rule treats

Table 5. Estimated Participation and Misreporting Rates under Different Approaches to Combining ADMIN and ALERT into a "True" Participation Measure

Decision Rule when ADMIN and ALERT Differ	Sample Size	Particip-ation Rate (%)	False Negative Rate (%)	False Positive Rate (%)
Rule 1: Always use ADMIN unless missing	2446	31.95	11.25	8.80
Rule 2: Always use ALERT unless missing	2446	32.30	11.10	8.31
Rule 3: Drop if disagreement	1898	28.25	10.98	4.53
Rule 4: More weight to matches	2446	34.81	11.57	5.46
Rule 5: More weight to nonmatches	2446	29.44	10.71	11.41

Notes: Observations are weighted to account for the complex sampling design of FoodAPS.

households as "true" nonparticipants if at least ADMIN or ALERT confirms nonparticipation. Otherwise, the household is considered a participant unless both are missing.

Table 5 presents estimates of participation, false negative, and false positive rates under each of the above decision rules. Note that each of the above decision rules leads to different measures of "true" participation, which we compare to self-reported participation to estimate reporting error rates. The estimated participation rates range from 28.25% (Rule 3) to 34.81% (Rule 4). This is a spread of 6.56, which represents 23% of the low end of the range. The estimated false negative rates range from 10.71% (Rule 5) to 11.57% (Rule 4), for a spread of 0.86 percentage points, or 8%. The false positive rates vary more substantially, from 4.53% (Rule 3) to 11.41% (Rule 5), for a spread of over 150%.

Some patterns also emerge. First, as expected, giving the benefit of the doubt to matches (Rule 4) leads to a relatively high estimated participation rate, and keeps the rate of false positives low but at the expense of a high rate of false negatives. The reverse is true when we give the benefit of the doubt to nonmatches (Rule 5). Perhaps, more surprising is that dropping cases where there is any ambiguity (ADMIN and ALERT disagree or either are missing; Rule 3) results in the lowest estimated participation rate, lowest rate of false positives, and second-lowest rate of false negatives. In other words, once we restrict the sample to households for whom the administrative measures are likely quite accurate, we see less disagreement with self-reported participation. There is a particularly large reduction in the number of cases in which the respondent reports participation but the administrative data disagree. This implies that some of the estimated misreporting observed under other decision rules is not actually misreporting at all, but instead reflective of flaws in the administrative variables. It is also noteworthy that the sample shrinks so much—2446 to 1898, or 29%—under Rule 3, underscoring that the amount of ambiguity, and therefore scope for error, in the administrative measures is substantial.

Preferred Approach to Combining REPORT, ADMIN, and ALERT

Given the ambiguity and sensitivity documented above, it is reasonable to ask whether linked administrative data can still be used to obtain insights beyond what could be done with self-reported information alone. The conservative Rule 3 should lead to a very accurate participation measure but at the cost of discarding nearly a third of the sample, which creates concerns about endogenous sample selection and external validity. The other decision rules avoid such a large reduction in sample size but at the expense of accuracy. The goal of this section is to implement a more detailed strategy for combining ADMIN and ALERT that utilize self-reports to help resolve ambiguous

Table 6. Extent of Disagreement among SNAP Participation Variables

Category	REPORT	ADMIN Baseline	ALERT Baseline	Observations
A	0	0	0	952
E	0	0	1	11
B	0	0	–	144
E	0	1	0	21
C	0	1	1	77
E	0	1	–	12
B	0	–	0	1
E	0	–	1	9
D	0	–	–	261
C	1	0	0	74
E	1	0	1	69
E	1	0	–	18
E	1	1	0	109
A	1	1	1	795
B	1	1	–	58
E	1	–	0	17
B	1	–	1	79
D	1	–	–	81
	Total			2788

Notes: Frequencies are based on the main sample augmented with observations with missing ADMIN or ALERT but not any other variable. A table cell with "1" denotes participant, "0" denotes nonparticipant, and "period" denotes missing data. The numbers of households for each category are as follows: A = 1747, B = 282, C = 151, D = 342, and E = 266.

cases, with the goal of leveraging insights from all three measures to produce reliable estimates while preserving sample size.[19]

To motivate this approach, Table 6 presents information about the extent of disagreement among the three measures as well as the extent of missing data in each variable.[20] Also, the first column reports how we classify disagreements into various categories for the purpose of developing our new SNAP participation variable, which we refer to as "Preferred SNAP." There is about 63% agreement among all three measures (i.e., all three variables either indicate participation or non-participation), which we label as Category A. In Category B, making up about 10% of households, any two of the three measures agree, while the third is missing. Category C respondents, which account for 5%, have both administrative measures agreeing but in conflict with the self-report. Households with only the self-reported participation variable who are missing both administrative measures (Category D) make up 12%, while the remaining 10% of respondents are lumped into miscellaneous types of disagreement in Category E.

The new, "preferred" measure of SNAP participation combines information from Categories A, B, and C and sets to missing observations in Categories D and E. For Category A, all three variables are in agreement, so we are comfortable setting the "true" participation variable equal to the associated value. For Category B, we are also comfortable making a determination, since, although one variable is missing, the other two agree. For Category C, we consider the self-reported participation value to be

[19] This approach is similar in spirit to the FoodAPS' SNAPNOWHH variable discussed previously in footnotes 11 and 12. However, our method is more conservative in that it demands a higher level of agreement across the SNAP measures before considering the "true" value of participation nonmissing.

[20] Note that there is essentially no missing data for self-reported participation: only 4 out of 4704 consenting households in the FoodAPS are missing this variable.

erroneous, since it opposes both administrative measures, and there is no particular reason to expect errors in the administrative variables to be correlated with each other. This maintains the preference for administrative records if the information from those records appears to be reliable. Next, those in Category D have nonmissing self-reported participation but are missing both administrative measures. We code their participation as missing given the established concerns in the literature about relying only on self-reports. Finally, we also set the participation status of respondents in Category E to missing. There are three types of Category E households: ADMIN and ALERT are nonmissing but disagree, ADMIN and REPORT disagree while ALERT is missing, and ALERT and REPORT disagree with ADMIN missing. In such cases of explicit disagreement, a determination cannot be reached without establishing a rank ordering among the measures.

Ultimately, our preferred measure is nonmissing for the entire main sample of 2108 respondents. Relative to the sample sizes using the various decision rules in Table 5, this is less than the 2446 observations obtained using decision rules that force an outcome even in ambiguous cases, but larger than the 1898 observations obtained under the conservative Rule 3. The estimated participation rate using the preferred measure is 30.92%, which is slightly higher than those obtained using ADMIN and ALERT separately (Panel A in Table 4) but well within the ranges established by the various sensitivity checks in Tables 4 and 5. The preferred SNAP participation measure leads to relatively low estimated rates of false negatives (8.53%) and false positives (3.99%), but this is partly by construction, since the self-reported value is factored into the coding process.

4. Econometric Analyses and Results

We next turn to our regression estimates of the associations of SNAP with food insecurity, weight outcomes, and dietary healthfulness. This section's goal is to illustrate the sensitivity of these estimates to the assumptions, introduced in the previous section, about how to code ADMIN and ALERT separately as well as how to assign "true" participation in cases of disagreement between them. We do not attempt to address the endogeneity of participation, because doing so with a single cross section of data such as the FoodAPS is daunting, and our focus here is to examine measurement issues rather than identify causal effects.[21] Negative selection into SNAP is well-documented in the literature, so our OLS estimates will likely be biased in the direction of less favorable outcomes. This means, for instance, that SNAP will appear to be less beneficial to the recipient either understating the benefits of SNAP or potentially even finding that participation is associated with greater food insecurity, higher BMI and obesity rates, and less healthy diets, or at least be attenuated, measurement issues aside.

Our regressions take the form

$$y_i = \beta_0 + \beta_1 \mathrm{SNAP}_i + \boldsymbol{\beta_2 X_i} + \varepsilon_i, \qquad (4.1)$$

[21] One approach to addressing the nonrandom selection into SNAP is relying on instrumental variables. Unfortunately, the usual state-level administrative policies used to study programs such as SNAP and WIC are not likely to be valid instruments with cross-sectional data as one would have to rely on variation across states. These program rules may be correlated with other state-level characteristics unrelated to participation decisions (see, for example, Bitler, Currie, and Scholz 2003; Bitler and Currie 2005). As shown by Almada, McCarthy, and Tchernis (2016), nonclassical measurement error can substantially alter IV estimates and cause them to fall outside of nonparametric upper bounds. Measuring SNAP participation as accurately as possible, therefore, would arguably be even more critical in an IV context than in the OLS context shown here.

where y_i is the outcome variable for individual/household i (separate regressions for each of the outcomes discussed in section 2), SNAP_i is an indicator of SNAP participation (separate regressions for each decision rule from section 3), $\boldsymbol{X_i}$ is a vector of the control variables from section 2, and ε_{is} is the error term.

Measurement error in a binary variable is necessarily nonclassical, so one cannot simply assume $\hat{\beta}_1$ to be biased toward zero (Kreider 2010; Kreider et al. 2012; Nguimkeu, Denteh, and Tchernis 2019). Measurement error in SNAP participation could potentially even lead the OLS estimator to be wrongly signed (Nguimkeu, Denteh, and Tchernis 2019). It might be reasonable to suspect that some of the inconsistencies among the administrative measures, such as the inability to match names with sufficient certainty, are as good as random. However, other inconsistencies, such as appearing in the caseload records but not using an EBT card in the past 30 days, arise from personal choices and may, therefore, be correlated with the error term, hence leading to endogenous misclassification.[22]

Common Sample

We begin our presentation of the regression results with Table 7, which uses the main sample and compares OLS estimates (linear probability model if the outcome is binary) using REPORT, the baseline version of ADMIN (as described in Table 2), and the baseline version of ALERT (as described in Table 3). Similar to Panel A of Table 4, the purpose here is to use a common sample to provide an apples-to-apples comparison of the results across the three measures. The first key result is that the results are qualitatively similar regardless of the SNAP participation measure used. SNAP participation is consistently associated with worse values of all six outcomes. Estimates for food insecurity and BMI are significant at the 1% level for all three SNAP measures, while those for very low food security are never significant. Mild discrepancies are observed for HEI and obesity, as two of the estimates are significant at the 1% level, while the third (using ALERT for HEI and ADMIN for obesity) is significant at the 5% level. For severe obesity, the estimates for REPORT and ADMIN are significant at the 5% level and 1% level, respectively, though the estimate for ALERT is insignificant.

The magnitudes are somewhat sensitive to the choice of SNAP measure. The associations between SNAP and food insecurity range from 6 to 7 percentage points, for a 16.67% spread. The estimates for very low food security vary between 2 and 2.7 percentage points, or 35%. SNAP reduces HEI by between 1.3 and 2 units, for a sizeable 54% difference. The results for BMI only vary from 1.05 to 1.17 units, or 11%, but greater sensitivity is observed for the dichotomous weight outcomes. The estimates for Pr(Obese) and Pr(Severely Obese) range from 5.7 to 7.9 and 2.1 to 3.9 percentage points, respectively, for spreads of 39 and 86%. Note also that the pattern of results is inconsistent with simple attenuation bias, in which case we would expect the magnitudes to be larger using the administrative SNAP measures than the self-report. For three of the outcomes, the magnitudes are actually largest using self-reported participation, and in only one case is the magnitude using self-reported participation the smallest. This is consistent with the reporting error being nonclassical (which can yield an expansion bias), but is also consistent with the administrative measures not being any more reliable than the self-report (i.e., there is some attenuation bias regardless of the measure used).

[22] Moreover, Bound, Brown, and Mathiowetz (2001) discuss the possibility that the measurement error may not be nondifferential, where measurement error is not independent of the outcomes of interest.

Table 7. Regression Results using Each Participation Measure Separately

	Food Insecurity	Very Low Food Security	Healthy Eating Index	Body Mass Index	Obese	Severely Obese
Self-reported	0.066*** (0.020)	0.027 (0.018)	−1.680*** (0.615)	1.045*** (0.343)	0.079*** (0.024)	0.039** (0.018)
ADMIN baseline	0.060*** (0.020)	0.022 (0.018)	−2.071*** (0.601)	1.166*** (0.344)	0.057** (0.023)	0.035* (0.018)
ALERT baseline	0.070*** (0.020)	0.020 (0.018)	−1.292** (0.605)	1.114*** (0.340)	0.061*** (0.023)	0.021 (0.019)

Notes: Statistics are from main sample of 2108 observations. Heteroscedasticity-robust standard errors are in parentheses. Observations are weighted to account for the complex sampling design of FoodAPS.
***Statistically significant at the 1% level.
**Statistically significant at the 5% level.
*Statistically significant at the 10% level.

With all that said, it is reasonable to ask whether the variability in the estimated SNAP coefficients is "real" in a statistical sense. To that end, for each outcome, we conducted postestimation *t*-tests of the equality of the coefficients using the self-reported SNAP measure versus ADMIN, the self-report versus ALERT, and ALERT versus ADMIN. In most cases, we do not reject the null hypothesis of equality at even the 10% significance level. The only exception is that the estimates for HEI using ADMIN and ALERT are statistically different at the 5% level. Interestingly, some sizeable differences in magnitudes—including those for obesity and severe obesity—are not statistically different.[23] This could be because of the relatively modest sample size of the FoodAPS, but regardless, we cannot rule out that at least some of the larger observed differences in magnitudes are merely attributable to sampling error.

Different Classification Choices for ADMIN and ALERT Separately

Table 8 presents similar OLS results using the self-reported participation variable and the different classification rules for coding ADMIN and ALERT separately, as described in Tables 2 and 3. As in Table 7, for all outcomes, the signs are robust across SNAP measures. However, there are some noteworthy differences in terms of significance levels and magnitudes. For instance, the association between self-reported SNAP participation and very low food security is a sizeable and statistically significant 4 percentage points. In contrast, with the exception of Alternate Classification 1, the same association is insignificant using any classification scheme for the administrative measures, and the magnitudes are much smaller: 0.2 to 0.3 percentage points. Recall from Table 7 that using the self-report also led to an insignificant result for very low food security for the common sample, meaning that much of the sensitivity observed here is actually from the difference in the sample (i.e., adding back in 27 to 680 observations with nonmissing self-reports but a missing value of one or both administrative measures depending on the administrative classification). This underscores the external validity concerns raised by the large amounts of missing data for the administrative variables.

The results for HEI and severe obesity are also relatively sensitive. For HEI, the estimates using REPORT and ADMIN are large (−1.5 to −1.88 units) and significant. However, with the exception of Alternate Classification 1, they shrink considerably (−0.72 to −1.2) using ALERT and are insignificant in some cases. Accordingly, the spread between the smallest and largest magnitude for HEI is over 190%. For severe obesity in Table 8, the estimates range from 3.5 to 5.2 percentage points using REPORT and ADMIN, but they shrink to 1.2–2.5 percentage points and become statistically insignificant using ALERT. The spread between the highest and lowest estimates for severe obesity is therefore a substantial 333%.

For the other outcomes, the sensitivity across the different constructions of the SNAP variable is less notable, though there are a few outliers. In almost all cases, SNAP is associated with a statistically significant (at the 1% level) increase in food insecurity of between 5.3 and 6.8 percentage points. The only exception is ALERT Alternate Classification 1, for which the estimate drops to 4.1 percentage points and is significant at just the 10% level. Similarly, for BMI, almost all the estimates are significant at the 1% level and within a narrow range of 1.13–1.38 units. The only exceptions are Alternate Classification 1 for both ADMIN and ALERT; the estimate is 0.91 units and significant at the 5% level for the former, and 0.72 units and significant at the 10% for the

[23] For instance, statistical tests for differences for Pr(Obese) between REPORT and the two administrative measures yield *p*-values of 0.21 and 0.33 for ADMIN and ALERT, respectively. The *p*-values for pairwise comparisons of estimated associations for all outcomes are available from the authors upon request.

Table 8. Regression Results using Each Participation Measure Separately

	Sample Size	Food Insecurity	Very Low Food Security	Healthy Eating Index	Body Mass Index	Obese	Severely Obese
Self-reported	2788	0.059*** (0.018)	0.040** (0.016)	−1.500*** (0.541)	1.272*** (0.302)	0.094*** (0.021)	0.043*** (0.016)
ADMIN baseline	2340	0.055*** (0.019)	0.019 (0.017)	−1.663*** (0.569)	1.328*** (0.325)	0.065*** (0.022)	0.048*** (0.017)
ADMIN alternate 1	2340	0.068*** (0.022)	0.033* (0.019)	−1.880*** (0.710)	0.912** (0.380)	0.043 (0.026)	0.035* (0.020)
ADMIN alternate 2	2340	0.056*** (0.020)	0.002 (0.018)	−1.484** (0.577)	1.202*** (0.333)	0.057*** (0.022)	0.045** (0.018)
ADMIN alternate 3	2206	0.060*** (0.020)	0.012 (0.018)	−1.640*** (0.597)	1.377*** (0.344)	0.066*** (0.023)	0.051*** (0.018)
ADMIN alternate 4	2165	0.053*** (0.020)	0.023 (0.018)	−1.417** (0.599)	1.249*** (0.341)	0.059** (0.023)	0.052*** (0.019)
ALERT baseline	2214	0.064*** (0.019)	0.023 (0.018)	−1.201** (0.591)	1.154*** (0.333)	0.064*** (0.023)	0.021 (0.018)
ALERT alternate 1	2214	0.041* (0.023)	0.058*** (0.020)	−2.080*** (0.738)	0.724* (0.398)	0.056** (0.027)	0.025 (0.021)
ALERT alternate 2	2214	0.062*** (0.020)	0.009 (0.018)	−0.907 (0.596)	1.159*** (0.335)	0.065*** (0.023)	0.022 (0.018)
ALERT alternate 3	2135	0.067*** (0.020)	0.016 (0.018)	−1.090* (0.608)	1.218*** (0.342)	0.068*** (0.023)	0.023 (0.019)
ALERT alternate 4	2788	0.052*** (0.018)	0.016 (0.016)	−0.717 (0.529)	1.133*** (0.302)	0.059*** (0.021)	0.012 (0.017)

Notes: Statistics based on the main sample augmented with observations with missing ADMIN or ALERT but not any other variable. Heteroscedasticity-robust standard errors are in parentheses. Observations are weighted to account for the complex sampling design of FoodAPS.
***Statistically significant at the 1% level.
**Statistically significant at the 5% level.
*Statistically significant at the 10% level.

latter. For obesity, in all but two cases, SNAP is significant at the 5% level or better, with coefficient estimates of between 5.6 and 6.8 percentage points. The outliers are the self-reported SNAP measure, for which the estimate is notably larger (9.4 percentage points), and ADMIN Alternate 1, for which it is notably smaller (4.3 percentage points) and statistically insignificant.

While there is therefore some evidence of meaningful sensitivities in the magnitudes of the estimated associations across the 11 different constructions of the SNAP variable, the differences are only statistically significant at the 10% level or better in four cases, all for very low food security. These significant differences are for the estimated coefficients using ADMIN Alternate 2 versus ALERT Alternate 1 (p = 0.03), ADMIN Alternate 3 versus ALERT Alternate 1 (p = 0.07), ALERT Alternate 1 versus ALERT Alternate 2 (p = 0.06), and ALERT Alternate 1 versus ALERT Alternate 4 (p = 0.09).[24]

Different Classification Choices when Combining the SNAP Measures

Finally, Table 9 presents regression results using the five decision rules discussed in "Different Classification Choices when Combining ADMIN and ALERT" section, as well as our preferred consolidation rule from the "Preferred Approach to Combining REPORT, ADMIN, and ALERT" section (Preferred SNAP). Additionally, we consider a version of our preferred measure that imputes the missing values from Categories D and E (Preferred SNAP [with imputation]). We perform multiple imputations under the assumption that the likelihood of missing data is correlated with observables but conditionally independent of unobservables, usually referred to as a "Missing at Random (MAR)" assumption.[25]

The first five rows show the results using the ad hoc decision rules, while the last two rows use our preferred measure both with and without imputation. Again, the signs are robust to the different SNAP measures, but there are important differences in significance levels and magnitudes. For instance, the association between SNAP and very low food security is significant and large (3.4 percentage points) using Rule 4 but insignificant in the other cases with a magnitude as small as 0.9. The difference between the largest and smallest estimates is therefore 280%. The estimate for HEI is usually significant and reaches as large as −1.69 units, but it is an insignificant −0.9 unit under Rule 2, for a spread of 88%. For severe obesity, significance levels are again mixed, with the estimates ranging from 3.2 to 4.7 percentage points (spread of 124%). However, these differences in magnitudes are never statistically significant, as they all have p-values all above 0.2.

Using the preferred measure, the results are very similar both with and without imputation. SNAP is predicted to increase the probabilities of being food insecure, having very low food security, being obese, and being severely obese by 6.7, 2.7, 7.2, and 4.5 percentage points, respectively.

[24] Several differences have p-values in the 0.1–0.2 range. For very low food security, these are for REPORT versus ADMIN Alternate 2 (p = 0.11), REPORT versus ALERT Alternate 2 (p = 0.19), ADMIN Baseline versus ALERT Alternate 1 (p = 0.12), ADMIN Alternate 4 versus ALERT Alternate 1 (p = 0.18), ALERT Baseline versus ALERT Alternate 1 (p = 0.18), and ALERT Alternate 1 versus ALERT Alternate 3 (p = 0.11). For HEI, these include ADMIN Alternate 1 versus ALERT Alternate 4 (p = 0.19) and ALERT Alternate 1 versus ALERT Alternate 4 (0.13). For obesity, there is only one: REPORT versus ADMIN Alternate 1 (p = 0.13). For severe obesity, these are for REPORT versus ALERT Alternate 4 (p = 0.17), ADMIN Baseline and ALERT Alternate 4 (p = 0.13), ADMIN Alternate 2 versus ALERT Alternate 4 (p = 0.17), ADMIN Alternate 3 versus ALERT Alternate 4 (p = 0.12), and ADMIN Alternate 4 versus ALERT Alternate 4 (p = 0.1).

[25] We implement the multiple imputation procedures using Stata's *mi impute* and *mi estimate* commands, with 50 multiply imputed samples.

Table 9. Regression Results Combining Participation Measures through Various Rules

	Sample Size	Food Insecurity	Very Low Food Security	Healthy Eating Index	Body Mass Index	Obese	Severely Obese
Rule 1: always use ADMIN unless missing	2446	0.050*** (0.019)	0.022 (0.017)	−1.538*** (0.558)	1.352*** (0.320)	0.068*** (0.021)	0.047*** (0.017)
Rule 2: always use ALERT unless missing	2446	0.060*** (0.019)	0.02 (0.017)	−0.901 (0.561)	1.308*** (0.316)	0.070*** (0.022)	0.035** (0.017)
Rule 3: drop if disagreement	1898	0.075*** (0.021)	0.028 (0.019)	−1.689*** (0.652)	1.297*** (0.363)	0.063** (0.025)	0.032 (0.020)
Rule 4: more weight to matches	2446	0.061*** (0.019)	0.034** (0.017)	−1.482*** (0.567)	1.410*** (0.316)	0.071*** (0.022)	0.045*** (0.017)
Rule 5: more weight to nonmatches	2446	0.050*** (0.019)	0.009 (0.017)	−0.989* (0.558)	1.276*** (0.321)	0.069*** (0.021)	0.038** (0.017)
Preferred SNAP	2108	0.069*** (0.020)	0.026 (0.018)	−1.298** (0.607)	1.475*** (0.337)	0.073*** (0.023)	0.043** (0.018)
Preferred SNAP (with imputation)	2788	0.067*** (0.020)	0.027 (0.017)	−1.401** (0.626)	1.447*** (0.328)	0.072*** (0.023)	0.045** (0.018)

Notes: Statistics based on the main sample augmented with observations with missing ADMIN or ALERT but not any other variable. Heteroscedasticity-robust standard errors are in parentheses. Observations are weighted to account for the complex sampling design of FoodAPS.
***Statistically significant at the 1% level.
**Statistically significant at the 5% level.
*Statistically significant at the 10% level.

SNAP also increases BMI by 1.45 units and reduces HEI by 1.4 units. SNAP is significant at the 5% level or better for all outcomes except very low food security.

5. Conclusion

This article leverages the availability of self-reported and two different administrative measures of SNAP participation in the FoodAPS to investigate several issues related to SNAP and measurement error. We first present evidence that the two administrative SNAP variables suffer from considerable ambiguity and disagree with each other almost as much as they disagree with self-reported participation. We then demonstrate that different methods of coding the two administrative variables separately as well as various approaches to combining their resulting preferred versions into a single "true" participation measure can lead to meaningfully different estimated participation and misreporting rates. Next, we examine sensitivity to assumptions about the administrative variables across OLS estimates of the associations of SNAP with food insecurity, body weight, and healthfulness of food purchases. There are some instances of meaningful differences in coefficient estimates and levels of statistical significance across various constructions of the SNAP variable. However, in most cases, the coefficient estimates are reasonably similar, and the differences between the estimates are in most cases not statistically significant.

Our work serves as a cautionary tale for using administrative records uncritically under the assumption that they represent the "gold standard" with regard to measurement. While some of the difficulties we observed with the linked administrative variables may be unique to FoodAPS, others likely generalize to other settings. For instance, challenges with obtaining data from all states and differences in data quality across states are hardly unique to SNAP caseload files, as many programs (such as Medicaid and public schools) are operated at the state or local levels and standards for data collection may differ across different geographic areas. Additionally, probabilistic matching between survey respondents and verified program participants would be necessary for other contexts as well, since it is unlikely that both sources include universal identifiers such as social security numbers. Moreover, the fact that matches to EBT transaction data were not attempted for individuals who (perhaps erroneously) reported not participating in SNAP points to the broader tradeoff between rigor and budgetary/practical constraints during data collection. When faced with a choice between nationwide surveys and administrative records that are only available for certain areas or individuals and potentially flawed for others, it is not obvious that the administrative data are preferable.

With all that said, our work also provides an example of how researchers can leverage the availability of multiple imperfect administrative and self-reported measures of a program participation variable to reach clear conclusions. Our approach involves demonstrating robustness to an array of sensitivity analyses, plus developing a single, comprehensive measure based on all available information. This allows us to obtain "preferred" results, both for participation rates and regression estimates, as well as a range of possible plausible values. Similar strategies could potentially be utilized in other contexts as well.

Nonetheless, our study suffers from several limitations that should be addressed in future work. For instance, while we propose a method that intuitively should minimize measurement error, there is no way to directly test whether it indeed accomplishes that objective or whether other strategies could be superior. Additionally, we purposefully do not address endogenous SNAP participation because of inherent difficulties in pursuing standard IV methods with a single cross-sectional data with a relatively small sample size. Furthermore, because of the modest sample size, it is difficult to tell whether the lack of statistically significant differences across regression estimates using the various SNAP measures truly indicates the lack of "real" differences. Much is therefore left to be learned about both the impacts of SNAP and best practices for measurement when multiple flawed indicators of program participation are available.

Appendix: 10-Question Food Security Question in FoodAPS

Question	Description
E2	In last 30 days, worried food would run out before we got more money
E3	Food ran out and had no money to buy more, in last 30 days
E4	Could not afford to eat balanced meals, in last 30 days
E5	Adults skipped or cut size of meals b/c not enough money, in last 30 days (Y/N) Universe: Answered "Sometimes not enough to eat" or "Often not enough to eat" description of food sufficiency question within last 30 days, OR answered "Often true" or "Sometimes true" to E2, E3, or E4
E5a	Number of days adults skipped/cut meal size b/c not enough money, last 30 days Universe: Answered "Yes" to E5
E6	Eat less than felt you should b/c not enough money, in last 30 days (Y/N) Universe: Answered "Sometimes not enough to eat" or "Often not enough to eat" description of food sufficiency question within last 30 days, OR answered "Often true" or "Sometimes true" to E2, E3, or E4
E7	Ever hungry but did not eat b/c not enough money, in last 30 days (Y/N) Universe: Answered "Sometimes not enough to eat" or "Often not enough to eat" description of food sufficiency question within last 30 days, OR answered "Often true" or "Sometimes true" to E2, E3, or E4
E8	Lose weight b/c not enough money for food, in last 30 days (Y/N) Universe: Answered "Sometimes not enough to eat" or "Often not enough to eat" description of food sufficiency question within last 30 days, OR answered "often true" or "Sometimes true" to E2, E3, or E4.
E9	Skip food all day b/c not enough money for food, in last 30 days (Y/N) Universe: Answered "Yes" to E5, E5a, E6, E7, or E8
E9a	How often adults skipped food all day b/c not enough money, in last 30 days universe: Answered "Yes" to E9

Acknowledgments

This research was supported by USDA grant no. 59-5000-5-0115 to the National Bureau of Economic Research, entitled, "Using FoodAPS for Research in Diet, Health, Nutrition, and Food Security." The views expressed are those of the authors and not necessarily those of the Economic Research Service, Food and Nutrition Service, or the U.S. Department of Agriculture. We thank the editor Madeline Zavodny, anonymous referees, Marianne Bitler, Janet Currie, Christian Gregory, John Kirlin, Diane Schanzenbach, and audiences at the Association for Public Policy Analysis and Management Fall Research Conference, National Bureau of Economic Research Conferences on Using FoodAPS for Research in Diet, Health, Nutrition, and Food Security, and the USDA Economic Research Service for helpful comments.

References

Almada, Lorenzo N., Ian McCarthy, and Rusty Tchernis. 2016. What can we learn about the effects of food stamps on obesity in the presence of misreporting? *American Journal of Agricultural Economics* 98:997–1017.

Almada, Lorenzo N., and Rusty Tchernis. 2018. Measuring effects of SNAP on obesity at the intensive margin. *Economics and Human Biology* 31:150–63.

Bitler, Marianne P., Janet Currie, and John Karl Scholz. 2003. WIC eligibility and participation. *Journal of Human Resources* 38:1139–79.

Bitler, Marianne P., and Janet Currie. 2005. Does WIC work? The effects of WIC on pregnancy and birth outcomes. *Journal of Policy Analysis and Management* 24:73–91.

Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. Measurement error in survey data. *Handbook of Econometrics* 5: 3705–843.

Courtemanche, Charles J., Joshua C. Pinkston, Christopher J. Ruhm, and George L. Wehby. 2016. Can changing economic factors explain the rise in obesity? *Southern Economic Journal* 82:1266–310.

Currie, Janet. 2003. US food and nutrition programs. In *Means-tested transfer programs in the United States*, edited by Robert A. Moffitt. Chicago, IL: University of Chicago Press, pp. 199–290.

Denteh, Augustine. 2017. The effect of SNAP on obesity in the presence of endogenous misreporting. Unpublished paper.

Gundersen, Craig, and Victor Oliveira. 2001. The food stamp program and food insufficiency. *American Journal of Agricultural Economics* 83:875–87.

Gundersen, Craig. 2015. SNAP and obesity. In *SNAP matters: How food stamps affect health and well being*, edited by J. Bartfeld, C. Gundersen, T. Smeeding and J. P. Ziliak. Redwood City, CA: Stanford University Press, pp. 161–85.

Hofferth, Sandra L. 2004. Persistence and change in the food security of families with children, 1997-1999. E-FAN-04-001. Economic Research Service, U.S. Department of Agriculture. Available www.ers.usda.gov/publications/efan04001/

Hoynes, Hilary W., and Diane Whitmore Schanzenbach. 2016. U.S. food and nutrition programs. In *Economics of means-tested transfer programs in the United States*, volume 1, edited by R. Moffit. Chicago, IL: University of Chicago Press, pp. 219–301.

Huffman, Sonya Kostova, and Jensen, Helen H.. 2003. Do food assistance programs improve household food security? Recent evidence from the United States. Center for Agricultural and Rural Development Working Paper No. 03-WP 335.

Kreider, Brent. 2010. Regression coefficient identification decay in the presence of infrequent classification errors. *The Review of Economics and Statistics* 92(4):1017–23.

Kreider, Brent, John V. Pepper, Craig Gundersen, and Dean Jolliffe. 2012. Identifying the effects of SNAP (food stamps) on child health outcomes when participation is endogenous and misreported. *Journal of the American Statistical Association* 107(499):958–75.

Mabli, James, Ohls, Jim, Dragoset, Lisa, Castner, Laura, and Santos, Betsy. 2013. Measuring the effect of Supplemental Nutrition Assistance Program (SNAP) participation on food security. Mathematica Policy Research for the U.S. Department of Agriculture, Food and Nutrition Service.

Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan. 2015. Household surveys in crisis. *Journal of Economic Perspectives, American Economic Association* 29:199–266.

Meyer, Bruce D., Mittag, Nikolas, and Goerge, Robert M. 2018. Errors in survey reporting and imputation and their effects on estimates of food stamp program participation. NBER Working Paper No. 25143.

Meyer, Bruce D., and Mittag, Nikolas. 2018. Misreporting of Government Transfers: How Important are Survey Design and Geography? IZA Discussion Paper No. 12038. Available at SSRN: https://nam04.safelinks.protection.outlook.com/?url=https%3A%2F%2Fssrn.com%2Fabstract%3D3318813&data=02%7C01%7Ccourtemanche%40uky.edu%7C7dfcd67fdf664b12cdaa08d6bcee07ae%7C2b30530b69b64457b818481cb53d42ae%7C0%7C0%7C636904127857402341&sdata=zOyUF5cvWi0Nhm7k1ZWShPs%2BXR3np0vN4BK7swntulw%3D&reserved=0"https://ssrn.com/abstract=3318813

Meyerhoefer, Chad D., and Yuriy Pylypchuk. 2008. Does participation in the food stamp program increase the prevalence of obesity and health care spending? *American Journal of Agricultural Economics* 90:287–305.

Mittag, Nikolas. 2013. A method of correcting for misreporting applied to the food stamp program. US Census Bureau Center for Economic Studies Paper No. CES-WP-13-28.

Mittag, Nikolas. 2016. Correcting for misreporting of government benefits. IZA Discussion Paper No. 10266.

Mykerezi, Elton, and Bradford Mills. 2010. The impact of food stamp program participation on household food insecurity. *American Journal of Agricultural Economics* 92(5):1379–91.

Nord, Mark, and Prell, Mark A.. 2011. Food security improved following the 2009 ARRA increase in SNAP benefits. USDA, Economic Research Service, Economic Research Report No. 116.

Nguimkeu, P., A. Denteh, and R. Tchernis. 2019. On the estimation of treatment effects with endogenous misreporting. *Journal of Econometrics* 208(2):487–506.

Schmidt, Lucie, Lara Shore-Sheppard, and Tara Watson. 2016. The effect of safety net programs on food insecurity. *Journal of Human Resources* 51(3):589–614.

U.S. Department of Agriculture. 2012. *Building a healthy America: A profile of the Supplemental Nutrition Assistance Program*. Washington, DC: Office of Research and Analysis, Food and Nutrition Service, U.S. Department of Agriculture.

Volpe, Richard, Abigail Okrent, and Ephraim Leibtag. 2013. The effect of supercenter-format stores on the healthfulness of consumers' grocery purchases. *American Journal of Agricultural Economics* 95:568–89.

Van Hook, Jennifer, and Kelly Stamper Balistreri. 2006. Ineligible parents, eligible children: Food stamps receipt, allotments and food insecurity among children of immigrants. *Social Science Research* 35(1):228–51.

Wilde, Parke, and Mark Nord. 2005. The effect of food stamps on food security: A panel data approach. *Review of Agricultural Economics* 27(3):425–32.