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

Misreporting of Government Transfers: How Important Are Survey Design and Geography?

Bruce D. Meyer* and Nikolas Mittag†

Recent studies linking household surveys to administrative records reveal high rates of misreporting of program receipt. We use the FoodAPS survey to examine whether the findings of these studies of general household surveys using one or two states generalize to a survey with a narrow focus and across many states. First, we study how reporting errors differ from other surveys. We find a lower rate of false negatives (failures to report true receipt) in FoodAPS, likely partly due to the shorter recall period of FoodAPS. Misreporting varies with household characteristics and between interviewers. Second, we examine geographic heterogeneity in survey error to assess whether we can extrapolate from linked data from a few states. We find systematic differences between states in unconditional error rates but no evidence of substantial differences conditional on common covariates. Thus, extrapolating error rates across states may yield more accurate receipt estimates than uncorrected survey estimates.

JEL Classification: C81, D31, I32, I38

1. Introduction

Survey data are crucial for academic research as well as to make informed policy decisions. Yet, the accuracy of the information obtained from surveys depends on the accuracy of survey responses. Unfortunately, several indicators of survey quality, such as the nonresponse rate and estimates of the extent of misreporting show that the accuracy of surveys is declining. As discussed in Meyer et al. (2015), the extent of misreporting is particularly large for questions on government transfer programs. Recent studies have linked survey data from major household surveys to state administrative records on transfer programs (see Bound et al. 2001 for an overview). These studies reveal high rates of underreporting, showing that sometimes more than half of true Supplemental Nutrition Assistance Program (SNAP) recipients do not report receipt in the survey data. The errors are systematically related to other variables in the surveys, so that they severely bias studies of poverty and program receipt as well as analyses of the safety net and its effectiveness (see e.g., Bollinger and David 1997; Cerf Harris 2014; Meyer et al. 2018; Meyer and Mittag 2019).

* University of Chicago, NBER, and AEI, Chicago, IL, 60637, USA.

† CERGE-EI, Prague 1, 110 00, Czech Republic; E-mail: nikolasmittag@posteo.de; corresponding author.

Prior linkage studies linked administrative data for at most a few states to general economic household surveys in which SNAP receipt was not a central topic,[1] which raises two questions this study seeks to address. First, do findings on survey error from general economic household surveys generalize to surveys with a narrow focus on SNAP receipt? Studies such as Celhay et al. (2018b) point out the importance of survey design, but the similarity of the previously linked surveys limits the extent to which they can analyze how survey design affects survey error. Second, are results from studies with a few states generalizable nationally and are corrections or extrapolation based on these geographic subsamples likely to improve estimate accuracy? Both issues hinge on the extent of geographic variation in survey error.

In this study, we use data from the National Household Food Acquisition and Purchase Survey (FoodAPS) survey linked to administrative SNAP records from multiple states to examine the extent and nature of survey error in reported SNAP receipt, specifically whether respondents report receipt in the month before the interview. We examine how this survey error differs between the specialized FoodAPS survey that emphasizes SNAP and nutrition programs and general economic surveys. We first examine the extent and nature of survey error in FoodAPS and compare it to previous studies. At 18.3%, the rate at which recipient households fail to report receipt is lower than in previously linked surveys. At 1.2%, the rate at which households not classified as recipients by the linked administrative records report receipt is similar to those in prior studies. We emphasize the role of survey design features, such as reference periods and interviewers, but also assess whether common predictors of survey error differ between FoodAPS and previously linked surveys. Recall length affects errors and a shorter recall period contributes to the lower FoodAPS error rates. We do not find major differences between FoodAPS and the previous surveys in terms of the key predictors of survey error. We provide evidence that false negative rates vary across interviewers both unconditionally and when controlling for variables related to interviewer assignment. Finally, we use the linked FoodAPS data from multiple states to study geographic heterogeneity in survey error. We reject that unconditional differences between reported and administrative receipt status are the same across states. Nevertheless, we do not find evidence of geographic variation in survey error conditional on demographic characteristics, so that extrapolation across geography still improves the accuracy of state-specific SNAP receipt rates over estimates based on survey reports.

These results extend findings from prior linkage studies in several ways that can help survey users and producers assess and improve estimate accuracy. By examining survey error in a specialized survey, FoodAPS, we are not only able to demonstrate similarities in the patterns of survey error across surveys, but also that the nature and extent of survey error clearly depends on survey design. Applied researchers can use these insights to gauge the reliability of their data and the applicability of corrections for survey error. If survey error is sufficiently similar, information from this study can also be incorporated in corrections for survey error. The FoodAPS data allow us to advance our understanding of the effects of survey design features such as reference periods and the importance of interviewers, which can help survey producers to improve survey accuracy. Contrary to prior linkage studies, we are able to analyze the extent of geographic heterogeneity in survey error. We thereby provide first evidence on our ability to learn and extrapolate from studies that only link one or two states. Overall, our results extend our understanding of survey error and provide guidance for both survey producers and survey users in choosing an appropriate way of dealing with problems of data accuracy.

---

[1] See for example, Bollinger and David (1997), Marquis and Moore (1990), Taeuber et al. (2004), Cerf Harris (2014), Kirlin and Wiseman (2014), Meyer et al. (2018), and Celhay et al. (2018a).

The next section summarizes our data sources, data linkage, and definitions. Section 3 describes the differences between the survey reports of SNAP receipt and the linked administrative receipt variable. Section 4 studies the determinants of these differences and the role of survey design. Section 5 examines geographic heterogeneity. Section 6 concludes.

2. Data and Linkage

This section describes the creation of the linked data, our sample, and the SNAP receipt variables. We first describe the sources of the survey and administrative information. Next, we summarize how the two data sources are linked and how linkage issues limit the sample we use in the analyses later. Finally, we discuss how we define SNAP receipt variables based on survey and administrative data that are sufficiently comparable to study survey error. The information provided in the survey only allows us to study survey error in whether the date of last receipt was correctly reported as being in the month before the interview or not. Accurately matching this variable in the administrative data requires further reducing the sample of households for which we can validate SNAP reports. Figure 1 provides an overview of the choices that lead to our analysis sample.

Survey Data

The FoodAPS survey collects data from 27 states on all household food purchases, nutrition information, and health. The sample is stratified and weighted to yield nationally representative estimates for research on health and obesity, food insecurity, and food assistance. The survey particularly highlights SNAP receipt and recipients. This emphasis dictates the survey design, as FoodAPS uses two separate sampling frames: a SNAP recipient and a nonrecipient frame. The recipient frame was constructed from addresses of individuals receiving SNAP in February 2012 provided by the states. The nonrecipient frame was constructed by removing the recipient addresses from an address-based sampling list constructed from the U.S. Postal Service Delivery Sequence file. Roughly 30% of addresses were sampled from the recipient frame; the remainder of the addresses were sampled from the nonrecipient frame.

To account for changes of receipt status after the construction of the sampling frame,[2] FoodAPS asked about SNAP receipt in a screener interview conducted by telephone. The main
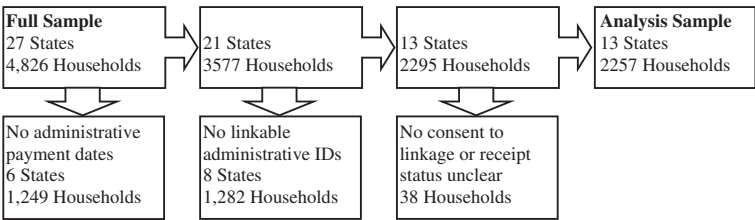

Figure 1. Overview of Choices Leading to the Analysis Sample.

---

[2] Such changes may arise, because sampled households may enroll or leave SNAP or move between the construction of the sampling frame and the interview (April 2012–January 2013). In addition, five states did not provide the address data to construct the SNAP recipient frame, so there was only one sampling frame in these states. See the FoodAPS user guide (USDA ERS 2016a) for further detail.

point of the screener interviews was to mitigate nonresponse bias and determine eligibility of the household to participate in FoodAPS. Screener information on SNAP receipt and income was also used to define four strata or target groups. The first target group included all SNAP recipients. The other three target groups contained nonrecipients stratified by income relative to the poverty line. A sample of 4826 households was interviewed before and after a week of recording food acquisition between April 2012 and January 2013. Computer-aided in-person interviews were conducted with the main food shopper in the household. All information on SNAP receipt and most of the remaining information used in this study stems from the initial interview. Only a few covariates such as household income were collected in the final interview. See the Appendix for a brief summary of the interview procedure and content.

The first SNAP question in FoodAPS was "Do you or does anyone in your household receive benefits from the SNAP program?" If the respondent indicated yes, questions on the date of the last receipt of SNAP benefits, the amount received and whether the amount received was higher, lower, or equal to the usual amount received followed. Those who did not indicate current SNAP receipt were asked questions about SNAP receipt in the past including a question on the date of their last SNAP payment. Note that the initial question did not define a reference period but left the respondent some leeway in defining what constitutes "receiving benefits from the SNAP program." To study survey error, we need to match the definitions in the survey question exactly, which is not possible if the question depends on the respondent's interpretation. Therefore, we cannot study the initial question on current SNAP receipt. Rather, we examine the following question on the date of the last payment, because the administrative data allow us to exactly match the information this question asks as we discuss later. The FoodAPS documentation provides further information on the survey design.

There are several survey design reasons why misreporting in FoodAPS may differ from that in the general economic surveys used in previous studies. Celhay et al. (2018b) provide a detailed discussion and empirical evidence on how survey design features (such as salience, reference periods, recall, stigma, and respondent cooperation) may affect survey error. Many of these sources of survey error differ between major household surveys and FoodAPS. SNAP receipt was a salient topic in FoodAPS. Respondents were asked about SNAP receipt both in the screener interview and early in the initial interview. The questions regarding SNAP receipt also differed from those in major household surveys. Most importantly, the survey asked about payments in the recent past. Thereby, FoodAPS had a much simpler reference period than most major household surveys that ask about receipt in the past year or even the past calendar year. Most large surveys ask just one question on receipt and possibly a follow-up question on amounts received. The FoodAPS interview contained a separate section on nutrition assistance that included a sequence of questions on SNAP receipt, making the topic more salient. The salience of SNAP may not only have affected survey accuracy by altering respondent behavior, but also by affecting interviewers. Interviewers were aware of the importance of SNAP receipt status and had information on SNAP receipt status as reported in the screener interview.

Administrative Records and Data Linkage

The FoodAPS survey data were combined with administrative data on SNAP from two sources: administrative records on SNAP payments and SNAP electronic benefit (EBT) card transactions. The administrative records on SNAP payments, called the caseload data, were obtained from state SNAP agencies. The transaction data, referred to as the ALERT data, were provided by the U.S. Department of Agriculture (USDA) and contain a record for each payment with a SNAP

EBT card. The FoodAPS documentation (USDA ERS 2016a,b) provides detailed information on the content of these files and how they were linked to the survey data.

Unfortunately, the administrative records differ between states. Six of the 27 participating states did not provide caseload data with payment dates. Therefore, only the ALERT data on SNAP use was linked to the FoodAPS survey. Consequently, no reliable administrative data on the date of the last payment is available in these states. Information on the timing of the last payment is crucial to the central goal of this study, which is to examine whether survey respondents accurately reported receipt of SNAP in the past month. Therefore, we excluded the six states for which this information is not available. Of the remaining 21 states, we exclude eight states that provided case-load and ALERT data, but the ALERT data did not contain linkable identifiers. Thus, rather than linking the combined caseload and ALERT data to the survey, the two data sources had to be linked to the survey separately. In addition to the probabilistic link of the caseload data summarized later, transactions from the ALERT data were linked to the survey data using a different probabilistic record linkage procedure described in the Appendix.[3]

Restricting the sample as summarized in Figure 1 involves a trade-off between a large, repre-sentative sample and data accuracy. Excluding some states and households means that our sample is not representative of the entire United States and (due to stratification and reweighting) not neces-sarily representative of the population in the states in our sample. Nevertheless, rather than obtaining nationally representative estimates, our main goal is to study survey error and to what extent we can obtain nationally representative estimates in the presence of survey error that may vary across states. Consequently, we only analyze data from the 13 states that provided the most accurate and homogeneous data to minimize confounding survey error with variation in administra-tive data and linkage quality.[4] Courtemanche et al. (2018) and Kang and Moffitt (2018) use the data from all states and examine their accuracy and usefulness further. They also find differences in the linked administrative variable between the 13 states that provided caseload and ALERT data with unique IDs and the remaining states. Providing less information and linking the two administrative data sources separately likely induced additional error, so that the linked administrative variable is less accurate for the states we exclude. In addition, studying geographic variation in survey error requires the accuracy of the linked administrative variable not to vary between states. Courtemanche et al. (2018) show that variation between states also stems from differences in data quality or link-age rather than from differences in reporting only. For the sample we use, we have no evidence of systematic differences in the accuracy of the linked data between states. Consequently, these states provide us with the largest sample in which the accuracy of the linked data should be homogeneous enough to attribute variation in the differences between the administrative and survey variables to differences in survey error rather than variation in administrative data quality.

Even for the 13 states in our sample, the data do not contain identifiers to link the administra-tive data to the survey, so they were linked probabilistically as described in the FoodAPS documen-tation. The two data sources were linked directly, rather than linking both sources to a population register as done in many prior record linkage studies. Thus, we cannot determine the extent of or adjust for missed links, because we cannot distinguish survey records that did not link to the admin-istrative data because they do not receive SNAP, from survey records that did not link due to

---

[3] See Fellegi and Sunter (1969), Copas and Hilton (1990), and Winkler (2014) for detailed discussions of probabilistic record linkage and the FoodAPS documentation for detail on the implementation.

[4] For one of the 13 states, linkable identifiers were not provided for about half the state, because the state uses two different processing systems for two geographic regions. We only include the subsample with identifiers.

missing or incorrect personal information. See Meyer and Mittag (2018) for further discussion and Meyer et al. (2018) for arguments why such missed links likely understate net differences between the data sources. Twenty-one households in our sample did not provide consent to record linkage. We exclude these households from our sample throughout. For the remaining households, matching records in the caseload data were searched based on first name, last name, phone number, and street address (including apartment number). If a household was matched to a SNAP case, the date of the most recent SNAP payment according to the caseload data was added to the FoodAPS data.

Comparable Administrative and Reported Definitions of SNAP Receipt

Examining survey errors requires comparing survey responses to a measure of truth or a more accurate variable that measures the same concept. See for example, Groves and Lyberg (2010) for a discussion. Consequently, we need to define a measure of SNAP receipt based on the survey reports and construct an administrative variable that matches this definition as closely as possible.

As discussed earlier, the question on which the variable whether anyone in the household receives SNAP (SNAPNOWREPORT) is based does not indicate a reference period.[5] Consequently, we cannot match the definition of SNAPNOWREPORT in the administrative data. Instead, we define current receipt based on the reported date of the last payment. Our survey measure of current SNAP receipt is an indicator whether the reported date of last SNAP receipt occurred on the interview day or in the 32 days before the interview.[6] According to our survey variable, 691 out of 2257 survey households are SNAP recipients, which yields a (weighted) receipt rate of 11.7%. There are several other ways to define SNAP receipt. Table A1 provides an overview of both survey and administrative SNAP receipt definitions. Changing the time period of reported SNAP receipt to 31 or 33 days before the interview reclassifies less than five households. Neither change has a visible effect on the rates of reported receipt and survey error. On the other hand, our variable differs from SNAPNOWREPORT for 59 households. Fifty-three are reclassified as recipients, which increases the estimated reported receipt rate by one percentage point.

To assess errors in the survey reports, we need to define a second SNAP receipt variable based on the linked administrative data that matches our survey definition of SNAP receipt. Both the caseload and ALERT data contain exact dates of payments and transactions, so we can match the definition of the survey variable in each of the two linked administrative data sources exactly. We consider a household to be a current SNAP recipient according to the administrative data if either the linked caseload or ALERT data indicate a payment or a transaction on the interview day or in the 32 days before the interview. As Table A1 shows, using a 31 or 33-day period before the interview day does not affect the variable at all.

The survey question asks about the date of the last benefit receipt, which would ideally be captured by the SNAP payments from the caseload data alone. Even though this definition is less congruent with the survey question than one based on the caseload payments only, we use the additional information on transactions from the ALERT data to define administrative SNAP receipt for two

---

[5] The date of last receipt among those who answered this question affirmatively varies between households. Some report dates up to 6 months in the past. Thirty-nine households report a date of last receipt more than 32 days ago, even though this means that they either recently became nonrecipients or received a more recent payment. This suggests that the interpretation of the question indeed varied between respondents. The FoodAPS household codebook (USDA ERS 2016b) provides further detail on this issue.

[6] We use 33 days to ensure that the time period cannot fall between two monthly payments. SNAP payments are dispersed monthly on fixed dates, so a period of 31 days should always include a payment to every recipient. We consider 33 days to make sure we do not miss any regular recipients, because they may have received a payment later on the interview day or because some states do not distribute SNAP payments on Sundays and thus vary payment dates by a day.

reasons. First, the linked caseload variable is not error-free and incorporating information from the ALERT data appears to correct some errors. Table A1 supports this point. There are 119 households with ALERT transactions, but no SNAP payment in the reference period. Transactions require prior payments, so these households likely received payments to their EBT card during the reference period that were not recorded in the linked caseload data due to errors. The only other explanation would be that these households used to receive SNAP, but currently do not, and spent SNAP benefits from prior months during the reference period.[7] This explanation seems unlikely, because most SNAP benefits are spent within a few days of receipt and are rarely carried over to the next month. In addition, there are clearly too many households of this type to be explained by SNAP exit or interruptions.[8] Consequently, the ALERT information on usage in the reference period indicates that the household also received SNAP benefits in the reference period that were missed by the caseload data or link. A second reason to consider these households to be administrative recipients is that we are mainly interested in whether households report receiving support from SNAP. Households who report SNAP receipt this month, because they used the EBT card even though there was no payment this month are a minor problem compared to the large share of households who do not report SNAP receipt. Therefore, we prefer to classify as correct reporters the few households that appear to misinterpret the survey question and report SNAP use rather than benefit receipt.

Defining administrative SNAP receipt based on the most recent administrative dates allows us to match the variable definition of the survey. Unfortunately, how the survey producers determined the most recent date is unclear and this process resulted in some dates being after the initial interview. FoodAPS only merged at most one payment date to each survey household. Thus, we do not know whether these households should be classified as recipients, because they also received SNAP before the interview, or as nonrecipients, because they enrolled in SNAP between the initial interview and the date recorded in the administrative data. A total of 262 administrative recipient households have dates after the initial interview according to at least one of the administrative data sources. It seems implausible that 12% of all interviewed households enroll in SNAP shortly after the interview. Table A2 tabulates the frequency of days passed between the interview and the recorded date, showing that most dates after the initial interview are shortly after the initial interview. This result supports statements by the survey producers that the linked dates after the interview likely arise from using an algorithm to pick the most recent date that may erroneously consider an administrative receipt date after the initial interview to be the most recent date if more days passed between the last date before the initial interview than between the initial interview and the first date after the initial interview.

Nonetheless, some of the households with administrative dates after the initial interview may have enrolled in SNAP after the initial interview and should therefore not be classified as administrative recipients. We combine information on receipt before the interview from several sources to identify households that are unlikely to have enrolled in SNAP between the initial interview and the recorded administrative date. For 145 of the 262 households with a date after the initial interview in either the caseload or ALERT data, the other administrative data source indicates that the household already received SNAP before the initial interview. Of the remaining households, 76 are from the SNAP recipient frame. Consequently, we know that someone at this address received SNAP in February *and* later in the interview year, which makes it very likely that the household also received

---

[7] It is important to note that for our sample, the caseload and ALERT data were linked deterministically, so disagreement between the two sources cannot arise from mislinking one source, but not the other.

[8] As Table A1 shows, recording these households as nonrecipients would lead to an implausibly low rate of SNAP receipt of 10.4% that would indicate substantial over-reporting of 12%.

SNAP at the time of the interview.[9] Finally, for 24 of the remaining observations, the administrative data indicate a payment or transaction within a week of the initial interview. It usually takes several days to process SNAP applications and make payments and hence for transactions to be possible. Thus, it appears more likely that these households were already recipients at the time of the initial interview than that they managed to enroll in SNAP and receive benefits in less than a week. We drop from the sample the remaining 17 households for which we cannot determine whether they already received SNAP at the time of the initial interview.[10]

The households with administrative dates of SNAP receipt or use after the initial interview also lack a measure of time between last receipt and initial interview, a variable we use in a later analysis. We impute a receipt date prior to interview assuming that it occurred 30 days before the recorded date. We only use these imputed variables in our analyses of recall and document that our findings still hold without the imputed observations. Figure 2 shows that including these households with imputed administrative dates of last receipt also removes an anomaly in the frequency distribution of time since last administrative receipt. Interview dates were not aligned with SNAP disbursement dates, so one would expect the number of days between the two dates to be roughly uniformly distributed. The shaded bars in Figure 2 show the frequency of days as originally recorded, that is, without imputed days after the interview. The frequency of interviews between 23 and 30 days is much lower than one would expect based on a uniform distribution. The contour-bars in Figure 2 also include dates after the initial interview replaced by the recorded date minus 30 days.[11] Including these imputed dates adds the "missing" interviews between 23 and 30 days, because most recorded dates after the interview are in the week after the interview as Table A2 shows. Imputing these dates thereby makes the frequency distribution closer to the uniform distribution we would expect, which again underlines that dates after the interview are likely driven by a coding problem that considered dates shortly after the interview to be the most recent date.

Finally, the FoodAPS data set provides a SNAP receipt variable, SNAPNOWHH, that is a combination of administrative and survey data on reported receipt. The main problem in using SNAPNOWHH to study survey error is the way it combines administrative and survey reported receipt, which mutes almost all overreporting. See the Appendix for further information.

3. Comparing Administrative and Reported SNAP Receipt

Our final sample consists of 2257 of the 4826 households in the FoodAPS data. This sample provides us with the most reliable data to study survey error by minimizing the extent of error in the linked administrative variable. This choice makes it more plausible that the extent and cross-state variation in the differences between survey and administrative variables point to survey errors rather than linkage errors. Nonetheless, some errors in the administrative variable certainly remain, making it crucial to

---

[9] This rule may misclassify a small number of households who either interrupted a SNAP spell at the time of the interview or moved into an address from the SNAP recipient frame, did not receive SNAP at the time of the interview, but enrolled in SNAP shortly after the interview date.

[10] Classifying these 17 households as administrative recipients increases the false negative rate by 4.4 percentage points and thereby amplifies our results. Another alternative is to classify only households for which at least one of the administrative data sources indicates receipt in the 32 days before the initial interview as administrative recipients. This rule increases the false positive rate by 1.5 percentage points and reduces the false negative rate by 2.1 percentage points. Thus, most households reclassified as nonrecipients by this rule report receipt, making them likely recipients.

[11] This rule only adds observations to the frequency plot, so that the contour-bars always cover the shaded bars.
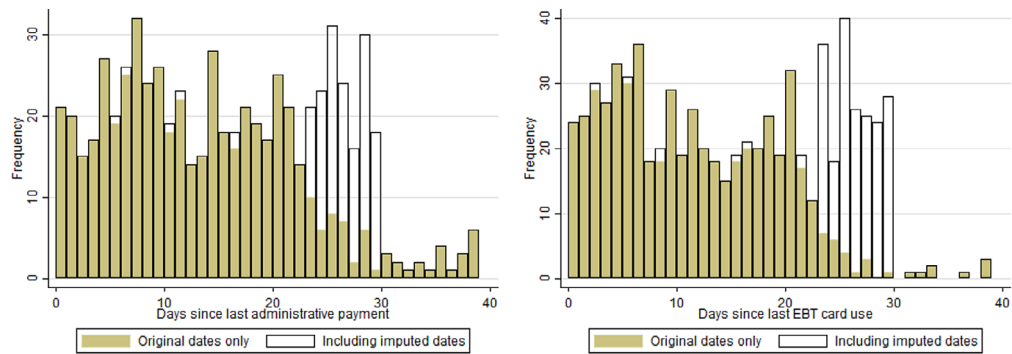
Figure 2. Frequency of Days Since Last SNAP Payment and EBT Card Use.
Notes: Unweighted observation counts of days between initial interview and last payment (left panel) and last EBT card use (right panel). [Color figure can be viewed at wileyonlinelibrary.com]

examine whether any patterns that we attribute to survey error below could also be caused by errors in the linked administrative variable. Table A3 provides summary statistics for our analysis sample. Of the 2257 households in our sample 691 are SNAP recipients according to our survey receipt variable and 768 are SNAP recipients according to our administrative receipt variable. The weighted receipt rates are 11.7% and 13% according to the survey and administrative variables. This comparison implies a net reporting rate (the ratio of the number of recipients according to the survey reports and the linked administrative variable) of 90%, which is much higher than the net reporting rates Meyer et al. (2015) report for various other household surveys.

Table 1 summarizes the differences between the survey and the administrative receipt variables in FoodAPS and compares these differences to prior linkage studies. We define SNAP receipt based on the last date of receipt being in the past month, so subject to the caveat that differences may also be due to errors in the administrative data, we examine whether respondents correctly report a date of last receipt in the past month. The first entry in Table 1 shows the (weighted) rate at which households who receive SNAP according to the administrative variable fail to report SNAP receipt (are false negatives)—18.3% in our FoodAPS sample. Some of these households may be nonrecipients that are erroneously linked to recipient households. Still, it seems reasonable to assume that the probability of failing to link a true recipient household that does not report is higher than the probability of erroneously linking a nonrecipient household to a SNAP case. If so, the true false negative rate is likely to be higher than the 18.3% we estimate. See Meyer et al. (2018) for further discussion.

The second entry in the first row of Table 1 is the (weighted) rate at which households who do not receive SNAP according to the administrative data report SNAP receipt (are false positives)—1.2% in our FoodAPS sample. This estimate likely overstates the frequency at which nonrecipient households report SNAP receipt, because some true recipient households that report receipt may not have been linked to their administrative SNAP case by error. Seventy-five percent of these false positives were sampled from the SNAP recipient frame. This finding is consistent with prior evidence that a large share of those who overreport program receipt received the program at some point outside the reference period (Celhay et al. 2018b).

The bottom part of Table 1 provides false positive and false negative rates from prior linkage studies. The first row contains error rates from Meyer et al. (2018), who linked administrative SNAP records from Illinois (2000–2004) and Maryland (2000–2003) to three major U.S. household surveys, the American Community Survey (ACS), the Current Population Survey (CPS), and the Survey of Income and Program Participation (SIPP). The next row reports error rates from Celhay

et al. (2018a), who link the same surveys to administrative SNAP records from New York for 2007–2013. The false negative rate in our FoodAPS sample is lower than the false negative rates in these prior studies. It is substantially lower than the rates in the ACS and CPS from either study and still slightly lower than in the SIPP. The false positive rate is more similar to prior studies. It is relatively high compared to the false positive rates in the earlier study of Illinois and Maryland, but lower than the false positive rates found later in New York. Consequently, comparing these false positive rates points toward overreporting being similar or less frequent in FoodAPS than in the general economic household surveys evaluated by prior studies.

4. The Determinants of Survey Errors

In this section, we use multivariate analyses to document systematic variation in survey error that provides evidence on the nature of errors and the origins of the differences we document earlier. We examine several survey design features that may account for survey error in FoodAPS and may help to explain why the error rates in FoodAPS differ from the three general economic household surveys in Table 1. In terms of survey design, the ACS, CPS, and SIPP differ substantially from FoodAPS. All three surveys are general economic surveys in which questions on SNAP and nutrition assistance only play a minor role. Being general economic surveys makes them similar in survey design, but there are differences between the three surveys that likely affect reporting accuracy. See Celhay et al. (2018b) for a detailed analysis. We first examine the role of recall. We then describe how survey error varies with demographic characteristics. Finally, we amend the multivariate models with information on FoodAPS interviewers to analyze whether differences between interviewers contribute to variation in survey accuracy.

Recall

Reference periods differ in length and complexity between the surveys in Table 1. The ACS asks about SNAP receipt in the 12 months before the interview. The CPS collects information on SNAP receipt and amounts received in the previous calendar year in spring of each year. The SIPP is a panel survey that asks about monthly SNAP receipt in the four months preceding the interview. Both the length and the complexity of the recall period are known to affect reporting accuracy. The last row of Table 1 already hints at recall playing a role in explaining the differences. It restricts the sample of Celhay et al. (2018a) to current administrative recipients, that is, households that receive SNAP in the month of interview (and in the reference period of the survey in the case of the CPS) according to the administrative data. The recall problem faced by these households is more similar to the one faced by SNAP recipients in FoodAPS in that both are asked to recall an event that happened in the past month.[12] The rates at which current administrative recipients fail to report are indeed lower than in the overall population. The reporting rates of current administrative recipients in the ACS and SIPP are similar to the reporting rate of administrative recipients in FoodAPS— slightly lower in the SIPP (17%) and slightly higher in the ACS (21%). While the false negative rate

---

[12] They still differ in that FoodAPS respondents are asked to recall a date and that respondents of the other surveys are asked to recall any payments over a longer time period rather than only in the past month.

Table 1. Reported and Administrative SNAP Receipt in FoodAPS and Major Household Surveys

	False Negative Rate			False Positive Rate		
FoodAPS	18.3%			1.2%		
Previous Studies Survey	ACS	CPS	SIPP	ACS	CPS	SIPP
Sample						
Illinois, Maryland (2002–2005)	33.1%	49.0%	22.8%	0.7%	0.8%	1.6%
New York (2007–2013)	25.7%	42.1%	19.4%	1.2%	2.0%	1.5%
New York (current recipients)	20.9%	38.3%	17.1%			

Notes: The false positive rate is the percentage of administrative nonrecipient household who report receipt. The false negative rate is the percentage of administrative recipient households who do not report receipt. The error rates for Illinois and Maryland are from Meyer et al. (2018). The New York error rates are from Celhay et al. (2018a). The last row restricts the sample to those who received SNAP during the reference period and at the time of the interview according to the linked administrative receipt variable. All estimates use household weights.

in the CPS drops to 38%, it remains substantially higher than in FoodAPS and the other surveys, when restricting the sample to current administrative recipients.

These lower error rates among current recipients point toward recall errors, but could also be driven by differences in the time period and state under study, or the salience of receipt between current and former recipients, or other differences. The administrative variables in FoodAPS provide us with the number of days since the last payment and the number of days since the last transaction. This information enables us to examine the role that time elapsed since these events has on responses. Contrary to prior studies where the elapsed time depends on a respondent's SNAP receipt history, here the variation is due to the timing of the interview relative to the last SNAP payment or use. This variation, though only over the month before interview, is likely exogenous, adding to the credibility of the evidence.

Figure 3 plots how the reporting rate among administrative recipients varies with these two measures of time elapsed.[13] Both the dashed line that is based on days since last administrative receipt and the solid line that is based on time since last use remain fairly stable at a reporting rate slightly above 80% for the first 20 days. The reporting rate declines rapidly with the number of days since last use after that. On the other hand, reporting rates do not seem to decline in the number of days passed since the last payment, even after 20 days. Figure 3 confirms prior results showing that the probability of reporting an event is declining in the length of time since the event occurred (e.g., Sudman and Bradburn 1973; Groves et al. 2009; Celhay et al. 2018b). We find that the probability of reporting receipt declines in time since last EBT card use rather than time since last SNAP payment, which suggests that recall of recent SNAP receipt is driven by recent SNAP use, rather than the receipt of payments.

Table 2 reports estimates from multivariate (single-equation) Probit models of the probability of survey errors.[14] The estimates in columns 1 and 2 are the determinants of the probability that an administrative recipient household fails to report SNAP in the survey, that is, is a false negative. In addition to conditioning on time since use and receipt, the model includes demographic and economic characteristics. The results confirm that it is time since last SNAP use rather than last payment that predicts failure to report SNAP receipt, holding other household characteristics constant.

---

[13] The figure excludes those with imputed administrative dates of last receipt or use. Including these observations does not change the figure much, though the decline in the reporting rate with use is slightly less pronounced.

[14] The multivariate analyses use a sample of 2256 observations throughout, because one observation is missing a value of one of the covariates. The results are not affected by dealing with the missing data in other ways.
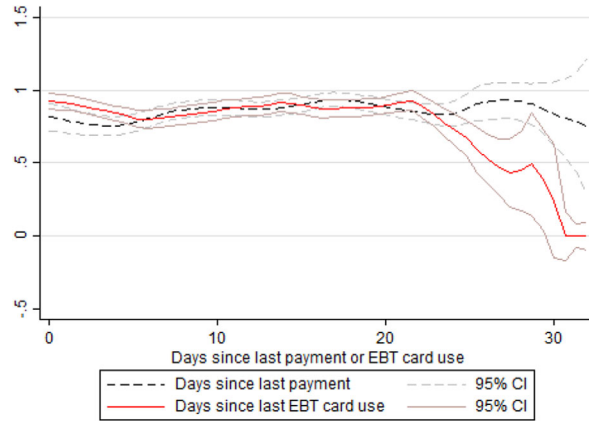
Figure 3. Reporting Rate by Time Since Last SNAP Payment and EBT Card Use.
Notes: Nonparametric (local polynomial) regression of reported SNAP receipt on days since last administrative SNAP payment or EBT card use. The sample is restricted to administrative SNAP recipients. [Color figure can be viewed at wileyonlinelibrary.com]

The number of days since the last ALERT transaction is significant and increases the probability of a false negative by 0.56 percentage points per day. The effect of the number of days since the last payment is insignificant and small at 0.14 percentage points per day.[15]

That time since last SNAP use rather than last payment drives the recall results may be due to people being more aware of card use than of payment receipt. The difficulty recalling a specific event (a SNAP payment) may not depend on the time since that event, but on the time since the last time the respondent was reminded of this event. This result suggests that the fact that FoodAPS attempted to interview the main food shopper may have contributed to the lower false negative rates. The main food shopper is more likely to possess and hence use the EBT card, which according to our analyses reduces recall problems. Our finding that card usage rather than payment receipt matters also points to another potential mechanism behind the common finding that larger households are more likely to fail to report program receipt: The larger the household, the less likely it is that a given respondent recently used the EBT card and hence the more likely that the household fails to report.

Demographic Characteristics

We would like to examine if the predictors of survey error differ between general economic household surveys and a specialized survey like FoodAPS, by comparing the determinants of FoodAPS errors to those in the literature. In addition to the determinants of false negative reports, Table 2 also reports estimates of a Probit model for the determinants of false positive reports, the probability that nonrecipients according to the administrative data report SNAP receipt. Meyer et al. (2018) and Celhay et al. (2018a) examine the role of household characteristics using more precise estimates from larger linked samples.

Overall, we find similar characteristics predict survey error as in the general economic surveys Meyer et al. (2018) and Celhay et al. (2018a) examine. Households with higher incomes and more

---

[15] The estimates are slightly larger if the imputed observations are excluded, so the results do not depend on the recoding of the administrative dates of receipt after the initial interview.

Table 2. The Determinants of False Negative and False Positive Reports, Probit Coefficients and Marginal Effects

	False Negatives		False Positives	
	Coefficient (1)	Marginal Effects (2)	Coefficient (3)	Marginal Effects (4)
One adult, no children	0.2248 (0.3167)	0.0482 (0.0677)	−0.1450 (0.3384)	−0.0034 (0.0078)
One adult with children	−0.1517 (0.3522)	−0.0325 (0.0753)	0.2168 (0.3549)	0.0050 (0.0083)
Multiple adults, no children	−0.3336 (0.2463)	−0.0715 (0.0539)	−0.0708 (0.2883)	−0.0016 (0.0067)
Number of members 18 or older	0.1830 (0.1188)	0.0392 (0.0256)	0.1372 (0.0891)	0.0032 (0.0022)
Number of members under 18	−0.1335* (0.0793)	−0.0286* (0.0172)	0.0469 (0.1090)	0.0011 (0.0026)
Change in household size in last 3 month	−0.0339 (0.2410)	−0.0073 (0.0516)	0.1189 (0.2680)	0.0028 (0.0062)
Rural	−0.1117 (0.1824)	−0.0239 (0.0389)	0.2421 (0.3032)	0.0056 (0.0072)
Black non-Hispanic	−0.0688 (0.2334)	−0.0148 (0.0501)	0.2770 (0.2476)	0.0064 (0.0059)
Hispanic	0.4779** (0.2068)	0.1024** (0.0437)	0.7663** (0.3386)	0.0178* (0.0093)
Male	−0.2121 (0.2107)	−0.0455 (0.0446)	−0.0530 (0.2346)	−0.0012 (0.0054)
Disabled	−0.0223 (0.2744)	−0.0048 (0.0588)	−0.2290 (0.3169)	−0.0053 (0.0077)
Age ≥50	−0.2969 (0.1929)	−0.0636 (0.0411)	0.0688 (0.2965)	0.0016 (0.0070)
High school graduate	0.3329 (0.2066)	0.0713 (0.0447)	−0.4100* (0.2313)	−0.0095* (0.0054)
Some college	0.3353 (0.2122)	0.0718 (0.0459)	−0.3302 (0.2401)	−0.0077 (0.0056)
College graduate and beyond	0.4704* (0.2766)	0.1008* (0.0587)	−0.6097* (0.3111)	−0.0142* (0.0078)
Interviews in English	0.2136 (0.2554)	0.0458 (0.0541)	−0.1862 (0.3593)	−0.0043 (0.0084)
Non-U.S. citizen	0.3648 (0.2752)	0.0782 (0.0587)	−0.5821 (0.3632)	−0.0135 (0.0090)
Household income divided by poverty line	0.1560* (0.0909)	0.0334* (0.0195)	−0.3574*** (0.1343)	−0.0083*** (0.0032)

(Continues)

Table 2. Continued

	False Negatives		False Positives	
	Coefficient (1)	Marginal Effects (2)	Coefficient (3)	Marginal Effects (4)
Household income divided by poverty line squared	−0.0029 (0.0025)	−0.0006 (0.0005)	0.0091*** (0.0030)	0.0002*** (0.0001)
Employed	−0.1853 (0.1919)	−0.0397 (0.0411)	−0.3323 (0.2552)	−0.0077 (0.0063)
Unemployed	−0.5083** (0.2410)	−0.1089** (0.0517)	0.4586* (0.2724)	0.0107* (0.0064)
Reported housing assistance receipt	−0.5920*** (0.2241)	−0.1268*** (0.0478)	0.2905 (0.3054)	0.0068 (0.0070)
Reported WIC receipt	−0.0766 (0.2369)	−0.0164 (0.0508)	−0.0807 (0.2978)	−0.0019 (0.0070)
Reported welfare, child support, alimony receipt	−0.6742*** (0.2180)	−0.1445*** (0.0482)	0.5619* (0.3327)	0.0131 (0.0080)
Days since last payment (incl. Imputed negative dates)	0.0067 (0.0054)	0.0014 (0.0012)		
Days since last EBT card use (incl. Imputed negative dates)	0.0262*** (0.0083)	0.0056*** (0.0017)		
Admin. Payment after interview	−0.1413 (0.2097)	−0.0303 (0.0446)		
Admin. EBT card use date after interview	0.2043 (0.2139)	0.0438 (0.0460)		
Admin. Payment date missing	0.5776** (0.2716)	0.1238** (0.0578)		
Admin. EBT card use date missing	0.0909 (0.4489)	0.0195 (0.0960)		
Constant	−2.1009*** (0.4946)		−1.6200** (0.6832)	
Mean of dependent variable (error rate)	0.183		0.012	
Observations	768		1488	

Notes: Demographic characteristics refer to the respondent. The omitted family type is multiple adults with children, the omitted education category less than high school and the omitted employment category is out of the labor force. All estimates use household weights. Standard errors in parentheses.

***$p < 0.01$
**$p < 0.05$
*$p < 0.1$

education are more likely to be false negatives, but less likely to be false positives. Thus, in line with prior studies, we find these households report less receipt, rather than uniformly report less well. Conversely, those who are unemployed, or report receipt of other programs commit fewer false negative and more false positive errors. They report receipt more frequently. With reductions in false negative rates by 13–15%, reporting other programs strongly predicts correct reporting by SNAP recipients. As in Celhay et al. (2018a), we also find Hispanic respondents commit more false positives and more false negatives, that is, errors in reporting SNAP are more frequent regardless of true receipt status. Contrary to Celhay et al. (2018a), who find this result to be even more pronounced among black respondents, we do not find systematic differences in the error rates of black respondents and the overall population.

Contrary to these prior studies, we only find weak effects of household composition and no differences in reporting by gender, age, or disability status. These last results may be due to less precise estimates from our smaller sample, rather than differences in the nature of the survey error. The signs and magnitudes of most coefficients, even if insignificant, are well aligned with prior studies. Consequently, our results indicate that survey error varies systematically with demographic characteristics, but this relationship is similar in previously validated general economic surveys and a specialized survey like FoodAPS.

Interviewers

Interviewers play a key role in the survey process, so their impact on error rates has received considerable attention in the literature on survey design.[16] Interviewers may affect response accuracy directly. They may decrease accuracy by not properly conducting interviews or conversely improve accuracy by following up on responses that seem likely to be erroneous. Some interviewers may probe more in states where SNAP is known under a different name. They may also have more subtle effects on answers by suggesting certain answers or creating stigma. FoodAPS interviewers were likely aware of the importance of SNAP for the survey and had information on the response to the screener question on SNAP receipt. The degree to which interviewers made use of this information and emphasized the importance of SNAP may have varied between interviewers, so studying interviewer heterogeneity in FoodAPS is particularly interesting.

The FoodAPS data include interviewer identifiers, which enables us to examine how error rates vary between interviewers. We first compute false positive and false negative rates by interviewer. The 2257 (initial) interviews in our sample were conducted by 92 interviewers, but many interviewers only conducted a few interviews. To exclude very noisy estimates of the error rates, we only examine the false negative rates of the 50 interviewers who conducted 7 or more interviews with administrative recipients and the false positive rates of the 69 interviewers who conducted 7 or more interviews with administrative nonrecipients. The excluded interviewers who conducted few interviews have a much higher average false negative rate—16 percentage points higher than the average rate among those with 7 or more interviews.[17] This difference shows that error rates vary between interviewers. Still, we cannot distinguish between several potential reasons for the higher error rates among those with few interviews, which may be due to interviewer experience or interviewer retention and assignment policies. Turning to interviewers with 7 or more interviews, we

---

[16] Bruckmeier et al. (2015) study the role of interviewers in transfer receipt and provide references.

[17] There are no false positives among the interviews with nonrecipients conducted by those with few interviews, but we cannot reject that this error rate is the same as the average false positive rate.

first test whether error rates are equal across interviewers. We reject that false negative rates are equal across interviewers with a *p*-value below 0.0001. We do not reject that false positive rates are equal across interviewers (*p*-value of 0.12). That we do not find any effects on false positives may indicate that false positive rates are not greatly influenced by interviewers, but may also be due to estimation noise from a much lower false positive rate.

Differences in unconditional error rates may be due to interviewer characteristics and behavior but may also arise from systematic differences in the respondents to which interviewers are assigned. We document that error rates vary systematically with respondent characteristics above. FoodAPS interviewers were assigned to households based on geography, so the demographic composition of interviewers' assigned areas likely causes differences in error rates between interviewers. To mitigate this problem, we analyze error rates controlling for demographics by adding interviewer dummies to the multivariate Probit models in Table 2. We estimate the conditional effect of interviewer *i* on the false positive and false negative rate as the marginal effect of the dummy for interviewer *i* from these two models.[18] To mitigate the incidental parameter problem, we include only one dummy for all interviews conducted by interviewers with less than seven interviews.

Controlling for demographics, the false negative rate for interviewers with few interviews remains substantially higher than among those with more interviews: The Probit marginal effects suggest that it is 8 percentage points higher than the average.[19] We now only weakly reject that interviewers with many interviews are homogeneous in their conditional false negative rates with a *p*-value of 0.08. This weak result combined with the large difference between those with few interviews suggests that overall interviewers differ systematically in their probability of producing a false negative response, even when assigned to households with the same demographic characteristics.

This result raises the question whether the differences in error rates between interviewers are large enough to make a substantive difference. Adding the interviewer dummies to an error Probit model with the covariates only substantially improves prediction accuracy.[20] It does not lead to meaningful changes in the estimated coefficients on the covariates of interest. Examining the magnitude of the interviewer effects is complicated, because the estimated interviewer effects confound true differences between interviewers and estimation noise. For false negatives, we find an interquartile range of the estimated interviewer effects large enough to leave scope for true interviewer effects of substantive importance.[21] On the other hand, comparing the dispersion of the estimated effects to the dispersion one would expect based on estimation error alone suggests that this difference could be estimation noise.[22]

Overall, our results provide evidence that interviewers differ systematically in their error rates, even when conditioning on covariates. This finding suggests that interviewers are heterogeneous and

---

[18] Using Probit marginal effects to calculate these error rates means that only the differences in error rates between interviewers can be estimated. Using a linear probability model instead of a Probit does not substantively alter our conclusions in this section.

[19] The results in the remainder of this section are based on models that exclude the interviews conducted by those with few interviews, but including them does not affect the results.

[20] For false negative rates, McFadden's pseudo-R^2 increases from 0.19 to 0.34, an 81% increase. For false positives, it increases from 0.30 to 0.50, a 67% increase.

[21] Excluding interviewers with no false negatives for comparability, the interquartile range is 18 percentage points for the estimated unconditional interviewer effects and 19 percentage points for the estimated conditional effects. Including interviewers with no false negatives yields an unconditional interquartile range of 17 percentage points and a conditional interquartile range of 12 percentage points.

[22] The variance of the estimated conditional interviewer effects is smaller than the average of the variance of the estimation error in all models.

that the differences are not entirely due to interviewer assignment, but our small sample does not provide strong evidence on the nature and magnitude of interviewer effects. We find interviewers with few interviews to have substantial error rates, but the results are neither precise enough to assess differences in the magnitude of the other interviewer effects, nor to establish how much of the variation between interviewers is due to interviewer heterogeneity and how much is due to differences in assignment.

5. Geographic Heterogeneity in Survey Error

Most previous record linkage studies are based on at most a few states. This research strategy makes the question of geographic heterogeneity in survey error crucial for two reasons. First, geographic heterogeneity limits the extent to which findings from one state apply more generally, that is, heterogeneity reduces what we can learn from such single-state studies. Second, if survey error is sufficiently stable across geography, record linkage studies from one state can be used to (partly) correct national estimates by extrapolating to the entire United States. FoodAPS linked data from multiple states, so we can examine the question of geographic heterogeneity directly. A key challenge in the presence of linkage error is to separate geographic variation in reporting from differences between states in the accuracy of the administrative data and linkage. We focus on the sample of states where the linkage process was the same. This choice makes it unlikely that the extent of linkage errors varies systematically across states, but we cannot conclusively distinguish variation in survey error from variation in linkage error.

We first test whether false positive and false negative rates differ between states. We reject that unconditional error rates are equal across the 13 states in our sample with a p-value of 0.002 for false negative rates and 0.003 for false positive rates. Because the extent of survey error depends on demographic characteristics that vary systematically across states, the more relevant question when extrapolating from one state for the purposes of multivariate imputation is the extent of differences after accounting for observable household characteristics (Mittag 2019). When controlling for demographics in a Probit analysis of the errors, we cannot reject that conditional false positive rates are equal (p-value: 0.12) and can only weakly reject that false negative rates are equal (p-value: 0.09).[23]

Examining how state fixed effects affect explanatory power indicates that most, but not all, between state variation is captured by observable covariates. For false negatives, a linear probability model with only state fixed effects explains 21% of the variation in the errors. Adding state dummies to a linear probability model with our demographic characteristics only increases explanatory power by 8%, from 33% to 35%. For false positives, explanatory power is low in all models. State fixed effects alone explain 2.1% of the variation but adding them to a linear probability model with demographic characteristics only increases explanatory power from 6.3% to 7.2%.[24] For comparison, adding the covariates to a linear probability model with state fixed effects leads to a much larger increase in explanatory power, 67% for false negatives and 240% for false positives.

The results so far do not rule out substantively sizeable differences. Analyzing the magnitude of differences raises the problem of separating differences between states from estimation noise. We find a sizeable interquartile range of the estimated false negative rates at 14 percentage points for the

---

[23] Using a linear probability model, the p-values are 0.28 for false positives and 0.14 for false negatives.

[24] In Probit models, adding state fixed effects to a model with covariates only increases McFadden's pseudo R^2 from 0.19 to 0.23 for false negative reports and from 0.3 to 0.36 for false positive reports.

unconditional false negative rates and 12 percentage points for the conditional false negative rates. The interquartile range of the estimated false positive rates is 1.7% unconditionally and 0.8% when conditioning on covariates.[25] Thus, we cannot rule out some meaningful differences. Nevertheless, comparing the variance of the estimates to the variance of the estimation error points toward these differences being estimation noise rather than geographic heterogeneity. In line with documenting unconditional heterogeneity, the variance of the estimates is larger than the average variance of the estimation error for both unconditional error rates. Yet, for both conditional error rates, the average variance of the estimation error exceeds the variance of the estimates by a factor of two or more.[26]

For researchers who want to obtain national estimates for the entire United States, this raises the question whether the loss of accuracy from using the error-prone survey data or the loss of accuracy from extrapolating across geography from a record linkage study in the presence of the heterogeneity we document is worse (Mittag 2019). To directly assess how this choice affects estimate accuracy and to further quantify the substantive importance of geographic heterogeneity, we conduct a leave-one-out extrapolation exercise. For each of the 13 states in our sample, we use the data from the other 12 states to estimate a Probit model of administrative SNAP receipt that conditions on our covariates and reported SNAP receipt. We use this model to predict the probability of SNAP receipt for every household in the left-out state. We estimate the SNAP receipt rate for each state as the average of these predicted probabilities.

We quantify the loss of accuracy by calculating mean squared error (MSE) of the state-specific receipt rates according to the linked administrative variable. The MSE from this extrapolation exercise is 0.000279. Estimating state-level SNAP receipt rates based on the predicted probabilities of SNAP receipt from a model estimated using data from all states yields an MSE of 0.000152. Thus, the increase in MSE from extrapolation (and using one state less in the estimation) is 0.000128, an 84% increase in error. A difference in MSE of the same magnitude would also arise from increasing the average variance of the estimation error in the state-specific receipt rates by 17%. This loss of precision roughly corresponds to reducing the number of households per state by 30.

In comparison, using the survey reports without any extrapolation yields an MSE of 0.000291. This is a 92% increase over state-level predictions without extrapolation and corresponds to an increase of the variance of the estimation error by 19% or a reduction of the sample by 33 households per state. Consequently, the loss of precision due to survey error is of a similar order of magnitude to the loss due to extrapolation. It should be kept in mind that a large part of the MSE from extrapolation is prediction variance, rather than bias from geographic heterogeneity. This added variance from extrapolation decreases with sample size, but is still large in a small survey like FoodAPS. We cannot separate prediction variance from bias, but the prediction variance appears to be large, as even the in-sample predictions from the survey reports lead to an MSE of 0.011, a four-fold increase over the survey reports. Yet, despite the large prediction variance due to the small sample and the relatively accurate reporting in FoodAPS, extrapolation across geography still yields a 4% increase in estimate accuracy and thereby improves over the survey reports.

---

[25] The conditional interquartile ranges are based on Probit marginal effects, when using a linear probability model, the range is unchanged at 12 percentage points for false negatives and slightly higher at 1.1 percentage points for false positives. The unconditional error rates are based on a linear probability model.

[26] For false negatives, this holds for both Probit and linear proability models. For false positives, the estimation error has a slightly smaller variance than the estimates, but both are smaller than 0.01 percentage points.

6. Conclusion

We use FoodAPS survey data linked to administrative SNAP records to examine the extent and nature of survey error and how errors differ between this specialized survey that emphasizes SNAP and nutrition programs and the general economic household surveys that prior studies have validated. We specifically focus on assessing the role of survey design and use the linked data from multiple states that FoodAPS provides to examine geographic heterogeneity.

We find false negative rates to be lower in FoodAPS at 18.3% than in the ACS and CPS and slightly lower than in the SIPP. Our results do not provide any evidence of false positive rates being higher in FoodAPS at 1.2% than in other surveys. Recall problems are an important cause of survey error. The shorter and simpler recall period of FoodAPS explains part of the lower false negative rate in FoodAPS compared to other surveys. The ability to recall SNAP receipt seems to be affected by time since last SNAP use rather than time since last SNAP receipt. This finding suggests that FoodAPS policy of asking the main food shopper may have contributed to lower false negative rates as well. We find that similar demographic characteristics predict reporting errors as prior studies. Both recipient and nonrecipient households that are poorer, have a less educated or unemployed respondent or report receipt of other programs are more likely to report SNAP receipt. Households with a Hispanic respondent report less well, that is, recipient households are more likely to fail to report receipt while nonrecipient households are more likely to report than are other households.

We document that error rates vary systematically between interviewers. Only part of this variation is explained by the demographic characteristics that explain survey error. This result suggests that interviewer heterogeneity affects survey accuracy, but our results on the substantive magnitude of interviewer differences are imprecise due to the small sample. We find that interviewers who only conduct a few interviews have substantially higher false negative rates than the average interviewer. Analyzing conditional variation in the error rates between frequent interviewers does not yield much evidence of heterogeneity beyond estimation noise, but also does not rule out interviewer effects of substantive importance. Finally, we examine geographic heterogeneity to gauge the generality of findings from prior linkage studies based on one state and to test extrapolation. We find evidence of unconditional differences in both false positive and false negative rates, but most or all of the geographic heterogeneity is due to differences in demographic characteristics. In line with this finding, extrapolation across geography in FoodAPS still improves accuracy over the survey reports, but only narrowly so.

Our results clearly show that survey design features such as the recall period and the choice of interviewers affect survey error. Yet, the results do not point to major differences in the nature of survey error between the previously validated general economic surveys and a survey with particular focus on SNAP and nutrition programs like FoodAPS. Error rates are a bit lower, which may be due to higher salience or interviewer training in a specialized survey. Much of the difference in error rates appears to be due to the shorter recall period rather than the specialization of the survey. We find that the demographic characteristics that predict survey errors in FoodAPS are similar to those found in prior studies, suggesting some generality to potential biases implied by past studies. We also find that while there is unconditional geographic heterogeneity, survey error appears to be similar across states conditional on demographic characteristics. Thus, we provide the first evidence that findings on the nature and correlates of survey error from one state are likely informative about survey error in other states.

This study also clearly shows that linked administrative data can have design complications and errors that mean they are not a panacea for the problem of survey error. Nevertheless, carefully analyzing the differences between the linked administrative variable and the survey reports can still

reveal valuable insights on the extent and nature of survey error, even in the presence of errors in both the administrative and the survey variable. These insights can help researchers, policy makers and survey producers to improve both survey production and usage. A better understanding of the nature of survey error can help survey producers to improve and researchers to gauge the reliability of their data. It can also help researchers to assess the robustness of their results to survey error by incorporating information on the nature of survey error in simulations (Millimet 2011) or bounds (Gundersen and Kreider 2008; Jensen et al. 2018). The models of survey error we estimate can also be used in corrections for measurement error. They provide information regarding the validity of common corrections, which often require errors to be independent of the true value of the variable or not to be predicted by a specific instrumental variable. The estimates from these models could also be used in validation-data based corrections that allow for arbitrary measurement error (Schenker et al. 2010; Davern et al. forthcoming; Mittag 2019). As the latter two papers discuss, information from one validation study can be used to correct for survey error in a different survey or sample if the errors are sufficiently similar. We document differences and similarities in survey error both across surveys and across states. A better understanding of the variation in survey error can help researchers decide whether corrections based on similar data are likely to improve estimates in specific cases. Thereby, our results provide useful information to improve the accuracy of survey data as well as estimates derived from them.

Appendix

FoodAPS Interviews

FoodAPS conducted two interviews with the main food shopper of the household. The initial interview collected information on household composition, demographics, and education. It included separate sections on SNAP and other food assistance programs as well as questions on usual food acquisition and consumption. After the initial interview, all members 11 years and older were asked to track food acquisition for a week using a diary and three phone calls to the FoodAPS study center to report food away from home. The final interview was conducted after these data had been collected, that is, roughly one week after the initial interview. The final interview collected information on consumption, expenditures, and income as well as food acquisition, preparation, and health-related questions.

Data Linkage

For our analysis sample, the two administrative data sources (caseload and ALERT records) were linked deterministically. The FoodAPS survey data were linked to the combined administrative data using the probabilistic linkage procedure of the caseload data described in Section 2. Among all states that provided caseload data, SNAP payments from the caseload data were matched to 1244 households. One hundred and thirty-six of these households were matched to multiple SNAP cases, either because their SNAP case identifier changed or because there are two SNAP cases in the same household. Of these matching households, 240 matches deemed uncertain by the probabilistic record linkage were matched based on a manual review to account for common differences in reporting of addresses and phone numbers (e.g., omitting the area code of the telephone number). The states that are not included in our sample provided ALERT data that needed to be linked to the survey data using a separate, second linkage procedure. This procedure matched transactions to respondents based on the store identifier, the amount spent, and the date of the transaction from the ALERT data and the responses to the food at home event questions. As pointed out earlier, to ensure comparability, we only analyze states that provided identical identifiers in the caseload and ALERT data and thus do not rely on data that uses this second probabilistic linkage procedure. In addition to making comparisons of survey errors across states questionable, the probabilistic linkage of the ALERT data appears to introduce additional errors in the linked administrative variable. For example, we find state-level false positive rates of up to 11% among these states, which points to problems in the administrative records or their linkage. A likely reason for this is that if neither SNAP transactions nor non-SNAP transactions at stores accepting SNAP EBT cards were reported during the week after the

Table A1. Alternative Definitions of Reported and Administrative SNAP Receipt

Receipt Definition	Number of Nonrecipient Obs. (1)	Number of Recipient Obs. (2)	Number of Obs. Recoded to Be Recipients (3)	Number of Obs. Recoded to Be Nonrecipients (4)	Receipt Rate (5)	Change in Receipt Rate (6)	False Positive Rate (7)	Change in FP Rate (8)	False Negative Rate (9)	Change in FN Rate (10)
Definitions of reported receipt										
1. Reported receipt ≤32 days prior to interview	1566	691	0	0	11.70%	0.00%	1.20%	0.00%	18.33%	0.00%
2. Reported receipt ≤31 days prior to interview	1570	687	0	4	11.64%	−0.06%	1.16%	−0.04%	18.55%	0.22%
3. Reported receipt ≤33 days prior to interview	1563	694	3	0	11.74%	0.04%	1.20%	0.00%	17.99%	−0.34%
4. Current receipt (SNAPNOWREPORT)	1519	738	53	6	12.71%	1.02%	1.43%	0.23%	12.04%	−6.29%
Definitions of administrative receipt										
1. Payment or EBT card use ≤32 days prior to interview	1489	768	0	0	13.04%	0.00%	1.20%	0.00%	18.33%	0.00%
2. Payment or EBT card use ≤31 days prior to interview	1489	768	0	0	13.04%	0.00%	1.20%	0.00%	18.33%	0.00%
3. Payment or EBT card use ≤33 days prior to interview	1489	768	0	0	13.04%	0.00%	1.20%	0.00%	18.33%	0.00%
4. Verified nonreceipt only (SNAPNOWHH)	1457	800	45	13	13.78%	0.74%	0.20%	−1.00%	16.35%	−1.98%
5. Payment ≤32 days prior to interview	1608	649	0	119	10.44%	−2.60%	3.13%	1.93%	14.79%	−3.53%
6. Payment or card use including all dates after interview	1489	785	0	17	13.69%	0.65%	1.20%	0.00%	22.71%	4.39%
7. Payment or card use, ignoring dates after interview unless date in other admin data source indicates current receipt	1598	659	0	109	11.13%	−1.91%	2.67%	1.47%	16.21%	−2.12%

Notes: This table summarizes how other definitions of SNAP receipt differ from our definition and how they affect rates of receipt and misreporting. The upper panel summarizes alternative ways to define SNAP receipt based on survey reports, the lower panel summarizes alternative definitions based on administrative information. The first definition in both panels is the definition we use in this article. The next two rows in each panel change the reference period by one day. Row 4 uses the variable SNAPNOWREPORT, which is defined as the survey response to the question about current receipt with ambiguous reference period. In the lower panel, row 4 uses SNAPNOWHH, which considers those who report receipt to be recipients unless recent receipt and current nonreceipt is indicated by the administrative data. Row 5 uses only the payment dates from the administrative data to define SNAP receipt. Row 6 includes the 17 observations with receipt dates after the interview for which current receipt cannot be verified. Row 7 considers those with caseload or ALERT dates after the interview date that are not confirmed to be recipients by the respective other data source to be nonrecipients. The first two columns provide observation counts of recipients and nonrecipients. Columns 3 and 4 provide observation counts of the number of recipient and nonrecipient households whose receipt status differs from our definition. The remaining columns provide (weighted) estimates of the rates of receipt and misreporting according to each variable definition as well as the difference (in percentage points) to the respective rate according to our definition.

Table A2. Frequency of Most Recent Recorded Administrative SNAP Payment and EBT Card Use After the Initial Interview

Number of Days Since Initial Interview	Number of Payment Cases	Number of EBT Card Use Cases
1	14	16
2	9	8
3	11	28
4	17	12
5	23	36
6	17	25
7	14	22
8 to 31 days	24	24
More than 31 days	17	27

Notes: Unweighted counts of observations with caseload payment dates or ALERT EBT card use dates after the day of the initial interview, when the question regarding SNAP receipt was asked.

Table A3. Summary Statistics for our Analysis Sample

	Mean	SE
Administrative SNAP receipt	0.34	0.01
Reported SNAP receipt	0.306	0.01
One adult, no children	0.216	0.009
One adult, with children	0.066	0.005
Many adults, no children	0.284	0.009
Number of members 18 or older	2.074	0.021
Number of members under 18	0.952	0.027
Change in household size in last three months	0.101	0.006
Rural	0.306	0.01
Black non-Hispanic	0.121	0.007
Hispanic	0.262	0.009
Male	0.259	0.009
Disabled	0.057	0.005
Age ≥ 50	0.392	0.01
High school graduate	0.282	0.009
Some college	0.33	0.01
College graduate and beyond	0.186	0.008
Interviews in English	0.863	0.007
Non-U.S. citizen	0.136	0.007
Household income divided by poverty line	2.27	0.053
Household income divided by poverty line squared	11.38	1.277
Employed	0.469	0.011
Unemployed	0.086	0.006
Reported housing assistance receipt	0.108	0.007
Reported WIC receipt	0.107	0.007
Reported welfare, child support, alimony receipt	0.086	0.006
Days since last payment (incl. Imputed negative dates)	8.932	0.519
Days since last EBT card use (incl. Imputed negative dates)	7.181	0.411
Administrative payment after interview	0.057	0.005
Administrative EBT card use date after interview	0.081	0.006
Administrative payment date missing	0.661	0.01
Administrative EBT card use date missing	0.65	0.01
Number of observations	2256	

Notes: Our analysis sample only includes the 13 states that provided caseload and ALERT data with linkable identifiers. From this sample, we exclude 21 households that did not provide consent to linkage and 17 households with unresolved receipt status due to dates of last payment or use recorded after the initial interview. Demographic characteristics refer to the respondent. All estimates use household weights.

interview, no probabilistic match to ALERT data was possible or attempted. Many households only use their SNAP benefits shortly after disbursement and are thus unlikely to use the EBT card every week. In addition, households may not have reported all EBT transactions in the survey. Thus, the additional ALERT link likely misses a substantial fraction of households with transactions in the month before the interview.

SNAP Receipt Variable Combining Administrative and Survey Information (SNAPNOWHH)

The FoodAPS data set includes a SNAP receipt variable, SNAPNOWHH, that is a combination of administrative and survey data on reported receipt. SNAPNOWHH classifies households as recipients if either caseload or ALERT data indicate receipt in the 36 days prior to the end of the survey week. Thus, it includes a slightly shorter period before the first interview than our administrative variable and adopts a more arbitrary classification rule for observations with dates of receipt after the interview. SNAPNOWHH also replaces the missing administrative information with survey reports for the households that did not provide consent to linkage, but we exclude these households from our sample. The main problem in using SNAPNOWHH to study survey error is the way it combines administrative and survey reported receipt. For those who report SNAP receipt, SNAPNOWHH overwrites the administrative receipt status with survey reported receipt status. The only exception to this rule is households who report receipt, but the administrative data verify prior, but not current receipt. Thereby, SNAPNOWHH rules out survey over-reporting of SNAP by households who never received SNAP or received it before the administrative records start in April 2012. The only way a household can commit a false positive error using SNAPNOWHH is when the administrative data confirm recent, but not current receipt. This situation occurs for six households in our sample, so comparing SNAPNOWHH to survey reports amounts to comparing survey reports to survey reports for almost all households that report SNAP receipt. Such an approach obviously mutes many differences between the two data sources, some of which are very likely due to survey error. For example, the false positive rate effectively drops to zero when using SNAPNOWHH as the administrative SNAP receipt variable. While a large fraction of overreporting is likely due to recent recipients (Celhay et al. 2018a), it is unlikely that *all* overreporting is due to recent recipients. Therefore, we compare a purely administrative-data based variable to survey reports, even though this likely overstates false positives due to failures to link some who truthfully report receipt. This possibility should be kept in mind when interpreting our results on false positives, which combine true false positives with linkage failures. Meyer et al. (2018) discuss the consequences of such errors.

References

Bollinger, C. R., and M. H. David. 1997. Modeling discrete choice with response error: Food stamp participation. *Journal of the American Statistical Association* 92(439):827–35.

Bound, J., C. Brown, and N. Mathiowetz. 2001. Measurement error in survey data. In *Handbook of econometrics*, volume 5, edited by J. J. Heckman and E. Leamer . Amsterdam: Elsevier, Chapter 59, pp. 3705–843.

Bruckmeier, K., G. Mueller, and R. T. Riphahn. 2015. Survey misreporting of welfare receipt—Respondent, interviewer, and interview characteristics. *Economics Letters* 129(April):103–7.

Celhay, P., Meyer, B.D. and Mittag, N. 2018a. Errors in reporting and imputation of government benefits and their implications. Unpublished manuscript.

Celhay, P., Meyer, B.D. and Mittag, N. 2018b. What leads to measurement error? Evidence from reports of program participation in three surveys. Unpublished manuscript.

Cerf Harris, B. 2014. Within and across county variation in SNAP misreporting: Evidence from linked ACS and administrative records. U.S. Census Bureau CARRA Working Paper 2014-05.

Copas, J. B., and F. J. Hilton. 1990. Record linkage: Statistical models for matching computer records. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 153(3):287–320.

Courtemanche, C., Denteh, A. and Tchernis, R. 2018. Estimating the associations between SNAP and food insecurity, obesity, and food purchases with imperfect administrative measures of participation. NBER Working Paper No. 24412.

Davern, M., B. D. Meyer, and N. Mittag. forthcoming. Creating improved survey data products using linked administrative-survey data. *Journal of Survey Statistics and Methodology*.

Fellegi, I. P., and A. B. Sunter. 1969. A theory for record linkage. *Journal of the American Statistical Association* 64(328): 1183–210.

Groves, R. M., F. J. Fowler, M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau. 2009. Survey methodology. In *Wiley series in survey methods*. New York: John Wiley and Sons.

Groves, R. M., and L. Lyberg. 2010. Total survey error: Past, present, and future. *Public Opinion Quarterly* 74(5):849–79.

Gundersen, C., Kreider, B. 2018. Food stamps and food insecurity what can be learned in the presence of nonclassical measurement error? *Journal of Human Resources*, 43(2):352–82.

Jensen, H., Kreider, B. Zhylyevskyy, O. 2018. Causal effects of multiple food assistance program participation on child food insecurity. Working Paper.

Kang, K.M. and Moffitt, R.A. 2018. The effect of SNAP and school food programs on food security, diet quality, and food spending: Sensitivity to program reporting error. Unpublished manuscript.

Kirlin, J.A., and Wiseman, M. 2014. Getting it right, or at least better: Improving identification of food stamp participants in the national health and nutrition examination survey. Working Paper.

Marquis, K. H., and J. C. Moore. 1990. Measurement Errors in SIPP Program Reports. In *Proceedings of the 1990 Annual Research Conference*. Washington, D.C.: U.S. Bureau of the Census, pp. 721–45.

Meyer, B.D. and Mittag, N. 2018. An empirical total survey error decomposition using data combination. NBER Working Paper 25737.

Meyer, B. D., and N. Mittag. 2019. Using linked survey and administrative data to better measure income: Implications for poverty, program effectiveness and holes in the safety net. *American Economic Journal: Applied Economics* 11(2): 176–204.

Meyer, B.D., Mittag, N. and Goerge, R. 2018. Errors in survey reporting and imputation and their effects on estimates of food stamp program participation. NBER Working Paper 25143.

Meyer, B. D., W. K. C. Mok, and J. X. Sullivan. 2015. Household surveys in crisis. *Journal of Economic Perspectives* 29(4):199–226.

Millimet, D. 2011. The elephant in the corner: A cautionary tale about measurement error in treatment effects models. In *Missing data methods: Cross-sectional methods and applications, Advances in Econometrics*, volume 27, edited by D. Drukker. Bingley, UK: Emerald Publishing.

Mittag, N. 2019. Correcting for misreporting of government benefits. *American Economic Journal: Economic Policy* 11(2): 142–64.

Schenker, N., T. E. Raghunathan, and I. Bondarenko. 2010. Improving on analyses of self-reported data in a large-scale health survey by using information from an examination-based survey. *Statistics in Medicine* 29:533–45.

Sudman, S., and N. M. Bradburn. 1973. Effects of time and memory factors on response in surveys. *Journal of the American Statistical Association* 68(344):805–15.

Taeuber, C., Resnick, D.M., Love, S.P., Stavely, J. Wilde, P. and Larson, R. 2004. Differences in estimates of food stamp program participation between surveys and administrative records. Working Paper, U.S. Census Bureau.

U.S. Department of Agriculture, Economic Research Service (USDA ERS). 2016a. *National Household Food Acquisition and Purchase Survey (FoodAPS): User's guide to survey design, data collection, and overview of datasets*. Washington, D.C.: U.S. Department of Agriculture.

U.S. Department of Agriculture, Economic Research Service (USDA ERS). 2016b. *National Household Food Acquisition and Purchase Survey (FoodAPS): Codebook: Household-level public use file, faps_household_puf*. Washington, D.C.: U.S. Department of Agriculture.

Winkler, W. E. 2014. Matching and record linkage. In *Wiley interdisciplinary reviews: Computational statistics*, volume 6. New York: John Wiley and Sons, pp. 313–25.