SSOAR

Open Access Repository

www.ssoar.info

Do supplemental list frames for subpopulations increase subpopulation sampling efficiency? Evidence from the National Household Food Acquisition and Purchase Survey

Zhang, Shiyu; Kirlin, John A.; Page, Elina T.; Zhang, Xingyou; West, Brady T.

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

gesis
Leibniz-Institut
für Sozialwissenschaften

Mitglied der
Leibniz-Gemeinschaft

Do supplemental list frames for subpopulations increase subpopulation sampling efficiency? Evidence from the National Household Food Acquisition and Purchase Survey

Shiyu Zhang, Survey Research Center, Institute for Social Research, University of Michigan-Ann Arbor, Michigan, USA
John A. Kirlin, Kirlin Analytic Services, Fredericksburg, VA, USA
Elina T. Page, Economic Research Service, United States Department of Agriculture, Washington, DC, USA
Xingyou Zhang, U.S. Bureau of Labor Statistics, Washington, DC, USA
Brady T. West, Survey Research Center, Institute for Social Research, University of Michigan-Ann Arbor, Michigan, USA

31.08.2020

Abstract

Multiple-frame sampling has been regarded as a device for increasing efficiency in identifying small subpopulations. However, there has been a lack of empirical evidence in supporting the efficiency of the multiple-frame approach and in guiding best practices. The current study focuses on a special scenario in which two frames were used to recruit sample members. Using paradata from the U.S. National Household Food Acquisition and Purchase Survey (FoodAPS), the current analysis focuses on recruiting households that received Supplementary Nutrition Assistance Program (SNAP) as a sub-goal of the survey sampling. SNAP households account for around one-fifth of the general U.S. population, compared to a survey goal of 30 percent of responding households. Our findings were consistent with theoretical expectations. Having and using additional SNAP list frames improved the efficiency of identifying SNAP households as opposed to screening a general address-based sample frame. This efficiency remained even as the SNAP list frames aged.

Keywords

Dual-frame sampling, recruitment efficiency, sampling efficiency, small subpopulations

Introduction

Surveys are often designed to produce estimates for different subpopulations. When the subpopulations are small, screening the entire population with a general frame may not be efficient for identifying eligible members of the subpopulation (Cowan, 1991; Kalton, 2009; Kalton & Anderson, 1986; Wagner & Lee, 2015). In response to this challenge, Hartley (1962, 1974) proposed that one could improve recruitment efficiency by adopting a multiple-frame approach and sampling subpopulation members from a separate list frame that is incomplete but contains a high percentage of eligible subpopulation members. The additional frame may boost the sample size of this subpopulation by making the recruitment less sporadic and thus more efficient (Kalton, 2003; Kalton, 2009; Kalton & Anderson, 1986; Lohr, 2009; Wagner & Lee, 2015). In the current study, we focus on a dual-frame setting and investigate whether having and using a list frame indeed provides extra benefits compared to screening a sample selected from a general frame.

The idea of using multiple frames to oversample small subpopulations has received substantial theoretical attention since its introduction and is currently implemented in several surveys (e.g., Bricker et al., 2017; Kalton, 2009; Lohr, 2009; Mecatti & Singh, 2014; Metcalf & Scott, 2009; Wagner & Lee, 2015; Srinath et al., 2004; Weerasekera et al., 2019). A great deal of research has been devoted to the development of point and variance estimators under the multiple-frame design (Chen, Stubblefield, & Stoner, 2020; Lohr & Rao, 2000; Lohr & Rao, 2006; Mecatti, 2007; Mecatti & Singh, 2014; Ranalli, Arcos, Rueda, & Teodoro, 2016; Rao & Wu, 2010; Sánchez-Borrego, Arcos, Rueda, 2019). Despite these important advances, very few studies have provided empirical support for the original motivation of adopting the multiple-frame approach. There is a dearth of evidence on whether the additional list frame can indeed increase operational and cost efficiency in identifying subpopulation members in comparison to screening the general population.

While it is certainly reasonable to assume that using a supplementary list frame would increase efficiency in identifying subpopulation members, this may not be unconditionally true. For instance, depending on the source of the list frame, the information could be inaccurate (West, Wagner, Hubbard, & Gu, 2015), which would largely undermine the usefulness of the additional frame. Likewise, depending on the specific subpopulation of interest, the group status could be time sensitive, which would make the list frame age quickly and lose its usefulness. Further, there may be a cost associated with obtaining the list frame, and this cost needs to be weighed against the efficiency benefits introduced in the sampling process. Thus, more empirical evidence is needed to support the assumptions and practices of adopting

a multiple-frame sample design with the intention of increasing efficiency in recruiting subpopulation members.

The Current Study

The current study investigates the utility of supplementary list frames in the U.S. National Household Food Acquisition and Purchase Survey (FoodAPS). Using paradata, we examine whether obtaining and using a list frame of recipients of the Supplementary Nutrition Assistance Program (SNAP) enabled more efficient identification of SNAP households in the FoodAPS recruitment process.

Based on the 2010 census enumeration (Lofquist, Lugaila, O'Connell, & Feliz, 2012), households with members receiving SNAP accounted for about 19% of all U.S. households (Gray & Eslami, 2014). With this proportion, while SNAP households are certainly a subpopulation, they are not necessarily rare and thus a general screening may suffice (Kalton, 2009; Kish, 1987). Since the added benefit of having an additional list frame depends on how rare the subpopulation is in the general population (Srinath, Battaglia, & Khare, 2004), we investigate whether the additional list frames brought additional benefit in sampling SNAP households, especially considering that these frames had aged and were not perfectly accurate.

Lists of SNAP recipient addresses were obtained from state SNAP agencies as supplemental list frames to boost the subsample size of SNAP households. As we explain in the Methodology section, only some states provided the list to be used as an additional frame; additionally, even in the states that did offer a list of SNAP recipients, the list was not consistently used throughout the entire data collection. This variation in having and using the SNAP list frames allows us to answer our first research question:

Did using SNAP list frames increase the efficiency in identifying SNAP households, in comparison to relying on general screening?

We defined *efficiency* in two ways. First, leveraging the fact that the data collection in different months and areas identified different numbers of SNAP households, field operations that identified a larger number of SNAP households are regarded as more efficient in recruiting SNAP households. Second, given that the number of contact attempts is a proxy of recruitment effort, we regard field operations that directed a larger percentage of total contact attempts to SNAP households as more efficient and goal-oriented in recruiting SNAP households from an effort-per-completed SNAP interview standpoint.

The aging of the list frame was another concern. Even if the list frames were accurate when they were freshly obtained from the state SNAP agencies in January 2012, they may have aged quickly. Being a SNAP participant is a fluid status. Over the four years from 2009-2012, 30% of SNAP participants were in the program for fewer than 12 months (Irving & Loveless, 2015). A substantial number of individuals on the initial list of SNAP participants may no longer be eligible in the latter period of the field operation. Even if participation lasted longer, household mobility might reduce the effectiveness of a list of addresses as time passes. These considerations motivate our second research question:

Did the usefulness of the SNAP list frames decrease over time as the frame aged?

Methodology

FoodAPS overview

In 2012 and 2013, FoodAPS collected information on foods acquired and their quantities and prices for a nationally representative sample of U.S. households. Households with a member receiving SNAP benefits were classified as one of the target groups that the survey intended to oversample.

FoodAPS had a three-stage sampling design. In the first stage, 50 primary stage units (PSUs) (U.S. counties or groups of contiguous counties) were selected. The 50 selected PSUs were located in 27 States; Alaska and Hawaii were excluded from the sample universe. In the second stage, within each of the 50 sampled PSUs, eight secondary sampling units (SSUs) were selected. The SSUs were defined as census block groups or groups of contiguous census block groups. The fact that the SSUs consisted of census block groups allowed us to draw information from the 2012 Census Planning Database (CPD) to characterize these SSUs and construct the covariates. In the third stage, addresses were sampled from the SSUs for screening interviews. The address selection utilized two frames—a commercial list of addresses obtained from the United States Postal Service (CDSF) and a list of SNAP household addresses obtained from state SNAP agencies. The two frames were first merged to identify duplicate addresses, which were then excluded from the postal service file to create two mutually exclusive frames.

The selected PSUs were located in a total of 27 states, but only 22 of the state SNAP agencies provided a list of SNAP addresses in time to be included in the sampling process. Even though the address lists were provided to the FoodAPS contractor at no financial cost, extracting and providing the lists placed an unknown cost on the already burdened state agencies. The difference in SNAP frame availability motivated our analyses: was the process of identifying SNAP households more efficient when a SNAP frame was available?

For determining household eligibility for participation, a 5-minute screening interview was conducted with sampled households at 13,445 addresses. Field interviewers did not know from which frame a particular address had been sampled, in part because there was no guarantee that the original occupants had not moved or changed program status by the time the screening interview was completed. For each address, field interviewers made up to eight in-person contact attempts at different times of the day and on different days of the week to try screening the household. Out of the 13,445 addresses, the eligibility/ineligibility status was determined for 11,812, and these 11,812 households with confirmed status were the basis of our analyses. Of the 11,812 households, 2,375 of them (20.1%) were SNAP households. For detailed information about the recruitment process of FoodAPS, please refer to FoodAPS documentation at www.ers.usda.gov/foodaps.

Dataset construction

The focus of the current study is the recruitment process for SNAP households during the third stage of sampling. To evaluate the ease of identifying SNAP households, we tracked the screening activities for each month of survey operations within each SSU. Each row of our

analytical dataset represents one month of screening activity within one SSU and is referred to as an *SSU-month unit*. There are 2,319 SSU-month units in total available for our analyses. Our two outcome variables were defined at this level using the screening activities associated with SNAP households within each SSU-month unit (see below).

Different SSUs have different concentrations of SNAP participants. Areas in the 22 states that provided the additional SNAP frame could naturally differ from areas in the other 5 states that did not provide the SNAP frame. Thus, the ability to find SNAP households efficiently during the screening process could be confounded by areal characteristics. To control for these potential confounding effects, we extracted information from the 2012 Census Planning Database (CPD) to characterize the SSUs and used these SSU characteristics as covariates in our models (see below). We selected two types of variables that are likely to be associated with the outcome of recruiting SNAP households: 1) variables that reflect areal socioeconomic status and thus are related to concentration of SNAP households, and 2) variables that reflect the difficulty of approaching and recruiting respondents in that area.

Outcome variables

Number of SNAP households

Our first outcome variable is the number of self-reported SNAP households screened in each SSU-month unit. However, directly comparing the number of screened SNAP households across SSU-month units would not be appropriate because the overall number of households screened varied across different SSU-month units. If some SSU-month units screened more households in general (i.e., more exposure), then we would reasonably expect them to have screened more SNAP households specifically (i.e., more event counts). Therefore, when using the number of SNAP households as the outcome variable, we coupled it with the total number of households screened in each SSU-month unit to *offset* the fact that more households being screened could mean more SNAP households being screened. By including the offsets and thus incorporating the exposure underlying the event counts, we are modeling the *rate* of screening SNAP households, with the number of screened SNAP households being the numerator and the number of total screened households being the denominator.

Number of contact attempts associated with SNAP households

The second outcome variable is the number of contact attempts associated with SNAP households in each SSU-month unit. For simplicity, we refer to this variable as the "number of SNAP contact attempts". Parallel to the discussion above about the *number of SNAP households*, when using the *number of SNAP contact attempts* as the outcome variable we include the number of contact attempts for all households in that SSU-month unit as an offset variable. This compensates for the fact that more contact attempts associated with SNAP households may have resulted from more contact attempts associated with all households in those SSUs.

Covariates

### *Percentage of screened households coming from a SNAP frame* (i.e., *SNAP frame usage*)

For evaluating the usefulness of the SNAP frames, we constructed a variable—the percentage of screened households coming from a SNAP frame in each SSU-month unit—to indicate not only whether the SNAP frame was available, but also whether the SNAP frame was in active use in each SSU-month unit. We note that *households coming from a SNAP frame* do not equate to *SNAP households* because the SNAP frames were not perfectly accurate or timely.

The values of this covariate range from 0 to 100 percent. Three scenarios underlie these values:

1) For data collection in the five states that did not provide a SNAP frame, their corresponding SSU-month units all have a value of 0 for this covariate because none of their screened households could come from a SNAP frame.

2) Although the other 22 states provided a list of SNAP addresses to be used as a supplementary frame, not all SSUs in these 22 states were actively using the SNAP frame in every month. For the 1,939 SSU-month units corresponding to these 22 states, 802 of the SSU-month units had 0% of screened households coming from the SNAP frame.

3) The remaining 1,137 SSU-month units had more than one screened household coming from the SNAP frame.

The three scenarios and value assignments of the covariate *percentage of screened households coming from a SNAP frame* are summarized in Figure 1. This variable was rescaled so that one unit corresponds to 10 percentage points. For visualization purposes only (Figure 2 below), we constructed a separate categorical indicator to indicate these three scenarios. We refer to this categorical indicator as *SNAP frame availability/usage indicator*.

Figure 1. A graphical explanation of the three scenarios and value assignments for the covariate *percentage of screened households coming from a SNAP frame*
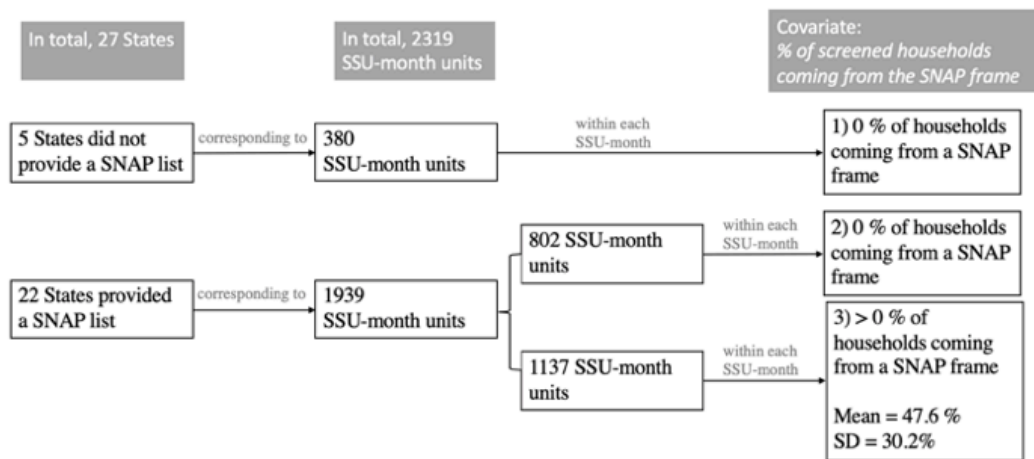
Time

The FoodAPS was fielded for ten months from mid-April 2012 to mid-January 2013. We examined whether time had a moderating effect on the usefulness of the SNAP frame. Time was included as a categorical variable. We collapsed the 10 data collection months into three groups: the first group from April to June (reference category); the second group from July to September; and the third group from October to January.

SSU characteristics

The SSUs in FoodAPS's second-stage sampling consisted of one or more contiguous census block groups. The 2012 CPD contains characterizing information for each of the census block groups and provides denominators that can be used to turn counts into ratios. To characterize the FoodAPS SSUs, we either aggregated the counts across census block groups to attain SSU-level counts which were then turned into ratios, or we took a weighted average of block groups' characteristics within each SSU. For a detailed description of how the covariates were constructed, please refer to Appendix A. All the SSU characteristics were centered on their means when included in the regression models as predictors. All the SSU characteristics that are percentages were rescaled so that one unit corresponds to 10 percentage points. Table 1 below summarizes the SSU characteristics:

Table 1. The SSU characteristics used as covariates

1. Percentage of people with a college degree or higher	the percentage of people with at least a college degree in each SSU, where a higher value would suggest higher socio-economic status for that SSU
2. Percentage of people who are below poverty level	the percentage of people who were classified as living below the federal poverty level in each SSU given their total family income within the last year, family size, and family composition
3. Percentage of people who moved	the percentage of people who moved from another residence in the US or Puerto Rico within the last year in each SSU
4. Percentage of non-English speaking households	the percentage of households in which no one aged 14 years and older spoke English very well in each SSU
5. Percentage of households that receive public assistance	the percentage of households that received public assistance income (i.e., Temporary Assistance to Needy Families or a state's general assistance program) in each SSU
6. Percentage of renter households	the percentage of households that were not owner-occupied in each SSU
7. Median income	the annual median household income of each SSU
8. Mail return rate	the number of 2010 census mail returns received out of the total number of valid occupied housing units (intended to capture the general survey cooperativeness of each SSU)

Statistical Analysis

We fitted zero-inflated negative binomial regression models using the *number of SNAP households* and the *number of SNAP contact attempts* as the outcome variables, respectively. Both outcome variables are count variables with a substantial portion of zero values (49%).

The zero-inflated negative binomial regression model is a mixture model consisting of two parts. The zero-inflation part predicts 0 count values (rather than a non-zero count value, as in the usual case) with a logistic regression; the negative binomial part models the non-zero counts with a negative binomial regression. Although covariates included in the two parts of the model do not need to be the same, we specified the same covariates for the two parts of the model (except for a small difference in the specification of the offset, as discussed below) because our interest in the SNAP frame's usefulness applies to both parts of the model. For a short description of zero-inflated negative binomial regression models, please refer to Appendix B.

As stated above when we introduced the two outcome variables, for the negative binomial part of each model, we included *the total number of households screened* in each SSU-month unit as an offset for the outcome variable *number of SNAP households*, and *the total number of contact attempts for all households* as an offset for the outcome variable *number of SNAP contact attempts*. With the offset term in our model, we were essentially parameterizing the percentages (i.e., $\frac{\text{number of SNAP households}}{\text{total number of households}}$ and $\frac{\text{number of SNAP attempts}}{\text{total number of attempts}}$) (see Appendix B). The consideration that a larger exposure results in more event occurrence also applies to the zero-inflation part of the model. The larger the exposure, the less likely we are to observe a zero. Because it is not sensible to statistically specify an offset term for a logistic model, we included a log-transformed version of the *total number of households screened* as a covariate in the zero-inflation part of the model for modeling the *number of SNAP households*. Similarly, we included a log-transformed version of the *total number of contact attempts* as a covariate in the zero-inflation part of the model for modeling the *number of SNAP contact attempts*.

Results

Descriptive results

The analytical dataset of the current study is at the SSU-month level. The two outcome variables (*number of SNAP households* and *number of SNAP contact attempts*) as well as our main covariate (*SNAP frame usage*) are defined at this SSU-month level and capture the sampling outcomes and activities within each SSU-month. The series of covariates that we constructed from the CPD to characterize the SSUs are at the SSU level. A given SSU was therefore assigned only one set of values for these characteristics, and these values were repeatedly assigned to the SSU's corresponding SSU-month units in our dataset. We bring this distinction between the SSU level and the SSU-month level to the reader's attention because, in this descriptive section, we report the descriptive statistics of the variables at the level where they are defined. The means, ranges, and standard deviations of the variables are presented in Table 2. We also compared the mean estimates based on the units with and without the SNAP frame.

Table 2. Descriptive statistics for the study variables based on the full sample, the units with a SNAP frame and the units without a SNAP frame.

	Overall				With SNAP frame	Without SNAP frame	
Variables defined at the SSU-month level (n= 2319 SSU-month units)	Mean	SD	Median	Range	Mean	Mean	T-test for difference in means
Number of SNAP households (outcome 1)	1.02	1.55	1	0 – 21	1.07	0.75	t=4.99, p<0.001
Number of total households (offset 1)	5.08	5.14	4	1 – 93	5.26	4.17	t=4.57, p<0.001
Number of SNAP contact attempts (outcome 2)	4.55	7.04	2	0 – 60	4.84	3.08	t=5.51, p<0.001
Number of total attempts (offset 2)	23.59	23.66	17	2 – 289	24.67	18.10	t=6.97, p<0.001
% of households coming from a SNAP frame (i.e., SNAP frame usage) (covariate)	23.36%	31.84%	0%	0% – 100%			
Variables defined at the SSU level (n= 394 SSUs)	Mean	SD	Median	Range	Mean	Mean	T-test for difference in means
% of people who are college and above educated (covariate)	25.44%	19.30%	19.06%	0% – 94.89%	25.99%	22.92%	t=1.15, p=0.25
% of people who are below poverty level (covariate)	17.14%	13.73%	13.90%	0% – 85.2%	16.50%	20.04%	t=-1.86, p=0.07
% of people who moved in (covariate)	17.01%	11.18%	13.99%	1.20% – 74.82%	16.81%	17.91%	t=-0.72, p=0.47
% of non-English speaking households (covariate)	0.14%	0.93%	0%	0% – 12.71%	0.17%	0.00%	t=3.07, p=0.002
% of households that receive public assistance (covariate)	2.64%	3.44%	1.58%	0% – 22.54%	2.48%	3.38%	t=-1.75, p=0.08
% of renter households (covariate)	36.44%	24.51%	29.85%	0% – 100%	36.26%	37.23%	t=-0.30, p=0.76
Median annual income (thousands) (covariate)	51.53	28.74	44.61	2.50 – 250.00	52.82	45.67	t=2.22, p=0.03
Mail return rate (covariate)	78.62%	8.07%	80.2%	45.75% – 93.80%	78.63%	78.54%	t=0.08, p=0.94

The correlations between the variables are presented in Table 3. The correlations between SSU characteristics were calculated at the SSU level (marked gray in Table 3). The remaining correlations were calculated at the SSU-month level. Based on the bivariate relationships, there was indeed a positive association between having and using SNAP frames and identifying SNAP households, as suggested by the positive correlation between the number of SNAP households and the SNAP frame usage (r= 0.33). Similarly, having and using SNAP

frames was positively associated with a larger amount of contact effort being directed to the SNAP households (r=0.31). All SSU characteristic covariates, except for percent of non-English households, were significantly associated with the two outcome variables, suggesting that it was easier to identify SNAP households in some SSUs than in others.

Table 3. Pairwise Correlations of the study variables.

	1	2	3	4	5	6	7	8	9	10	11
1 Number of SNAP households	1.00										
2 Number of SNAP contact attempts	0.88***	1.00									
3 SNAP frame usage	0.33***	0.31***	1.00								
4 College	-0.26***	-0.24***	-0.26***	1.00							
5 Poverty	0.21***	0.23***	0.22***	-0.35***	1.00						
6 Move in	0.09***	0.09***	0.04*	0.06	0.51***	1.00					
7 non-English	-0.02	0.01	0.03	0.06	-0.02	-0.03	1.00				
8 Public assistance	0.15***	0.15***	0.10***	-0.30***	0.28***	0.04	0.11*	1.00			
9 Renter	0.15***	0.18***	0.13***	-0.10*	0.60***	0.64***	0.16**	0.26***	1.00		
10 Income	-0.25***	-0.24***	-0.24***	0.68***	-0.62***	-0.33***	0.01	-0.27***	-0.48***	1.00	
11 Mail return	-0.16***	-0.19***	-0.16***	0.22***	-0.55***	-0.45***	-0.15**	-0.24***	-0.69***	0.37***	1.00

Modeling results

Predicting the number of SNAP households

The estimated parameters in the zero-inflated negative binomial model for the outcome variable *number of SNAP households* are presented in Table 4. As all interaction terms were not significant (Model 2), we interpret Model 1 as our final model for this dependent variable.

After adjusting for the various SSU characteristics, the coefficient for having and using SNAP frames was significant, for both the zero-inflation part and the negative binomial part of the model. In terms of the negative binomial part, the higher the percentage of households coming from a SNAP frame in each SSU-month, the more SNAP households were screened (Estimate= 0.144, 95% CI= [0.120, 0.159]). Mirroring this association, the higher percentage of households coming from a SNAP frame in each SSU-month unit, the less likely it was that no SNAP households were screened (Estimate= -2.468, 95% CI= [-4.207, -0.729]).   Time also had a negative effect on the identification of SNAP households. In the negative binomial part of the model, the last period from October to January screened significantly fewer SNAP households compared to the period from April to June (Estimate= -0.195, 95% CI= [-0.304, -0.087]).

Table 4. Estimated parameters in the zero-inflated negative binomial models.

	Number of SNAP households		Number of SNAP contact attempt		
	Model 1	Model 2	Model 1	Model 2	**Model 2'**
	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
Negative Binomial					
Intercept	-1.899(0.058)***	-1.823(0.081)***	-1.285(0.051)***	-1.246(0.066)***	-1.287(0.051)***
SNAP frame usage	0.144(0.008)***	0.130(0.013)***	0.078(0.007)***	0.066(0.011)***	0.075(0.007)***
July-Sep	-0.062(0.056)	-0.140(0.096)	-0.161(0.053)**	-0.203(0.080)*	-0.165(0.053)**
Oct-Jan	-0.195(0.055)***	-0.319(0.108)**	-0.111(0.054)*	-0.176(0.094)	-0.085(0.054)
July-Sep * SNAP frame usage		0.015(0.018)		0.007(0.017)	
Oct-Jan * SNAP frame usage		0.024(0.018)		0.020(0.017)	
College	-0.057(0.021)**	-0.054(0.022)*	-0.036(0.020)	-0.036(0.020)	-0.036(0.020)
Poverty	0.009(0.0255)	0.006(0.026)	0.071(0.025)**	0.071(0.025)**	0.071(0.025)**
Move in	-0.033(0.026)	-0.034(0.026)	-0.029(0.025)	-0.029(0.026)	-0.029(0.026)
Non-English	-0.531(0.248)	-0.518(0.349)	0.093(0.245)	0.083(0.246)	0.075(0.246)
Public assistance	0.023(0.062)	0.028(0.062)	0.015(0.061)	0.009(0.061)	0.007(0.0616)
Renter	0.019(0.016)	0.019(0.016)	0.005(0.015)	0.007(0.015)	0.007(0.015)
Annual income	-0.011(0.002)***	-0.011(0.002)***	-0.005(0.002)*	-0.005(0.002)*	-0.005(0.002)*
Mail return	-0.040(0.039)	-0.038(0.039)	-0.055(0.038)	-0.050(0.038)	-0.051(0.038)
Zero-inflation					
Intercept	-1.383(0.419)**	-1.202(0.418)**	0.095(0.127)	0.398(0.153)**	0.382(0.153)**
SNAP frame usage	-2.468(0.888)**	-3.722(3.454)	-0.385(0.035)***	-0.638(0.099)***	-0.633(0.101)***
July-Sep	0.052(0.424)	-0.122(0.439)	0.377(0.160)*	0.194(0.187)	0.208(0.186)
Oct-Jan	-0.173(0.581)	-0.5135(0.623)	1.006(0.169)***	0.378(0.218)	0.408(0.218)
July-Sep * SNAP frame usage		1.404(3.608)		-0.067(0.160)	-0.074(0.159)
Oct-Jan * SNAP frame usage		2.067(3.600)		0.368(0.106)***	0.356(0.107)***
College	0.354(0.159)*	0.366(0.160)*	0.167(0.050)***	0.175(0.051)***	0.175(0.051)***
Poverty	-0.484(0.349)	-0.489(0.354)	-0.141(0.077)	-0.150(0.078)	-0.151(0.078)
Move in	-0.095(0.271)	-0.075(0.266)	0.110(0.079)	0.101(0.081)	0.102(0.0817)
Non-English	0.183(2.291)	0.474(2.543)	1.294(0.616)*	1.259(0.630)*	1.266(0.631)*
Public assistance	-1.977(1.269)	-1.883(1.227)	-0.133(0.194)	-0.174(0.201)	-0.173(0.202)
Renter	0.341(0.184)	0.341(0.185)	0.066(0.045)	0.076(0.046)	0.076(0.046)
Income	0.013(0.016)	0.011(0.015)	0.018(0.005)***	0.017(0.005)***	0.017(0.005)***
Mail return	1.478(0.451)**	1.487(0.441)***	0.378(0.116)**	0.385(0.119)**	0.386(0.119)**
Log(total)	-0.311(0.219)	-0.282(0.219)	-0.844(0.077)***	-0.734(0.081)***	-0.734(0.082)***

Note. For the outcome variable number of SNAP households, log(total) is the log-transformed number of total households screened. For the outcome variable number of SNAP contact attempts, log(total) is the log-transformed number of total contact attempts. * $p<0.05$, ** $p< 0.01$, *** $p< 0.001$.

Predicting the number of SNAP contact attempts

The estimated parameters in the zero-inflated negative binomial model for the outcome variable *number of SNAP contact attempts* are also presented in Table 4. The interaction between time and SNAP frame usage was not significant for the negative binomial part, but was significant for the zero-inflation part of the model. We fitted a Model 2′ (2 Prime) that retains the interactions only for the zero-inflation part of the model as the final model for this outcome variable.

The pattern regarding our first research question about the effect of having and using SNAP frames stayed the same. As shown in the negative binomial part of the model, having a larger percentage of households coming from a SNAP frame was associated with a significantly larger portion of contact attempts directed toward SNAP households (Estimate= 0.075, 95% CI= [0.061, 0.089]). Time had a significant negative effect. Compared to the period from April to June, fewer contact attempts were directed toward SNAP households in the period from July to September (Estimate= -0.165, 95% CI= [-0.270, -0.060]), though the difference was not significant for the period from October to January.

Considering the zero-inflation part of the model, the interaction between time and percentage of households coming from a SNAP frame was significant. For the reference period from April to June, having and using SNAP frames significantly decreased the likelihood of having no contact attempts directed toward SNAP households (Estimate= -0.633, 95% CI= [-0.830, -0.435]). This effect of the SNAP frame usage was significantly weakened for the period from October to January (the interaction estimate= 0.356, 95% CI= [0.147, 0.565]), although the overall effect of the SNAP frame usage remained significant for the period from October to January (Estimate= -0.276, 95% CI= [-0.346, -0.206]; values not included in the table).

These findings consistently suggest that having and using the SNAP frames resulted in significant benefits in terms of increasing the efficiency of SNAP household identification. To visualize the results, we plot the percentage of SNAP households screened over all households screened in each month for the three scenarios summarized in the *SNAP availability/usage indicator* (Figure 2, left panel). The red dashed line represents activities in the SSU-months corresponding to the five states that did not have a SNAP frame (note that survey activities ended in the five states without a SNAP frame in December), the blue dotted line represents the activities in the SSU-months that did have the SNAP frame but did not seek to recruit SNAP households (*SNAP frame usage* = 0), and the green solid line represents the

activities in the SSU-months that used the SNAP frame (the *SNAP frame usage* > 0). The green solid line generally hovered above the red dashed and blue dotted lines, indicating the benefit of using the SNAP frame in terms of increasing the efficiency of SNAP household identification. In parallel, we plot the percentage of SNAP contact attempts over all contact attempts in each month for the three scenarios summarized in the *SNAP availability/usage indicator* (Figure 2, right panel). We see a very similar pattern in this figure.
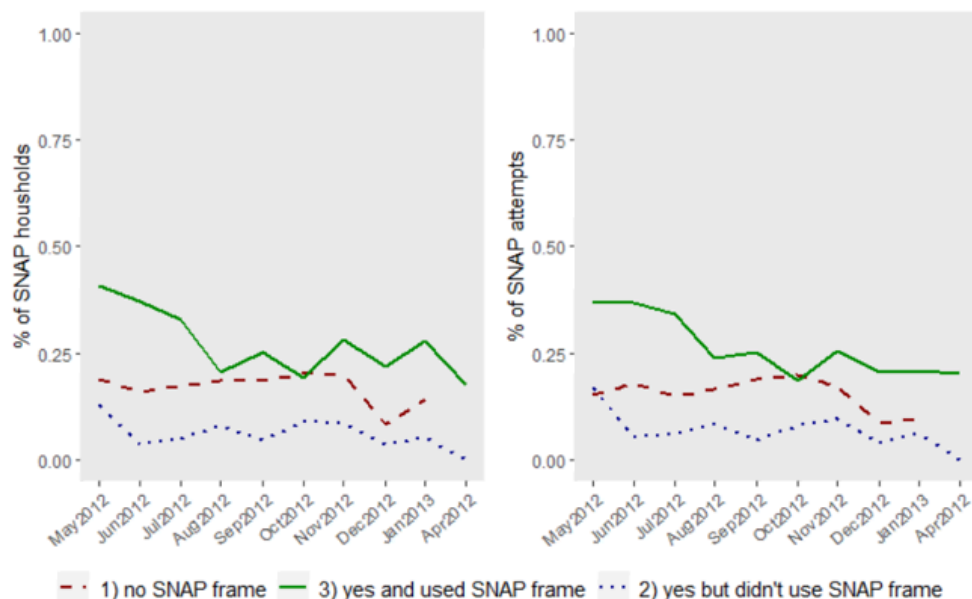

Figure 2. Left panel: percentage of SNAP households as a function of the *SNAP availability/usage indicator*, with "% of SNAP households" defined as $\frac{number\ of\ SNAP\ households}{numer\ of\ all\ households}$.
Right panel: percentage of SNAP contact attempts as a function of the *SNAP availability/usage indicator*, with "% of SNAP contact attempts" defined as $\frac{number\ of\ SNAP\ attempts}{numer\ of\ all\ attempts}$.

Discussion

Our first research question concerned whether the use of additional SNAP list frames improved the efficiency of identifying SNAP households as opposed to solely relying on general screening. We have shown that using the SNAP frames was associated with screening more SNAP households and a lower likelihood of screening zero SNAP households. Specifically, if we fix the values of all other covariates to zero (recall that we centered the covariates around their means), then a 10 percentage-point increase in SNAP frame usage was related to a 1.41 times increase (i.e., a 41% increase) in the expected number of SNAP households screened. These findings suggest that the additional list frames aided in the identification of SNAP households from the general population.

Considering the number of contact attempts, using the SNAP frames was associated with more contact attempts directed toward SNAP households and a lower likelihood of zero contact attempts directed toward SNAP households. Specifically, if we fix the value of all other covariates to zero, then a 10 percentage-point increase in SNAP frame usage was related to a 1.49 times increase (i.e., a 49% increase) in the expected number of contact attempts directed toward SNAP households. These findings suggest that when the additional list frames were used, a significantly larger amount of contact attempts were directed to the subpopulation of interest, which indicates more efficient effort directed at recruiting SNAP households.

Our second research question concerned whether the usefulness of the SNAP frames decreased over time because the list frames aged. Based on the interaction terms between the SNAP frame usage and the time variables, we found some evidence supporting this possibility, but the effect was not strong. Within FoodAPS's 10-month fielding period, the utility of the supplementary list frames persisted.

This study is one of the few that have empirically investigated the efficiency of multiple-frame surveys. Our findings confirm that general screening activities can identify SNAP households, as shown by the SNAP households screened in the five states whose state SNAP agency did not provide the supplemental SNAP frames. Nonetheless, by making comparisons between operations that worked with and without the additional list frames, our findings demonstrate that using additional list frames could improve screening efficiency and reduce survey screening costs in comparison to the general screening.

One caveat is that our results represent a case study. Our findings may not generalize if other studies focus on subpopulations of different sizes or if the list frames have different error structures. However, our results have shown the benefit of using additional frames even though SNAP households were not rare in the population (19%). Others who recruit rarer subpopulations may expect larger gains with additional list frames (Srinath et al., 2004).

Additionally, our results are based on observational data. The availability of SNAP frames was not part of a randomized experimental design. Even though our analyses controlled for a series of SSU characteristics that were likely to be related to SNAP household concentrations, we might not have accounted for all factors that differed between SSUs from states that did and did not provide the SNAP frame (and thus could have confounded the relationships of the presence of a SNAP frame with the recruitment outcomes).

Another limitation of the current analyses is that our analytical dataset was based on the households that were screened and reported their SNAP status in the screening interviews. This is not ideal because only 66% of the contact attempts were associated with the screened households. The remaining futile attempts could not be analytically included in the construction of one of the outcome variables (the number of SNAP contact attempts) because we have no information on the cases that they targeted. In addition, we rely on the screened households to construct our major covariate, the *percentage of screened households coming from a SNAP frame*, because we do not know in which SSUs those non-screened units resided. Our findings could be affected if the households that did not respond were systematically different from the households that did respond. We unfortunately cannot avoid or evaluate this problem because an address-based sampling frame was used in the FoodAPS and no information is available on the non-contacted cases. Future studies may address this problem if they draw samples from a frame that contains information on all of the frame units.

Appendix A.
Construction of SSU characteristics as covariates

Table A. How we constructed the SSU-level characteristics using the block group level information provided by 2012 Census Planning Database.

Covariate	Construction
Percentage of people with a college degree or higher	Within each SSU, we took the sum across census block groups to obtain the number of college-educated people within each SSU; we then divided these college-educated SSU sums by the total number of people who were 25-years-and-above within the SSUs, which resulted in the percentages of people with a college degree or higher in each SSU.
Percentage of people who are below poverty level	We summed the number of people living below the federal poverty level across census block groups within each SSU, and then divided these values by the total number of people whose poverty status was determined in each SSU.
Percentage of people who moved	We summed the number of movers across census block groups within each SSU, and then divided these values by the total number of people who were older than one year of age in each SSU.
Percentage of non-English speaking households	We summed the number of non-English speaking households across census block groups within each SSU, and then divided these values by the total number of occupied households in each SSU.
Percentage of households that receive public assistance	We summed the households that received public assistance across census block groups within each SSU, and then divided these values by the total number of occupied households in each SSU.
Percentages of renter households	We summed the number of non-owner-occupied households across census block groups in each SSU, and then divided these values by the total number of occupied households in each SSU.
	We took a weighted average across the median incomes of census block groups within each SSU. The weights were

Median income	calculated as , with census block group size equaling the number of occupied households in each census block group, and SSU size equaling the total number of occupied households in the SSU that the census block group nests in.
Mail return rate	The CPD reports a mail return rate for each census block group. We took a weighted average of the rates across census block groups within each SSU, with weights constructed in the same way as for median income.

Appendix B.
A short description of zero-inflated negative binomial models

If we conceptualize the observed count variable y_i as a product of the two latent variables z_i and y_i^*

$$y_i = z_i y_i^*$$

where z_i is a binary variable with values 0 for observations who have consistent zero count and 1 for all other observations and where y_i^* has a negative binomial distribution, then the probability of each observed count can be modeled as

$$P(y_i = k) = \begin{cases} P(z_i = 0) + P(z_i = 1, y_i^* = 0), & \text{if k} = 0 \\ P(z_i = 1)f(k), & \text{if k} = 1, 2, ... \end{cases}$$

where $P(z_i = 0)$ is modeled with a logit link function including covariates, and $f(k)$ is the negative binomial distribution with its mean being a linear parametric function of some (other) covariates.

For the negative binomial part of the model, we included an offset term to account for the fact that larger exposure (e.g., more households screened) results in more event occurrence (e.g., more SNAP households screened). With this offset term, our model's parameterization

$$log E(y_i^*) = \beta \boldsymbol{X_i} + log(\text{Offset}_i)$$

can be rewritten as

$$log E(y_i^*) - log(\text{Offset}_i) = \beta \boldsymbol{X_i}$$

$$log E(\frac{y_i^*}{\text{Offset}_i}) = \beta \boldsymbol{X_i}$$

where $\boldsymbol{X}$ refers to model predictors. Thus, we are essentially parameterizing percentages (e.g., $\frac{\text{number of SNAP households}}{\text{total number of households}}$.)

References

1. Bricker, J., Dettling, L. J., Henriques, A., Hsu, J. W., Jacobs, L., Moore, K. B., … & Windle, R. A. (2017). Changes in US family finances from 2013 to 2016: Evidence from the Survey of Consumer Finances. *Federal Reserve Bulletin*, *103*(3), 1-42.
2. Chen, S., Stubblefield, A., & Stoner, J. A. (2020). Oversampling of Minority Populations Through Dual-Frame Surveys. *Journal of Survey Statistics and Methodology*. Advance online publication. https://doi.org/10.1093/jssam/smz054
3. Cowan, C. D. (1991). Using Multiple Sample Frames to Improve Survey Coverage, Quality, and Costs. *Marketing Research*, *3*(4), 66-69.
4. Gray, K. F., & Eslami, E. (2014). Characteristics of Supplemental Nutrition Assistance Program Households: Fiscal Year 2012. *US Department of Agriculture*. Retrieved from: https://files.eric.ed.gov/fulltext/ED582002.pdf
5. Hartley, H. O. (1962). Multiple frame surveys. In *Proceedings of the social statistics section, American Statistical Association* (Vol. 19, No. 6, pp. 203-206).
6. Hartley, H. O. (1974). Multiple frame methodology and selected applications. *Sankhya*, *36*(997), 118.
7. Irving, S. K., & Loveless, T. A. (2015). Dynamics of economic well being: Participation in government programs, 2009-2012: Who gets assistance. *US Department of Commerce, Economics, and Statistics Administration, US Census Bureau*. Retrieved from: https://www.census.gov/content/dam/Census/library/publications/2015/demo/p70-141.pdf
8. Kalton, G. (2003). Practical methods for sampling rare and mobile populations. *Statistics in Transition*, 6, 491-501.
9. Kalton, G. (2009). Methods for oversampling rare subpopulations in social surveys. *Survey Methodology, 35*(2), 125-141.
10. Kalton, G., & Anderson, D. W. (1986). Sampling rare populations. *Journal of the Royal Statistical Society: Series A (General)*, *149*(1), 65-82.
11. Lofquist, D. (2012). Households and families: 2010. *US Department of Commerce, Economics and Statistics Administration, US Census Bureau*. Retrieved from: https://www.census.gov/prod/cen2010/briefs/c2010br-14.pdf
12. Lohr, S. L., & Rao, J. N. K. (2000). Inference from dual frame surveys. *Journal of the American Statistical Association, 95*(449), 271-

280.

13. Lohr, S., & Rao, J. K. (2006). Estimation in multiple-frame surveys. *Journal of the American Statistical Association*, *101*(475), 1019-1030.

14. Lohr, Sharon. 2009. Multiple Frame Surveys. In D. Pfefferman and C.R. Rao (Ed.), *Handbook of Statistics, Vol. 29A, Sample Surveys: Design, Methods, and Applications* (pp. 3–8). Amsterdam: Elsevier/North-Holland.

15. Mecatti, F. (2007). A single frame multiplicity estimator for multiple frame surveys. *Survey Methodology*, *33*(2), 151-157.

16. Mecatti, F., & Singh, A. C. (2014). Estimation in multiple frame surveys: A simplified and unified review using the multiplicity approach. *Journal de la Société Française de Statistique*, *155*(4), 51-69.

17. Metcalf, P., & Scott, A. (2009). Using multiple frames in health surveys. *Statistics in Medicine*, *28*(10), 1512-1523.

18. National Household Food Acquisition and Purchase Survey. (2016). National Household Food Acquisition and Purchase Survey (FoodAPS): User's Guide to Survey Design, Data Collection, and Overview of Datasets. *U.S. Department of Agriculture*. Retrieved from: https://www.ers.usda.gov/media/8804/0_foodaps-user-guide-puf.pdf

19. Ranalli, M. G., Arcos, A., del Mar Rueda, M., & Teodoro, A. (2016). Calibration estimation in dual-frame surveys. *Statistical Methods & Applications*, *25*(3), 321-349.

20. Rao, J. N. K., & Wu, C. (2010). Pseudo–empirical likelihood inference for multiple frame surveys. *Journal of the American Statistical Association*, *105*(492), 1494-1503.

21. Sánchez-Borrego, I., Arcos, A., & Rueda, M. (2019). Kernel-based methods for combining information of several frame surveys. *Metrika*, *82*(1), 71-86.

22. Srinath, K. P., Battaglia, M. P., & Khare, M. (2004, August). A dual frame sampling design for an RDD survey that screens for a rare population. In *Proceedings of the Survey Research Methods Section of the American Statistical Association* (pp. 4424-29).

23. Wagner, J., & Lee, S. (2015). Sampling rare populations. In Timothy P. Johnson (Ed.), *Handbook of health survey methods* (pp.77-106). John Wiley.

24. Weerasekera, D., Lynch, C., Clark, R., Gribben, B., Boustead, C., & Tee, N. (2019). Methodology Report 2017/2018: New Zealand Health Survey. *Ministry of Health, New Zealand Government*. Retrieved from: https://www.health.govt.nz/system/files/documents/publications/nzhs-methodology-report-2017-18-20190319-update.pdf

25. West, B. T., Wagner, J., Hubbard, F., & Gu, H. (2015). The utility of alternative commercial data sources for survey operations and estimation: evidence from the National Survey of Family Growth. *Journal of Survey Statistics and Methodology*, *3*(2), 240-264.