Combatting Attrition Bias using Case Prioritization in the Survey of Income and Program Participation

Kevin Tolliver[1], Jason Fields[1], Stephanie Coffey[1], Amanda Nagle[1]

[1]U.S. Census Bureau, 4600 Silverhill Rd., Washington, DC, 20233

Abstract
This paper details an experiment testing an adaptive survey design approach to improving sample representativeness in a national longitudinal face-to-face survey. The adaptive strategy, case prioritization, was employed to focus data collection resources on cases that we feel may have the highest impact on attrition bias in the Survey of Income and Program Participation. Achieving interviews with these cases can have an impact on estimates of program participation, the essential aim of the survey. The findings of the experiment suggest that prioritizing cases can help the survey retain these targeted cases.

Key Words: Adaptive Survey Design, Case Prioritization, Attrition, Field Management, R-indicators, Paradata

1. Introduction

Most longitudinal studies face challenges maintaining representativeness in the presence of sample attrition. Fumagalli, L., Laurie, H., & Lynn, P. (2013) discuss how smaller achieved sample sizes due to attrition can imperil post-data collection analyses, and, where attrition is tied to outcomes of interest, how attrition can be a source of nonresponse bias. The authors also stress the importance of evaluating different attrition reduction strategies as those strategies have different cost and effectiveness properties. Although historically, a popular strategy used to combat attrition is the use of monetary incentives (Church, 1993; Singer, E., Gebler, N., Raghunathan, T., van Hoewyk, J. & McGonagle, K., 1999; James, 1997; Laurie, H., & Lynn, P., 2008; Rodgers, 2002; Fumagalli, L. et. al. 2013), adaptive survey designs (ASDs) have become increasingly popular (Tourangeau, R., Brick, J., Lohr, S., & Li, J., 2017) and has the potential to combat attrition in a longitudinal study (Lynn, 2017).

An adaptive survey design (ASD), has the potential to combat attrition because it can use information acquired between interviews and tailor protocol to improve survey outcomes. ASDs that tailor methods to individuals based on interim outcomes are "dynamic adaptive designs" (Schouten, Peytchev, and Wagner 2017). There are no discrete phases in data collection and the survey can elect to make course corrections continuously throughout data collection (Tourangeau et al., 2017). ASDs that tailor methods to individuals based solely on information prior to the start of data collection are called "static adaptive designs" (Schouten, Peytchev, and Wagner 2017).

One ASD course correction, case prioritization, may reduce the chance of nonresponse bias (Peytchev et al., 2010; Wagner, 2013; Lynn 2017). Case prioritization targets a subset of cases with pre-identified data collection features that are different from the typical features applied to the non-targeted population. For example, targeted cases could receive incentives, longer data collection periods, different modes of data collection, or even interviewers that are more experienced. Lynn (2017) refers to this as a "targeted design." Tourangeau et. al. (2017) and Lynn (2017) summarize some of the most recent developments in adaptive survey design. Tourangeau's 2017 paper, which references more than 30 papers on the topic since 2010, states that major changes to survey protocol are likely to reduce nonresponse bias, but the gains in response or reductions in the variation in response propensities have typically been "modest." Lynn (2017), who references more than 20 "targeted designs," reports that there have been experimental studies that have achieved desired outcomes, but few of the successful design features were means of tackling nonresponse bias. The evidence of adaptive survey design's impact in longitudinal studies to tackle nonresponse bias and attrition remains limited.

Coffey, S., Reist, B., & Miller, P. (upcoming) found that a dynamic adaptive design produced equal or better sample representation with lower cost in the National Survey of College Graduates (NSCG), a large longitudinal government survey. Studies have shown that response rates alone do not provide evidence of data quality, in particular with respect to nonresponse bias (Groves & Heeringa, 2006; Groves & Peytcheva, 2008). Peytchev, A., Riley, S., Rosen, J., Murphy, J., & Lindblad, M. (2010) prioritize low response propensity cases using incentives and Coffey et. al. (upcoming) tailor modes of collection to prioritize under-represented cases using R-indicators with incentives to attempt to reduce nonresponse bias.

In this paper, we discuss the adaptive survey design in use and being refined for the Survey of Income and Program Participation (SIPP), a national household longitudinal study conducted annually by the US Census Bureau (U.S. Census, 2019). The SIPP samples households, but then becomes person-based as it follows interviewed original sample members over the length of the panel (*U.S. Census Bureau*, 2016). We use a similar case prioritization strategy to the approach Peytchev et. al. (2010) uses for the Community Advantage Panel Survey with the similar goal of reducing nonresponse bias. In their research, individuals of this face-to-face survey are assigned a priority score based on prior year information. Although Peytchev has null findings, he suggests that with response rates near 92 percent, there may be little room for improvement in nonresponse bias. The SIPP does not have response rates that high, therefore testing a similar strategy to this survey could lead to positive results. Furthermore, we modify this approach by using two response models; prioritizing cases that have low response propensity based on prior wave demographic information but high response propensity based on the current wave paradata. We employ R-indicators, contact history information, administrative records in order to prioritize cases to enhance locating, combat attrition of select cases, and reduce overall attrition bias.

The remainder of this manuscript is as follows: Section 2 briefly introduces SIPP and discusses the motivation, the intent, and the methods behind the case prioritization. The details of the models and instructions used are given in the Appendices. Section 3 describes how the experiments were conducted and gives results. This includes measures of attrition of target households, attrition bias, and overall attrition. Section 4 discusses our conclusions, how our conclusions relate to prior research, and research going forward.

2. Case Prioritization in the Survey of Income and Program Participation (SIPP)

2.1 SIPP Background and Overview

The SIPP is a national household level longitudinal study conducted annually by the US Census Bureau (U.S. Census, 2019). The main objective is to provide accurate and comprehensive information about income and program participation dynamics for individuals and households in the United States.

Respondents were interviewed once each year for four years during the course of the 2014 panel. Each interview period is called a "wave." During the first wave of the survey, interviewers attempt to complete the survey in person with all sampled households, although a mix of face-to-face and telephone modes may be used at the respondent's request. During SIPP interviews, data are collected about all household members. Self-response is preferred for household members 15-years old and over however, proxy reports are accepted. During subsequent waves, the SIPP becomes person-based, following wave 1 respondents to new addresses and interviewing all of the individuals at those addresses. Non-respondents from wave 1 are not retained in subsequent waves. SIPP, like most longitudinal data collections, emphasizes sample retention and representativeness as key programmatic objectives to support high data quality. We focus our adaptive design interventions to help direct data collection effort to cases which maybe more important to ongoing sample composition, representativeness, and improving data quality.

Because the SIPP follows wave 1 respondents to new addresses and interviews all of the individuals at those addresses, the SIPP operation is tasked with identifying that someone has moved, locating that individual, and obtaining their response from the new household in addition to all the individuals who remained at the original household during the next wave. These "mover" cases are often difficult to locate, which leads to higher attrition of these cases. Additionally, these cases are often not missing at random. Movers are 57 percent more likely to come from a household who received program participation than a household who has not.

During wave 2 of the 2014 panel (conducted in the 2015 calendar year), survey leadership observed that effort was applied to cases inconsistently, and that some cases had not received a single recorded contact attempt, even as the data collection window was closing. The following year (2016), in the third wave of the 2014 panel, a pilot adaptive survey design study was conducted to test the feasibility of case prioritization in SIPP, as a way to manage the effort applied to individual cases. In the fourth wave of the 2014 panel (2017 calendar year), a more formal experiment was conducted with the intention of improving data quality and combatting attrition bias. This manuscript gives high-level summaries of the pilot study in wave 3, but focuses on the wave 4 case prioritization experiment.

2.2 Case Prioritization Methodology

The primary goal of testing case prioritization in the SIPP was to reduce bias due to attrition in the key survey outcomes: the percent of individuals who receives benefits from the Temporary Assistance for Needy Families (TANF), Supplemental Security Income (SSI), Women Infants and Children (WIC), Supplemental Nutritional Assistance (SNAP), or general assistance programs, (U.S. Census, 2019). Participation in one of these government programs is the central interest of the SIPP, and it is fundamental to other estimates that come from this survey. Achieving this goal would mean a more balanced sample in future waves of SIPP, resulting in more consistent coverage of subdomains of interest, since non-respondent cases are less likely to respond in future waves of a longitudinal survey.

Our strategy for improving data quality and reducing attrition bias is to prioritize movers, households who are unable to be linked to administrative data, and under-represented demographic subpopulations. Each of these household types are essential to accurately capturing the estimates on income and program participation. We simultaneously deprioritize any household that belongs to an over-represented demographic subpopulation that shows high reluctance to an interview. This should help the interviewer reallocate their effort to households that may have more value to the final data product or more receptive of an interview, with the intent of helping the interviewer obtain their higher priority cases. In doing this, we try to obtain equivalent response rates across the different sub-populations in order to achieve a "representative" sample with minimal bias to attrition.

As noted before, households or individuals that move between survey waves are more at risk to attrit for the simple reason that they may be more difficult to locate in the next wave. A lower response rate among movers could result in increased bias due to attrition. Prioritizing any household who has a recent register in the National Change of Address Registry ("likely movers") help identify "mover" households earlier in data collection, which gives the interviewers more time to locate the movers and a better chance at completing the survey. As a result, one of our primary tools for combating attrition bias was to prioritize mover households.

While using administrative information is important for identifying likely movers, it is also used for other statistical purposes, most notably imputation. The SIPP has integrated administrative data into Sequential Regression Multiple Imputation models that impute for missing content areas in interviewed households (Benedetto, Motro, & Stinson, 2015; Giefer, Williams, Benedetto, & Motro, 2015). The U.S. Census Bureau assigns Protected Identification Keys (PIKs) to their survey participants, which are unique personal identifiers to link persons to administrative records for statistical purposes (Wagner & Layne, 2014). Approximately 13% of SIPP sample are unable to be linked and matched to administrative data for various reasons. These "non-PIK" households have increased importance in fieldwork because the administrative records cannot be used in imputation and are not available for other administrative statistical practices. Further, all information that represents these households in the final longitudinal data come from the survey respondents, increasing the information value for these households, and their priority. Any household that may contain an individual aged 16 years or older that could not be linked to administrative records continue to be important sources of information in the SIPP data.

Lastly, we use prior wave information related to program participation and the SIPP 2014 nonresponse bias report (Westra and Nwaoha-Brown, 2017), to prioritize and deprioritize certain households. Variables such as number of persons in household and urban area help identify different demographic subpopulations that may or may not require additional effort. The details of which variables are used in the model are in Appendix 1.

2.1 Accounting for Response

Beyond the primary goal, we had a secondary objective to maintain attrition rate as much as possible. We expect prioritizing under-represented households and mover households to have negative ramifications on overall attrition. The intended result is to have a more representative sample for future waves, rather than having a larger sample. However, our strategy for trying to preserve attrition was to use a secondary response model derived from contact history information. Whereas, the R-indicator uses prior wave 1 demographic information to estimate propensities among demographic subpopulations, this secondary model uses current wave contact history information in order to estimate propensities given the amount of effort already exerted each day. The contact history response was modeled with number of contact attempts; interviewee reluctance, interviewer strategy, and prior wave response (see Appendix 2). This model was run daily on all finalized households (no further work to be done), and used to predict responses of any non-finalized households.

During the final month of data collection cases with extremely low response propensity scores were stopped entirely in an attempt to get interviewers to put more effort on their other cases. The propensity to respond and the partial R-indicators generally act in countervailing directions. Coupled with a general preference to maintain as much of the longitudinal sample as possible, striking a balance between these criterions is one of the bigger operational challenges. When deciding how many and which cases to prioritize and deprioritize throughout data collection, the research team wanted to have one bivariate scatter plot that assesses the likelihood to respond given their contact history and their likelihood to be under-represented at the end of data collection.
The criteria for prioritized cases in our prioritization changed weekly, making the case prioritization dynamic. These thresholds were determined through conversation between the research team and survey leadership. The R-indicator values for each individual case

$$\widehat{R_\iota} = \sqrt{\frac{w_i}{W}}\left(\hat{\rho}_i - \hat{\bar{\rho}}\right)$$

given that the individual record has a balancing response propensity $\hat{\rho}_i$, a baseweight w_i, the sum of the baseweights W, the unconditional individual level R-indicator ($\widehat{R_\iota}$) measures how much the individual observation contributes to the full sample's imbalance.

This individual case-level approach effectively treats each individual household like its own category. Like the unconditional partial R-indicator for category (Schouten, B., Shlomo, N., & Skinner, C., 2011), negative values indicate under-represented households and positive values indicate over-represented households. Individual cases that belong to several under-represented groups should have a large negative value and any individual cases that belong to several over-represented groups should have a large positive value.

3. Experimental Analysis

The experiments tested the effects of the different prioritizations. Every interviewer received the same instructions, see Appendix 3. In wave 3, each interviewer was assigned to prioritization (T) or no prioritization (C) group. If the interviewer was in the prioritization group, they were eligible to see high, medium, and low priority status next to each case on their workload laptops. If the interviewer was in the no prioritization group, they only saw medium priority status next to each case on their laptop despite the "*true*" priority of the case. In this pilot study, the high priorities during the first three months of data collection corresponded to likely movers, movers identified, and households that could not be administratively linked. During the final month of data collection, the high priorities were no longer these cases. Instead, cases were prioritized by R-indicators and response propensities. Table 3.1 displays the prioritization schedule.

Table 3.1 Prioritization Schedule for Wave 3 Pilot Study

Month	Prioritization Group (T)	No Prioritization Group (C)
1	H – Likely Movers, Non-PIKs, Movers Identified M- Otherwise	M – Every case
2	H – Likely Movers, Non-PIKs, Movers Identified M- Otherwise	M – Every case
3	H – Likely Movers, Non-PIKs, Movers Identified M- Otherwise	M – Every case
4	H – Under-represented/Likely Respondent L – Over-represented/Unlikely Respondent M – Otherwise	M – Every case

In wave 4, each interviewer was randomly assigned to one of three experimental groups: static adaptive prioritization (T1), static and dynamic adaptive design prioritization (T2) which will be referred to as "dynamic adaptive" henceforth, or no prioritization control (CO) group. The interviewers were stratified on interviewer experience (new hire, returning interviewer), regional office (The Census Bureau has six regional offices), workload size (below median workload size, above median workload size). Any interviewer assigned to the static adaptive prioritization design received only high and medium priority case assignments. Throughout data collection, the high priorities corresponded to likely movers, identified movers, and households that could not be administratively linked (non-PIK). Any interviewer assigned to the dynamic adaptive prioritization received, high, medium, and low case assignments. At the start of data collection, the priorities only corresponded to likely movers, identified movers, and non-PIK households. Beginning the second month of data collection, model-based priorities were introduced and used concurrently with the other prioritizations, (model-based prioritization trumped the static adaptive prioritization for cases that belonged in multiple categories).

Table 3.2 Prioritization Schedule for Wave 4 Experiment

Month	Static Adaptive Prioritization Group (T1)	Dynamic Adaptive Prioritization Group (T2)	No Prioritization Group (CO)
1	H – Likely Movers, Non-PIKs, Movers Identified M- Otherwise	H – Likely Movers, Non-PIKs, Movers Identified M- Otherwise	M – Every case
2	H – Likely Movers, Non-PIKs, Movers Identified M- Otherwise	H – Likely Movers, Non-PIKs, Movers Identified, Under-represented/Likely Respondent	M – Every case

		L – Over-represented/Unlikely Respondent M – Otherwise	
3	H – Likely Movers, Non-PIKs, Movers Identified M- Otherwise	H – Likely Movers, Non-PIKs, Movers Identified, Under-represented/Likely Respondent L – Over-represented/Unlikely Respondent M – Otherwise	M – Every case
4	H – Likely Movers, Non-PIKs, Movers Identified M- Otherwise	H – Likely Movers, Non-PIKs, Movers Identified, Under-represented/Likely Respondent L – Over-represented/Unlikely Respondent M – Otherwise	M – Every case

On average, the interviewers had between 25-40 cases to finish in the four-month data collection period. The static adaptive group had approximately one-third of their cases that were high priority throughout data collection. The dynamic adaptive group however had approximately one-third of their cases that were high priority at the start of data collection, but throughout data collection, we progressively shifted the percent of high cases closer to 45 percent and the number of low cases to 20 percent.

3.1 Experimental Analysis Methodology

Although the case prioritization experiment was instituted at the interviewer-level, we evaluate the results at the survey-level, because our adaptive design goals involve survey-level estimates. One of the most challenging facets of the experiment was that, in the normal course of a SIPP field period, cases could be reassigned to other interviewers by regional and central field management for various reasons. This meant that cases could shift between treatment and control interviewer groups. Approximately 14 percent of cases had priority changes because of reassignments in wave 3 and 11 percent of cases had priority changes because of reassignments in wave 4, and thus experienced a treatment and control effect.

It was important to the experiment's success that the treatment interviewers avoid working cases autonomously and do as they were instructed. Although each of the interviewers were given the same set of instructions, there was no mandate to enforce the adaptive protocols Wagner (2013) provides some evidence that interviewers may ignore prioritization information. Defining compliance in this context is difficult because the instructions (Appendix 3) maybe left to interpretation. For this paper, we do not conclusively say if an interviewer followed protocols, but we make judgments of noncompliance if we can identify in the interviewer's contact history that they made no attempts on their high priority cases during the first month or no attempts on their high priority cases for more than half the weeks. This provides a conservative measure of noncompliance. There still needs to be more research conducted to determine an appropriate measure of compliance.

Because some households receive treatment and control effects, and some interviewers do not follow protocols are limitations to our analyses, we provide results in three different ways to offer sensitivity to our analyses:

- Omitting sample that may have treatment and control effects (approximately 11 percent of cases)
- Omitting interviewers who did not follow protocols (130 interviewers which totaled to 8 percent of cases)
- Including all cases, assuming the interviewer who finalized the case at the end of data collection (close out) is the interviewer who had the largest effect on the outcome of that case.

Many of our results displayed are in percent. The p-values reported correspond to equality of means test with Normal approximation of a Bernoulli distribution. A few of our results are not in percent. The p-values reported for these outcomes are computed using two-sample t-test with pooled variances.

3.2 Attrition Bias Analyses

The prioritization does not simply try to reduce attrition but rather tries to maintain the equivalent attrition rates across different subpopulations related to income and program participation. Our analyses examine our representative measure when monitoring the prioritization, attrition of mover households and non-PIK households, and overall attrition bias.

As a measure of variability in response propensities, the use of R-indicators (Schouten et. al., 2009) served as a tool for assessing the variability of attrition across subpopulations. If every subpopulation attrits at the same rate, the R-indicator theoretically should remain at one throughout data collection and the attrition bias would be zero.

Prioritizing using R-indicators led to a more representative respondent population, (R-indicator nearly 0.1 larger in wave 3 and wave 4). The chance of observing that difference by chance was less than 0.001 according to a non-parametric permutation test described by (Zieffler, Harring, & Long, 2011). In this test, under the null hypothesis there is no difference between the R-indicators produced by the treatment groups, the cases with priorities, and the R-indicator produced by the control, the non-prioritized group. By observing 10,000 permutations of how interviewers were assigned to the experimental groups and recording the number of permutations that had an R-indicator difference as high or more than what is observed, we are able to estimate the chance of seeing a difference that large due to chance. Figure 3.2.1 displays the differences in the three different ways described.

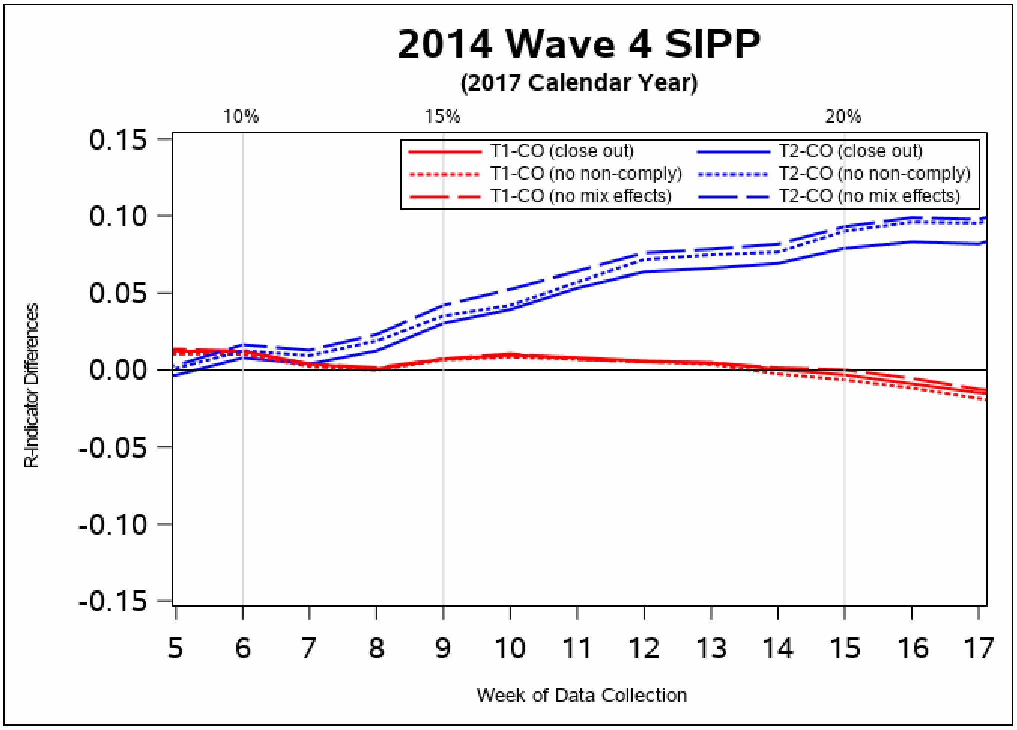

Figure 3.2.1: Full Sample R-indicator by Week of Data Collection

Targeting mover households and non-PIK households did not lead to a reduction of attrition rates among these households. The prioritization may have had small adverse effects on attrition rate.

Approximately 21 percent of household structures in SIPP sample have movers throughout the course of the panel, and movers attrit at a higher rate than their non-mover counterparts do. We define attritors by households who had completed or sufficient partial interviews in wave 1 and no interviews in wave 4. Successfully identifying and locating movers should significantly reduce attrition in a longitudinal survey, and may reduce attrition bias because movers are more likely to come from a household who have received government programs. Across the three differing analyses, targeting likely mover households identified approximately 0.65 more mover households per interviewer for the static adaptive treatment and 0.55 more mover households per interviewer for the dynamic adaptive treatment. See Table 3.2.1. Even though the attrition rate among these movers are slightly higher, this product results in approximately 10 percent more mover respondents. This was consistent with what we viewed in the 2016 pilot study as well. One possible explanation is that the effort interviewers put toward likely movers led to an increase of identified movers, but the additional heft of their workloads led to them not completing cases at the same rate.

Attriting household that have individuals that cannot be linked to administrative data have consequences on data quality. As noted before, SIPP uses administrative data for imputation. The non-mover households that are not linked to administrative data are our stationary targets. These make up approximately 13 percent of SIPP sample. In the wave 3 pilot study, the prioritization led to almost a two-percentage point increase in response rate among Non-PIK households. However, in the wave 4 experiment the prioritization led to an increase in attrition among these households. It is unclear why there was a positive effect in wave 3 and a negative effect in wave 4, but these results were largely insignificant both years. Table 3.2.1 summarizes the results for wave 4.

Table 3.2.1 Movers Identified per Interviewer, Unweighted Attrition Rate of Movers, and Unweighted Attrition Rate of Stationary High Priority Cases

Mthd	No Prioritization (CO)			Static Adaptive Prioritization (T1)			Dynamic Adaptive Prioritization (T2)		
	Movers Identified per Int.	Mover Attrit. Rate	Stationary Target Attrit. Rate	Movers Identified per Int.	Mover Attrit. Rate	Stationary Target Attrit. Rate	Movers Identified per Int.	Mover Attrit. Rate	Stationary Target Attrit. Rate
Close out	4.91	54.1 (N/A)	51.0 (N/A)	5.47 (0.108)	56.6 (0.068)	54.1 (0.026)	5.36 (0.189)	55.5 (0.229)	53.9 (0.033)
No Mixed Effect	4.44	47.5 (N/A)	50.1 (N/A)	5.23 (0.014)	51.7 (0.003)	52.2 (0.124)	5.11 (0.038)	49.9 (0.079)	51.5 (0.236)
No Non-Comp.	4.90	49.9 (N/A)	51.3 (N/A)	5.59 (0.006)	52.5 (0.006)	52.4 (0.276)	5.44 (0.134)	50.6 (0.348)	52.35 (0.295)

While the R-indicators assess representation among different demographic groups and attrition rates of the mover and stationary targets assess representation of these groups, ultimately our primary goal is to reduce the bias due to attrition. We define the Attrition Bias (AB) in this context as

$$AB(X) = \left(\frac{N_{Attr}}{N_{Wave1}}\right)\left(\hat{\theta}_{Wave4}(X) - \hat{\theta}_{Wave1}(X)\right)$$

$\hat{\theta}_{.}(X)$ refers to the percentage of households who received government assistance X in the offset$(\cdot)$. The offset *Attr* refers to households who had completed or sufficient partial interviews in wave 1 but not in wave 4, the offset *Wave4* refers to the estimate from all households who had completed or partial interviews in wave 4, and the offset *Wave1* refers to all households with completed or sufficient partial interviews in wave 1.

The dynamic adaptive prioritization resulted in a reduction in the magnitude of attrition bias in all of the five key outcomes, while the static adaptive prioritization resulted in a reduction of the magnitude of attrition bias in four of the five key outcomes. This pattern is consistent with the three differing analyses. These differences are significantly lower for SNAP, SSI, and WIC. The probability of all five outcomes having lower attrition bias by chance is 0.03125[1] and the probability of all four of the five variables having by chance is 0.0625. The survey leadership views these differences as non-ignorable.

Table 3.3.2 Attrition Bias Estimates

Key Outcome Variable	Method	CO (p-val)	T1 (p-val)	T2 (p-val)
GA	At Close out	0.088 (N/A)	0.077 (0.306)	0.005 (0.290)
	Omitting Mixed Effects	0.079 (N/A)	0.052 (0.297)	0.020 (0.292)

[1] Estimate comes from Pr(Successes <x), where a Success refers to Prioritization AB < No Prioritization AB and Successes ~ Binomial(5,0.5)

Key Outcome Variable	Method	CO (p-val)	T1 (p-val)	T2 (p-val)
	Omitting Non-Compliant	0.084 (N/A)	0.073 (0.305)	-0.016 (0.279)
SNAP	At Close out	1.260 (N/A)	0.919 (0.001)	0.945 (0.016)
	Omitting Mixed Effects	1.137 (N/A)	0.915 (0.001)	1.027 (0.018)
	Omitting Non-Compliant	1.243 (N/A)	0.820 (0.001)	1.081 (0.002)
SSI	At Close out	0.974 (N/A)	0.742 (0.009)	0.718 (0.010)
	Omitting Mixed Effects	1.018 (N/A)	0.753 (0.009)	0.704 (0.010)
	Omitting Non-Compliant	1.009 (N/A)	0.752 (0.010)	0.703 (0.010)
TANF	At Close out	0.111 (N/A)	0.032 (0.211)	0.038 (0.177)
	Omitting Mixed Effects	0.109 (N/A)	0.014 (0.208)	0.024 (0.178)
	Omitting Non-Compliant	0.103 (N/A)	0.016 (0.206)	0.032 (0.182)
WIC	At Close out	0.171 (N/A)	0.218 (0.070)	0.099 (0.059)
	Omitting Mixed Effects	0.124 (N/A)	0.195 (0.068)	0.078 (0.056)
	Omitting Non-Compliant	0.145 (N/A)	0.198 (0.069)	0.053 (0.057)

3.3 Overall Attrition Rates

This section analyzes if the effect of the prioritization on the overall attrition rate. Lowering overall response rate and consequently increasing attrition was initially seen as a risk, no significant differences in the overall attrition rate would be viewed as a positive result. Prioritizing cases increased the overall attrition rate by approximately 0.4 to 1.6 percentage points and decreased the number of completed interviews per interviewer by approximately 0.3 and 0.9, depending on cases omitted for the different analyses. Observing the method with no mixed effects and no-non compliant interviewers, we see the dynamic adaptive prioritization attrition rates are close to the no prioritization attrition rate and not significantly different. This provides evidence that the use of contact history response model may have benefited the overall attrition rate.

Table 3.3.1 Unweighted Attrition Rates

Method	No Prioritization (CO)		Static Adaptive Prioritization (T1)		Dynamic Adaptive Prioritization (T2)	
	Completions per Int.	Attrition Rate	Completions per Int.	Attrition Rate	Completions per Int.	Attrition Rate
Close out	14.6 (N/A)	43.0 (N/A)	13.7 (0.243)	44.3 (0.016)	13.9 (0.358)	44.1 (0.040)
No Mix Effect	14.6 (N/A)	42.9 (N/A)	13.9 (0.367)	44.1 (0.033)	14.2 (0.614)	43.3 (0.302)
No Non-Compliant	13.7 (N/A)	41.8 (N/A)	13.2 (0.500)	43.4 (0.003)	13.4 (0.654)	42.5 (0.0157)

4. Conclusions

We tested a similar strategy to Peytchev et.al. (2010) and saw a mix of positive and negative impacts. As expected, the prioritization led to between 0.3 and 0.9 fewer respondents per interviewer (depending on which analyses are used). However, it located 0.2 to 0.7 more movers that are more likely to be program participants. Unexpectedly, the prioritization retained fewer stationary targets. Ultimately, the experiment did achieve its primary goal of reducing attrition bias in nearly all key outcomes. The reductions in the variation in response propensities are modest as Lynn (2017) described, but not ignorable. We believe that because the response rates in SIPP are lower than the Community Advantage Panel Survey (Peytchev et. al. 2010) there was more room for improvement in attrition bias, or perhaps the use of a secondary contact history model mitigated some of the risks involved. Since many of the negative results were not statistically or practically significant, SIPP leadership has elected to move forward with case prioritization in the future.

This paper provides evidence of adaptive design's impact in this longitudinal study to tackle attrition bias. This means that there is belief that this methodology can significantly reduce the attrition of these cases in future panels.

To truly understand the impact of case prioritization, the research team plans to analyze how case prioritization affected interviewer behavior. In this face-to-face survey, interviewers are generally given discretion on how they work their cases. While there were instructions provided to the interviewers, the protocol were not mandated and there was no oversight into compliance with prioritization procedures, leaving some freedom for the interviewers on how they elected to work these cases. This means the real application of the prioritization is subject to interviewer compliance. While there is no formal definition for interviewer compliance, we develop a post-hoc definition of noncompliance and conduct a modest analysis develop a post-hoc definition omitting the most noncompliant interviewers.

A large limitation to the analysis is that some cases experience both a treatment and a control effect. SIPP reassigns cases to other interviewers for various reasons. Respondent relocation and reluctance, interviewer vacation, and new interviewer hires are some of the reasons households are reassigned. We estimate that this is probably about 11 percent of cases. The paper presents results assuming the interviewer who finalized the case was the interviewer who were actively worked the case and omitting cases that may have experienced control and treatment effects.

Since SIPP plans to continue the usage of case prioritization, the survey research team will need to research the best way of dealing with the limitations noted above. Future research plans include further analyzing the data, by seeing what effect did the instructions provided in Appendix 3 had on interviewer behavior and data collection costs.

Appendix 1

One of the main strategies of this case prioritization is targeting under-represented subpopulations as they relate to program participation. Having a balance of Wave 1 program participants is the most direct way to accurately capture any change in program participation. The research team also felt the importance to balance among those Wave 1 program participants who were not receiving program participation at the time of the interview, as this could be an indicator of possible change. This variable is considered Wave 1 Program Participation with a Gap.

In addition to those two variables, the research team uses the theory that balancing on variables related program participation can better balance the households who receive program participation. To identify these sociodemographic subpopulations, we considered more than 60 variables from SIPP frame data and prior wave data to model program participation. The survey director and the research team chose these variables. For the numeric variables, such as person count and age, we categorized them based on what makes the most sense for the survey. Ultimately, the research team decided on 12 variables (17 including five interactions).

Table A.1. R-indicator Variables

Variables	Description	Values
Wave 1 Program Participation	Did someone in household receive program assistance in Wave 1?	1 - Yes 0 – No
W1 PP Gap	Did someone in household receive program assistance the previous year but not at time of interview?	1 - Yes 0 – No
Person Count		0 - 0 1 - 1 2 - 2 3 - 3 or 4 4 - 5+
Employment Status 2+	At least one person has two or more jobs	1 - Yes 0 – No
Poverty Status	Household considered in poverty	1 - Yes 0 – No
Age of Oldest Household Member		1 - <25 2 - [25,45) 3 - [45,65) 4 - >65
Age of Youngest Household Member		1 - <3 2 - [3,10) 3 - [10,18) 4 - >18
Discouraged Worker		1 - Yes 0 – No
Marital Status		1 - Yes 0 – No
Tenure	Is this house rent/ owned?	1 - Owned 2 - Rent 3 - Other

Urban/Rural	Is this home in a urban or rural area	U - Urban R – rural
Female Householder	Is the householder female?	1 - Yes 0 – No

The interaction variables included in the model are previous year assistance and person count, previous year assistance gap and person count, tenure and person count, and tenure and poverty status, age of oldest individual and age of youngest individual.

Modelling Wave 2 and Wave 3 program participation with Wave 1 data, our model is extremely predictive (percent concordance is 96.2%; using cross-validation, the predictors correctly identified 83.1% of program participants and 97.1% of non-program participants). We emphasize in this section, that while the model is extremely predictive of program participation, it is not our intent to simply target program participants. The goal of this exercise is to dissect the Wave 4 frame, establishing subpopulations related to program participation, in an attempt to attain equal response rates among each subpopulation.

Appendix 2

In order to address the concern that the case prioritization may negatively affect response rates, we use a secondary model that primarily consists of current wave contact history that is used to predict response. This is done so that the interviewers do not waste too many resources on under-represented cases that are reluctant.

The contact history paradata that was included in the model were:

Variables	Description	Values
Previous Wave Response	Did someone in household respond the previous wave?	1 - Yes 0 – No
Previous Wave Contact Attempts	How many contact attempts were needed the previous wave?	
Mover	Is this house a mover?	1 - Yes 0 – No
Current Wave In-person Contact Attempts	How many in-person contact attempts have been made so far?	
Current Wave Phone Contact Attempts	How many phone contact attempts have been made so far?	

Current Wave Interviewee Contact Attempts	How many times as the interviewee reached out to interviewer?	
Interviewee Reluctance	Has the interviewer shown reluctance?	S- Strong T-Time B-Burden P-Privacy N-None
Interviewer Strategy	What strategies were implemented?	1- Scheduled Appointment 2- Left Card 3- No Strategy 4- Other

This model was run daily on every case that was completed, and used to predict every incomplete household. This model was first used at the start of the second month of data collection when approximately 20 percent of cases were completed.
As more cases were completed, the model had better predictions, but even at the start of the second month, the model correctly identified 70 percent of non-respondents. Accurately estimating non-respondents were extremely valuable, because a subset of those cases were made low priority and allowed interviewers to reallocate their resources to cases that were more likely to respond. Toward the end of data collection, these models correctly identified more than 90 percent of nonrespondents and 70 percent of respondents.

Appendix 3

Since interviewers are often left to their own discretion, it is imperative to the experiment's success that the interviewers do as they are instructed. The results presented are under the assumption that the interviewers are compliant with the following instructions:

> "High Priority cases should get first attention each day you work. A contact attempt should always be made within a week of a case being marked High Priority, if a good contact address or telephone number is available. You are encouraged to work High Priority cases as often as necessary to complete them faster. Some High Priority cases will be movers and will require every effort to rapidly locate the household members."
>
> "Low Priority cases are effectively 'on temporary hold,' meaning the sponsor does not want these cases to be worked temporarily."

Acknowledgements

We are thankful to Dr. Benjamin Reist and Dr. Gina Walejko for their valuable guidance and encouragement during the course of this research. We are also thankful to Dr. Peter Miller for taking the time to offer valuable input and comments to the paper. Both parties were very instrumental in the development of this final product.

References

Please verify that all references are included. Please verify that all references are included. Please verify that all references are included.

Please verify that all references are included. Please verify that all references are included. Please verify that all references are included. Please verify that all references are included.

Benedetto, G., Motro, J., & Stinson, M. (2015). *Introducing Parametric Models and Administrative Records into 2014 SIPP Imputations.* Paper presented at the Federal Committee on Statistical Methodology (FCSM) Research Conference, Washington, DC.

Brick, M.J., & Williams, D. (2013), Explaining Rising Nonresponse Rates in Cross-Sectional Surveys. *The ANNALS of the American Academy of Political and Social Science, 645*(1), 36-59.

Church, A. H. (1993), Estimating the effect of incentives on mail survey response rates: a meta-analysis. *Public Opinion Quarterly, 57*, 62–79.

Coffey, S., Reist, B., & Miller, P. (Upcoming). Interventions On-Call: Dynamic Adaptive Design in the 2015 National Survey of College Graduates. *Journal of Survey Statistics and Methodology*

Fumagalli, L., Laurie, H., & Lynn,P. (2013). Experiments with Methods to Reduce Attrition in Longitudinal Surveys. *Journal of the Royal Statistical Society, 176* (2), 499-519.

Giefer, K., Williams, A., Benedetto, G., & Motro, J. (2015). *Program Confusion in the 2014 SIPP: Using Administrative Records to Correct False Positive SSI Reports*. Paper presented at the Federal Committee on Statistical Methodology (FCSM) Research Conference, Washington, DC

Groves, R. (2006), Research Synthesis: Nonresponse Rates and Nonresponse Error in Household Surveys. *Public Opinion Quarterly, 70*, 646–675.

Groves, R. M., & Heeringa, S. G. (2006), Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs, *Journal of the Royal Statistical Society: Series A, 169*, 439–457.

Groves, R.M., & Peytcheva, E. (2008), The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-analysis, *Public Opinion Quarterly*, *72* (2), 167-189.

Hock, E. (1989), SIPP 1984: Estimates from the Panel File. Internal Census Bureau Memorandum from Waite to Butz.

Laurie, H., & Lynn, P. (2008). The Use of Respondent Incentives on Longitudinal Surveys. Institute for Social and Economic Research, ISER working papers.

Lynn, P. (2017), From standarised to targeted survey procedures for tackling non-response and attrition, *Survey Research Methods, 11* (1), 93-103.

James, T. L. (1997) Results of the wave 1 incentive experiment in the 1996 Survey of Income and Program participation. *Proc. Survey Research Methodology Sect. American Statistical Association*, 834–839.

Peytchev, A., Riley, S., Rosen, J., Murphy, J., & Lindblad, M. (2010), Reduction of Nonresponse Bias in Surveys Through Case Prioritization, *Survey Research Methods*, *4*, 21–29.

Rodgers, W. (2002) Size of Incentive Effects in a Longitudinal Study. *Mimeo* . Survey Research Centre, University of Michigan, Ann Arbor.

Schouten, B., Cobben, F., & Bethlehem, J. (2009), Indicators for the Representativeness of Survey Response, *Survey Methodology, 35*, 101–113.

Schouten, B., Cobben, F., Lunquist, P., & Wagner, J. (2016), Does More Balanced Survey Response Imply Less Non-response Bias?, *Statistics in Society, 179*(3), 727-748.

Schouten, B., Shlomo, N., & Skinner, C. (2011), Indicators for Monitoring and Improving Representativeness of Response, *Journal of Official Statistics, 27*, 231–253.

Schouten, B., Peytchev, A., Wagner, J. (2017), *Adaptive Survey Design.* Boca Raton, Florida: CRC Press.

Singer, E., Gebler, N., Raghunathan, T., van Hoewyk, J. & McGonagle, K. (1999) The effect on incentives in telephone and face-to-face surveys. *Journal of Official Statistics, 15*, 217–230.

Survey of Income and Program Participation: 2014 Panel Users' Guide. (2016). (1 ed.) Washington, DC: U.S. Census Bureau.

U.S. Census Bureau, Survey of Income and Participation Overview https://www.census.gov/sipp/

U.S. Census Bureau, Survey of Income and Participation Background and History https://www.census.gov/programs-surveys/sipp/

Tourangeau, R., Brick, J., Lohr, S., & Li, J. (2017), Adaptive and Responsive Survey Designs: A Review and Assessment. *Journal of the Royal Statistical Society, 180*(1), 203-223.

Wagner, D. & Layne, M. (2014), The Person Identification Validation System (PVS): Applying the Center for Administrative Records' Research and Applications (CARRA) Record Linkage Software. https://www.census.gov/content/dam/Census/library/working-papers/2014/adrm/carra-wp-2014-01.pdf

Westra, A., & Nwaoha-Brown, F. (2017). *Nonresponse Bias Analysis for Wave 1 2014 Survey of Income and Program Participation (SIPP) (ALYS-16)*. Washington, DC: U.S. Census Bureau Retrieved from https://www2.census.gov/programs-surveys/sipp/tech-documentation/complete-documents/2014/2014_SIPP_Wave_1_Nonresponse_Bias_Report.pdf

Zabel, J. (1998), An Analysis of Attrition in the Panel-Study of Income Dynamics and the Survey of Income and Program Participation with an Application to a Model of Labor Market Behavior, *The Journal of Human Resources, 33*(2), 479-506.

Zieffler, A. S., Harrning, J.R., & Long, J.D. (2011). Comparing Groups: Randomization and Bootstrap Methods using R. *John Wiley & Sons*. 117-137.