

*Sampling and weighting the
Understanding America Study*

*Marco Angrisani, Arie Kapteyn, Erik Meijer,
Htay-Wah Saw*

Paper No: 2019-004

**CESR-SCHAEFFER
WORKING PAPER SERIES**

The Working Papers in this series have not undergone peer review or been edited by USC. The series is intended to make results of CESR and Schaeffer Center research widely available, in preliminary form, to encourage discussion and input from the research community before publication in a formal, peer-reviewed journal. CESR-Schaeffer working papers can be cited without permission of the author so long as the source is clearly referred to as a CESR-Schaeffer working paper.

Sampling and weighting the Understanding America Study

Marco Angrisani, Arie Kapteyn, Erik Meijer, and Htay-Wah Saw*
Center for Economic and Social Research, University of Southern California

October 16, 2019

Abstract

The Understanding America Study (UAS) is a probability-based, nationally representative internet panel. This document describes the methods used for sampling and weighting UAS surveys. UAS members are drawn from multiple sampling frames, and new batches of are added regularly. For most batches, we use an adaptive sampling algorithm, by first drawing zip codes whose distributions complement the distributions in the panel and then randomly sampling addresses from the United States Postal Service's Delivery Sequence File within the selected zip codes. Accordingly, weights are constructed in two stages. In the first stage, we construct base weights that account for differences in the sampling strategies for different frames. In a second stage, we construct poststratification weights to adjust for selective nonresponse.

1 Introduction

The Understanding America Study (UAS) (Alattar, Messel, & Rogofski, 2018) is a nationally representative internet panel of the civilian noninstitutional population age 18 and over in the U.S. (50 states and D.C.). It started in 2014 and it has been expanded since. It consists of a number of subpanels that are drawn from different sampling frames. It uses an adaptive sampling approach to keep the subpanels representative of their target populations even in the face of selective nonresponse and attrition. Nevertheless, like almost every survey, the UAS uses sampling weights to adjust for imbalances in the distribution of demographics and socio-economic variables. This document describes the details of the sampling methods used and how the weights are constructed. See Tourangeau, Brick, Lohr, and Li (2017) for an overview of adaptive sampling and Lohr (2009) for a review of multiple frame samples and how to weight them.

*We like to thank Mick Couper and Jon Krosnick for their insightful and valuable comments throughout the development of the UAS weighting procedure described in this document.

Table 1: UAS batches and frames

Batch(es)	Frame	Weights
1	1. Nationally representative (simple random sample)	Yes
2, 3	2. Native American (stratified 2-stage)	Partly
4	3. Los Angeles County young mothers (list sample)	No
5–12	4. Nationally representative (2-stage, fixed)	Yes
13, 14, 18, 19	5. Los Angeles County (2-stage)	Yes
15, 16	6. California (2-stage)	Yes
17	7. Nationally representative (2-stage, pps)	Yes

Note. “fixed” indicates a fixed number of sampled households per zip code. “pps” indicates that the number of sampled households is proportional to the number of households in a zip code.

2 Sampling

The UAS is sampled in batches. As of June 2019, there are 19 batches. Table 1 gives an overview. As mentioned above, the UAS draws from multiple frames, but each batch draws from only one frame. Most batches use a two-stage sample design, in which zip codes are drawn first, and then households are randomly drawn from the sampled zip codes. The exceptions are batches 1 and 4, which drew simple random samples from lists. For batch 1, we used the list of all individuals in the database of the vendor ASDE. For batch 4, we used all addresses listed on birth certificates issued in Los Angeles County in the years 2009–2012, subsetted to a certain set of zip codes. When sampling from a list of addresses, we initially contact one person per household. Complete documentation about the UAS recruitment procedure can be found at <https://uasdata.usc.edu/page/Recruitment>.

We provide weights for all batches, except batch 4 and non-Native American households in batches 2 and 3.

The three nationally representative frames represent the same population (the whole U.S.). In terms of populations, California is a subset of the U.S., and Los Angeles County is a subset of California. So these are fully nested. Native Americans are also a subset of the U.S., and they partially overlap with the other subsets.

The remainder of this section discusses our adaptive sampling algorithm for drawing a new batch, given the composition of the UAS at the time of sampling.

2.1 Basic notation and definitions

We look at the marginal distributions of categorical variables and how adding a new batch changes these. This is without loss of generality, because the joint distribution of two or more categorical variables is the marginal distribution of a new categorical variable that has one category for each combination of categories for the original variables. Let the number of variables of interest be M , and be indexed by $m = 1, \dots, M$.

The marginal distribution of variable m in the existing panel consists of the fractions f_{mj} ,

$j = 1, \dots, J_m$. That is, f_{mj} is the number of panel members with value j of variable m divided by the total number of panel members. If desired, this (and all related concepts) can be defined on a subset of the panel, such as active panel members, active panel members living in the State of California, or active panel members living in Los Angeles County (these are the most common subsets we will consider). The number of relevant panel members is n_0 . Hence, the number of relevant panel members in category j of variable m is $n_{0mj} = n_0 f_{mj}$.

Analogously, the marginal distribution of variable m in the population of interest consists of the fractions g_{mj} , $j = 1, \dots, J_m$, and the marginal distribution of variable m in the population of zip code k consists of the fractions h_{kmj} , $j = 1, \dots, J_m$.

2.2 Sampling from a zip code

Suppose we draw a simple random sample of size n_1 from zip code k . In expectation, $n_{1kmj} = n_1 h_{kmj}$ of the individuals in the sample will have value j of variable m . Not all of these will eventually become panel members. Let r_{mj} be the response rate (fraction of those sampled who will become panel members) in category j of variable m .¹ Then, in expectation, this sample will add $n_{2kmj} = n_1 h_{kmj} r_{mj}$ individuals in category j of variable m . Hence, once the recruitment is completed, we expect to have $\hat{n}_{mj} = n_{0mj} + n_{2kmj}$ individuals in category j of variable m and $\hat{n} = n_0 + n_{2k}$ panel members, where

$$n_{2k} = \sum_{j=1}^{J_m} n_{2kmj},$$

which should not depend on m . Consequently, in expectation, the fraction of the sample that is in category j of variable m after drawing zip code k will be

$$\hat{f}_{kmj} = \frac{\hat{n}_{mj}}{\hat{n}} = \frac{n_0 f_{mj} + n_1 h_{kmj} r_{mj}}{n_0 + n_1 \sum_{\ell=1}^{J_m} h_{km\ell} r_{m\ell}}.$$

2.3 Desirability criteria

Our goal is to draw a probability sample with potentially unequal probabilities such that the panel likely becomes more representative of the population of interest. To operationalize this, we need a measure of representativeness and a function that translates this into a probability of drawing a zip code.

For the Native American batches, we used a measure of effective sample size (ESS) as our measure of representativeness. See Appendix B for its description. For the nationally representative batches (5–12, 17 at the time of writing), Los Angeles County batches (13, 14, 18, 19), and

¹We could allow for response rates that vary by zip code, but for the most part, we will assume a constant response rate.

California batches (15, 16), we used a Pearson chi-square (χ^2) measure, defined as

$$D(f, g) = \sum_{m=1}^M \sum_{j=1}^{J_m} \frac{(f_{mj} - g_{mj})^2}{g_{mj}}. \quad (1)$$

Thus, for each prospective zip code k , we compute the measure $D_k = D(\hat{f}_k, g)$. The probability of drawing zip code k is then a function of D_k . For the ESS criterion, larger values are more desirable, so the probabilities should be increasing in ESS. For the χ^2 measure, smaller values are more desirable, so the probabilities should be decreasing in D_k .

2.4 Probability of drawing a zip code

The probability of drawing a zip code should be a monotonically increasing function of its desirability score (ESS or χ^2). However, we have no theory indicating what the specific form of this relation should be. Choosing the “most desirable” zip code leads to a nonrandom algorithm and, in general, may be prone to distortions if this zip code is unusual on dimensions that are not considered. Hence, we prefer a relation that is “not too flat” and “not too steep.” That is, there is sufficient randomness, but not so much that the sampling is indistinguishable from simple random sampling, and the sample is not pushed toward representativeness. In Appendix A, we report results from some simulations in which we considered various criteria and various values of their tuning parameters. We found that the following algorithm gave satisfactory results:

1. Compute the quantile Q_k of D_k . That is, the zip code with the smallest value of D_k gets $Q_k = 1/K$, the next gets $Q_k = 2/K$, and so forth, and the least desirable zip code gets $Q_k = 1$.
2. Compute the utility function

$$U_k = -\frac{Q_k^{1-a}}{1-a},$$

where a is a tuning parameter. For all values of a , U_k is a decreasing function of Q_k . We found that $a = 1.5$ is satisfactory. With this value, U_k has positive values. It decreases steeply for small values of Q_k and more gradually for larger values of Q_k , thus favoring zip codes with small Q_k but giving fairly even chances for all other zip codes. Because Q_k is monotonically increasing in D_k , zip codes with small D_k (more desirable) are favored.

3. Compute $w_k = b + U_k$, where b is chosen such that the ratio between the highest and lowest value of w_k is equal to R , which is a tuning parameter that we can choose. Based on our simulations, we chose $R = 100$. Given our choice of functions and tuning parameters, all w_k are strictly positive.
4. Compute $p_k = w_k / \sum_{\ell=1}^K w_\ell$. The probability of drawing zip code k is p_k .

2.5 Practicalities and population size

Suppose that, starting from the distributions f , g , and $\{h_k\}$, we concatenate the fractions for the different dimensions m and then rescale such that the fractions sum to 1. That is, the pair (m, j) is mapped to a unique index μ , say, and $\tilde{f}_\mu = f_{mj}/M$, and analogously for the other distributions. The resulting vectors \tilde{f} , \tilde{g} , and $\{\tilde{h}_k\}$ then have all the properties of a one-dimensional distribution. If we use these in our computations, this leaves all results essentially unchanged. Specifically, this results in desirability measures $\tilde{D}_k = D_k/M$, but these result in the same quantiles Q_k as in the original definition and thus the same probabilities of drawing the zip codes. Doing this has some computational advantages.

For the nationally representative, LA County, and California batches, we use the distributions of the following characteristics:

- sex \times age \times race/ethnicity: $\{\text{male, female}\} \times \{18\text{--}34, 35\text{--}44, 45\text{--}54, 55\text{--}64, 65+\} \times \{\text{non-hispanic white alone, non-hispanic black alone, non-hispanic Native American alone, non-hispanic Asian alone, other non-hispanic, hispanic}\};$
- sex \times age \times marital status: $\{\text{male, female}\} \times \{18\text{--}34, 35\text{--}44, 45\text{--}54, 55\text{--}64, 65+\} \times \{\text{married, divorced/separated, widowed, never married}\};$
- sex \times age \times education: $\{\text{male, female}\} \times \{18\text{--}34, 35\text{--}44, 45\text{--}64, 65+\} \times \{\text{less than high school, high school/GED, some college/associate's degree, college graduate, postgraduate}\};$
- population size \times urbanicity: $\{\text{five quintiles}\} \times \{\text{rural, mixed, urban}\}$ (starting with batch 13);
- region: $\{\text{LA County, rest of CA, rest of the U.S.}\}$ (starting with batch 15).

These choices are partially based on what is available in the ACS 5-year zcta-level files, which we use to estimate the distributions h_k of the zip codes, and partially based on practicalities of what categories are expected to give reasonable cell sizes in the sample. Population size, urbanicity, and region are zip code level variables, but we compute their distributions weighted by population size to make them comparable to the corresponding distributions in the panel.

Initially, the algorithm had a tendency to draw small zip codes. A plausible explanation for this is the law of large numbers: if the U.S. population were randomly allocated to zip codes, the larger zip codes would tend to have “average” characteristics, whereas smaller zip codes would span a wider range from “undesirable” to “desirable”. The desirable small zip codes would be more likely to be drawn. Therefore, starting with the LA County 2 batch (batch 13), we have added zip code population size quintile by urbanicity as an additional variable used in evaluating desirability, which should help counterbalancing this. In batches 14, 18, and 19 (LA County 3–5), we have implemented a further variation on the sampling scheme that tackles this directly. Specifically, the sampling probabilities of the zip codes are now computed as $p_k = N_k w_k / \sum_{\ell=1}^K N_\ell w_\ell$, where N_k is the population size (initially number of residents age 18+, now number of households) of the

zip code. Thus, if w_k were constant, this would be proportional to size sampling. Because the w_k are not constant, this is not proportional to size sampling, but it is in the same spirit. In batches 15–17, instead of adjusting the sampling probability at the zip code level, we drew the number of addresses per zip code proportional to size, which has a similar effect but draws a more even fraction of the population from each sampled zip code.

For most batches, we drew zip codes without replacement (within and across batches). However, because the number of zip codes is small relative to sample size for the LA County batches, we draw them with replacement. The number of addresses we drew per zip code in each batch was 40, except for the LA County batches, where we drew 10 per zip code each time the zip code was drawn, and the batches where we drew the number of addresses proportional to size. There, the target *average* number of addresses was 40.

3 Base weights

The procedure to sample zip codes in the batches that use two-stage sampling has been refined over time, but we treat this as random variation within each frame. Hence, for constructing weights, we combine the batches that use the same frame as indicated in Table 1. Because the second stage of the two-stage sampling differs between batches 5–12 (40 addresses per zip code) and 17 (number of addresses proportional to number of households per zip code), we treat these as separate frames. There are some variations in the procedure between frames that we also take into account explicitly.

The procedure for constructing base weights follows the discussion in Lohr (2009, section 2.4). It consists of these steps:

1. Estimate the probability $\pi_{k|s}$ of drawing zip code k in sampling frame s with a logit model that uses zip code characteristics as regressors. The characteristics used are Census region, urbanicity, zip code population, and demographic distributions (gender, age, race-ethnicity, marital status, education). For the frames that use single stage sampling, we treat the whole list as a single zip code and thus $\pi_{k|s} = 1$. The Los Angeles County two-stage frame uses with-replacement sampling of zip codes, and this requires an adaptation that estimates the expected number of times a zip code is drawn instead (see Appendix E).
2. Estimate the probability $\pi_{h|sk}$ of drawing household h conditional on its zip code k being drawn, in sampling frame s . This is equal to the number of households sampled divided by the number of households in the zip code.
3. The marginal probability that household h from zip code k is sampled in frame s is then $\pi_{hk|s} = \pi_{h|sk}\pi_{k|s}$. Ignoring the negligible probability that a household is sampled multiple times (in multiple samples), the overall probability that household h from zip code k is sampled in any sample is $\pi_{hk} = \sum_s \pi_{hk|s}$. Note that computing this requires the computation of the counterfactual conditional probabilities of being sampled in a different frame from the one the household was actually drawn from. Some of these may be zero, for example,

the probability that a household from Colorado is drawn from the California frame is zero, but this does not hold for other batches or households.

4. The raw base weight is $w_{hk}^{\text{raw}} = 1/\pi_{hk}$.
5. We calibrate the raw base weights to match known subpopulation totals. To this end, we construct a minimal set of mutually exclusive and exhaustive subpopulations such that each frame is the union of one or more of these subpopulations. Let $g = 1, \dots, G$ denote these subpopulations.

Compute calibrated base weights as $w_{hk}^{\text{cal}} = A_g w_{hk}^{\text{raw}}$, where the correction factor A_g is such that the weights w_{hk}^{cal} for subpopulation g sum to the number of households in this subpopulation as estimated from the American Community Survey (ACS).

6. Compute relative base weights $w_{hk}^{\text{base}} = B w_{hk}^{\text{cal}}$, where the constant B is chosen such that the weights w_{hk}^{base} sum to the number of sampled households, and thus the average base weight in the drawn sample is 1.

In order to generate sample weights for each UAS survey, we compute poststratification weights, using w_{hk}^{base} as the base weight in a raking procedure. This second step is discussed in detail in Section 5.

Note that it is implicit in the procedure described above that the goal is to combine the samples from the different frames into a single sample with a single sampling weight. An estimator using this weighted sample is called a *single-frame estimator* (Lohr, 2009, section 2.4). This is not necessarily the preferred approach in terms of statistical efficiency. Most of the literature on multiple-frame sampling discusses estimators that are optimal combinations of separate estimators by frame and subpopulation (Lohr, 2009, section 2.2). Typically, because optimality depends on the variances and covariances of the estimators, for different parameters the optimal combination will be different. For our purposes, however, this is very impractical. We disseminate each survey as a single dataset with a single weight. Hence, the single-frame weight is preferred on practical grounds, while still statistically sound.

3.1 Incorporating the Native American batches

The Native American batches introduce some complications. In sampling these, the list of zip codes was reduced to zip codes with at least 5% of the population American Indian or Alaska Native (AIAN). Sampling was stratified by whether the zip code was in an ‘‘American Indian Area’’ according to the Census Bureau or not. Zip codes in which this information was missing, or that were mixed, were excluded. Also, zip codes with missing information about urbanicity were excluded.² The resulting set of zip codes formed the population from which zip codes were

²Zip codes that had missing information about population size or fewer than 100 residents were also excluded, but this holds for all batches that draw zip codes.

sampled. Sampling was done from two strata: stratum 1 contained American Indian Areas and stratum 2 other zip codes. Based on estimates of the Native American population in both strata, it was decided to draw 50 zip codes from stratum 1 and 371 from stratum 2, to obtain about one third of our sample from American Indian Areas, which roughly corresponds to the population fraction among AIANs. Within each stratum, a variant of our sampling algorithm was used that oversampled zip codes that would increase the effective sample size relative to cost the most. The variables used in this were Census Division (the nine state groupings used by the Census Bureau), urbanicity (urban, rural, mixed), population size, and fraction of the population that is AIAN. See Appendix B for more details.

The list of drawn zip codes was submitted to the vendor ASDE, which limited their database to these zip codes and drew a simple random sample of 10,000 individuals from this subset. Hence, in expectation, the sample sizes in each zip code are proportional to the population sizes. We used a random 5,000 out of these 10,000 sampled individuals for batch 2. For batch 3, we excluded the zip codes drawn for batch 2 and drew another sample in the same way.

For constructing the base weights, this has the following consequences:

- The probability of drawing a zip code is estimated separately for AI Area and non-AI Area zip codes.
- Covariates for estimating the probability of selecting a zip code within these strata are Census Division, urbanicity, log population size, and the log of the AIAN fraction of the population.
- The probability that a household from zip code k is sampled conditional on the set of sampled zip codes is approximately $2 \times (5,000/10,000) \times (10,000/N^*)$ where N^* is the population size of the combined $2 \times (50 + 371) = 842$ zip codes that were drawn.

Non-Native Americans were not eligible to become panel members.³ We accidentally invited some of them a year later to become panel members, which led to a number of non-Native Americans recruited through Native American batches joining the UAS. Because we are unable to attach a probability to this happening, these panel members receive a weight of zero and, for the purpose of creating base weights, the probability of recruiting non-Native Americans in Native American batches is considered to be zero.

3.2 Los Angeles County young mothers

This subsample of the UAS (batch 4) was a simple random sample from a list of women who had given birth in Los Angeles County between 2009 and 2012 in zip codes around restaurants participating in a healthy menu options project. The invitation was extended to anyone living at the address at the time of sampling. Because of the highly specific nature of the subsample, we do not provide weights for members recruited within this batch.

³At this stage, we used a wider definition of “Native American,” which included Native Hawaiians and Pacific Islanders in addition to AIANs.

Table 2: Subpopulations with their estimated number of households and batches that were (partially) drawn from them

Region	NA zip	Households			Batches
		Total	NA	Non-NA	
LA County	yes	29,518	1,716	27,802	1–3, 5–19
LA County	no	3,249,074	66,762	3,182,312	1, 5–19
Other CA	yes	133,710	7,325	126,385	1–3, 5–12, 15–17
Other CA	no	9,394,863	267,292	9,127,571	1, 5–12, 15–17
Non-CA	yes	2,562,811	352,386	2,210,425	1–3, 5–12, 17
Non-CA	no	102,345,040	1,771,871	100,573,169	1, 5–12, 17
<i>Subtotals and totals</i>					
CA incl. LAC	yes	163,228	9,041	154,187	
CA incl. LAC	no	12,643,937	334,054	12,309,883	
U.S. total	yes	2,726,039	361,427	2,364,612	
U.S. total	no	114,988,977	2,105,925	112,883,052	
LA County	any	3,278,592	68,478	3,210,114	
Other CA	any	9,528,573	274,617	9,253,956	
Non-CA	any	104,907,851	2,124,257	102,783,594	
CA incl. LAC	any	12,807,165	343,095	12,464,070	
U.S. total	any	117,715,016	2,467,352	115,247,664	

Note. NA = Native American

3.3 Subpopulation totals

Table 2 shows the subpopulation totals for the subpopulations g used, as discussed above. These totals have been estimated from the 2012–2016 5-year average ACS file.

3.4 Sampling over time and dynamic populations

The construction of the base weights implicitly assumes a one-time sample from a well-defined population. However, the UAS is a panel that has grown (and continues to grow) over time, and surveys in the UAS are done on a regular basis. This has various potential consequences for the sampling and weighting.

- **Dynamic populations:** mortality, “births” (i.e., individuals turning 18), and migration mean that populations are changing. We assume that this is slow enough that it does not affect the composition of our population(s) of interest. However, specific individuals regularly move into and out of populations, especially subpopulations. For example, individuals

sampled as part of the LA County batches may not live in LA County anymore. The base weights do not depend on specific individuals currently in a given survey, but rather the households that were sampled, whether or not they responded and became UAS panel members. Hence, the base weights are computed for the frame and population they were in at the time of sampling. In contrast, the poststratification weights for a given survey use current characteristics, population membership, and marginal distributions (from the CPS). A related issue is that zip codes are sometimes updated. Technically, we are not using zip codes in the sampling but *zip code tabulation areas* (zcta). These do not change very often, typically once every 10 years with a new census, and even then most zcta's stay the same. So this is not currently a problem and we do not expect this to become a problem in the next few years.

- Changing panel composition. Drawing a new batch changes the base weights of all batches. This also means that which base weights should be used in the raking of a specific survey depends on when the survey was administered, which batches were included in the sample of panel members that were invited for the survey, and other selection criteria. As far as the latter are concerned, UAS surveys typically use three different reference samples:
 - Nationally Representative batches only (NR)
 - All batches excluding Native American batches (All_noNA)
 - All batches including Native American batches (All)

To address this, we compute a separate set of base weights for surveys in which the reference sample is NR, All_noNA, or All. For each of these sets, we compute multiple base weight variables, depending on which batches were invited to the survey. Consider, for example, two surveys based on NR. Suppose that one only includes batch 1, while the other includes batch 1 and batches 5-12. These two surveys will have different base weights, reflecting that for both the reference sample was NR, but different batches were included.

Formally, for each survey $w_{hk}^{\text{base}} = w_{hk}^{\text{base}}(\Sigma, b)$, where $\Sigma \in \{\text{NR}, \text{All_noNA}, \text{All}\}$ and b indicates that only batches up to b were invited to the survey.

4 Sample Selection for UAS Surveys

The selection of UAS members who are invited to take a survey follows these steps:

1. Identification of reference sample Σ .

In this step, we determine whether respondents should be chosen from Nationally Representative batches only (NR), all batches excluding Native American batches (All_noNA), or all batches without restrictions (All)

2. Identification of included batches b .

Once the reference sample Σ has been determined, the exact batches from which survey invitees are chosen from is represented by all available batches within Σ at the time the sample selection is performed.

3. Selection of survey invitees.

Assuming that the target sample size for a specific survey is τ and the expected response rate is r , a random sample of size τ/r (approximated to the nearest integer) is selected within the reference sample Σ and available batches b . We use a constant response rate across surveys calculated as the average response rate across all UAS surveys in the last year.

Based on the above steps, the appropriate set of base weights for each specific survey, $w_{hk}^{\text{base}} = w_{hk}^{\text{base}}(\Sigma, b)$, is identified as a function of the reference sample Σ and the available batches b which survey invitees were drawn from.

Several “core” or “general” surveys are administered to all existing panel members. In this case, $\Sigma = \text{All}$ and b is all available recruitment batches in the UAS. As of September 2019, the appropriate set of base weights for a “core” or “general” survey would be $w_{hk}^{\text{base}} = w_{hk}^{\text{base}}(\text{All}, 19)$. It is worth reminding that $w_{hk}^{\text{base}} = 0$ for batch 4 and non-Native Americans recruited through Native American batches.

5 Poststratification

Due to selective nonresponse, the sample of actual respondents may have different characteristics than the population of interest, even after correcting for different sampling probabilities through the base weights. Therefore, surveys typically perform another step of weighting, matching distributions of (usually) demographic characteristics to external distributions. This is called *poststratification*.⁴ We do this as well in the UAS.

In this second step, we perform *raking* weighting (also known as iterative marginal weighting), starting from the base weights, w_{hk}^{base} , described above. With this, we assign poststratification weights to survey respondents such that the weighted distributions of specific socio-demographic variables in the survey sample match their population counterparts (benchmark or target distributions).

The benchmark distributions against which UAS surveys are weighted are derived from the Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC) administered in March of each year. We use the latest available CPS-ASEC at the time a survey is completed. Since the CPS-ASEC becomes available at the end of each year, typically there exists a mismatch

⁴Sometimes, the term poststratification is used in a more narrow sense, only applying to cell-based weighting as opposed to raking, which we actually perform. For example, Valliant, Dever, and Kreuter (2013, sec. 14.2) make this distinction and call the whole class of weighting methods that uses external information *calibration*, which then includes both poststratification and raking as special cases.

Table 3: Recoded variables for poststratification

Recoded variable	Categories
Gender	1. Male; 2. Female
Age	1. 18–39; 2. 40–49; 3. 50–59; 4. 60+
Born in the US	0. No; 1. Yes
US citizen	0. No; 1. Yes
Education	1. High school or less; 2. Some college; 3. Bachelor or more
Native American	0. No; 1. Yes
Race-ethnicity	1. White; 2. Black; 3. Others; 4. Hispanic
Census region ^a	1. Northeast; 2. Midwest; 3. South; 4. West
Urbanicity ^b	1. Rural; 2. Mixed; 3. Urban
Marital status	1. Married; 2. Separated/divorced/widowed; 3. Never married
Work status	1. Working; 2. Unemployed; 3. Retired; 4. On leave, disabled, other
Household composition	1. 1 Member; 2. 2 Members; 3. 3 or 4 Members; 4. 5 or More Members
Household income	1. <\$30,000; 2. \$30,000–\$59,999; 3. \$60,000–\$99,999; 4. \$100,000+

^aCensus regions are obtained from the respondent’s zip code of residence.

^bUrbanicity is obtained from the respondent’s zip code of residence.

of one year between the year of survey completion and the year of the CPS-ASEC used to derive benchmark distributions. We provide updated weights based on the same CPS-ASEC year as the one of survey completion to users on request.

Unless otherwise required by the aims of the survey and specified in the sample selection process, the reference population for UAS surveys is the U.S. population of adults, age 18 or older, excluding institutionalized individuals and military personnel.

5.1 Categorization and imputation of raking factors

For poststratification weighting purposes, we use demographic information taken from the most recent *My Household* survey, which is answered by all active UAS members every quarter. All socio-demographic variables in the *My Household* survey are categorical, but some, such as age, education, and income, take values in a relatively large set. We recode all socio-demographic variables considered for poststratification into new categorical variables with no more than 5 categories. The aim of limiting the number of categories is to prevent these variables from forming strata containing a very small fraction of the sample (less than 5%), which may cause sample weights to exhibit considerable variability.

The list of all recoded categorical variables that may be used to generate poststratification weights is reported in Table 3. As explained below, default poststratification weights are produced using a predefined set of socio-demographic variables. Upon request, custom poststratification weights can be produced using an alternative set of socio-demographic variables chosen among those listed in Table 3.

Because the raking procedure can only be applied to nonmissing data, we impute missing values of socio-demographic variables according to the following sequential scheme:

- Gender is obtained from administrative information.
- When age is missing, we use the age range variable from the *My Household* survey to impute age categories. If the age range is also missing, the gender-specific sample mode is used.
- Once age categories have been imputed (if missing), the variable with the fewest missing values is the first one to be imputed by means of a regression featuring gender and the age categories as regressors. This newly imputed variable is then added to the set of regressors to impute the variable with the second smallest number of missing values. The procedure continues in this fashion until the variable with the most missing values is imputed using information on all other available socio-demographic variables.

For binary indicators, such as Born in the US and US citizen, missing values are imputed using a logistic regression. For ordered categorical variables, such as education, household composition, and household income, missing values are imputed using an ordered logistic regression. For unordered categorical variables, such as marital status, race-ethnicity, and work status, missing values are imputed using a multinomial logistic regression. Census region and urbanicity are never missing, as they are obtained from respondents' zip codes of residence.

Each UAS survey data set including sample weights also contains a binary variable (imputation flag) indicating whether any of the recoded socio-economic variables used for poststratification has been imputed or taken from administrative data not available to data users. As of September 2019, this variable takes value 1 for 0.13% of UAS members.

5.2 Raking and trimming algorithm

As mentioned above, we use raking to generate poststratification weights. This procedure involves the comparison of target population relative frequencies and sample relative frequencies on a number of socio-demographic variables independently and sequentially. More precisely, starting from the base weights, at each iteration of the algorithm weights are proportionally adjusted so that the distance between survey and population marginal distributions of each selected socio-demographic variable (or *raking factor*) decreases. The algorithm stops when survey and population distributions are perfectly aligned. A maximum of 50 iterations is allowed for perfect alignment of survey and population distributions to be achieved. If the process does not converge within 50 iterations, no sample weights are returned and attempts using different raking factors are made.

Our raking algorithm trims extreme weights in order to limit variability and improve efficiency of estimators. We follow the general weight trimming and redistribution procedure described by Valliant et al. (2013, p. 388). Specifically, we define $N = N_w + N_{nw}$ the total sample size, where N_w is the number of respondents who receive a weight, and N_{nw} includes respondents belonging

to batch 4 and respondents from non-Native American households belonging to batches 2 and 3, who do not receive a weight. Indicating with w_i^{rak} the raked weight of respondent $i = 1, \dots, N_w$, the sample average of the raked weights is

$$\bar{w}^{\text{rak}} = \frac{1}{N_w} \sum_{i=1}^{N_w} w_i^{\text{rak}},$$

which is used to perform the following operations:

- We set the lower and upper bounds on weights equal to $L = 0.25\bar{w}^{\text{rak}}$ and $U = 4\bar{w}^{\text{rak}}$, respectively. These are in line with those often mentioned in the literature and adopted by other surveys (Battaglia, Hoaglin, & Frankel, 2009).
- We reset any weights smaller than the lower bound to L and any weights greater than the upper bound to U :

$$w_i^{\text{trim}} = \begin{cases} L, & \text{if } w_i^{\text{rak}} \leq L; \\ w_i^{\text{rak}}, & \text{if } L < w_i^{\text{rak}} < U; \\ U, & \text{if } w_i^{\text{rak}} \geq U. \end{cases}$$

- The amount of weight lost by trimming is

$$w^{\text{lost}} = \sum_{i=1}^{N_w} (w_i^{\text{rak}} - w_i^{\text{trim}}),$$

and distribute it evenly among the respondents whose weights are not trimmed.

While the raking weights match population distributions of selected variables, the trimmed weights typically do not. We therefore iterate the raking and trimming until poststratification weights are obtained that respect the weight bounds and align sample and population distributions of selected variables. This procedure stops after 50 iterations if an exact alignment respecting the weight bounds cannot be achieved. In this case, the raked weights will ensure an exact match of (weighted) survey relative frequencies to their population counterparts, but the weights may take values outside the interval defined by the prespecified lower and upper bounds.

5.3 Final poststratification weights

Indicate by w_i^{post} the final poststratification weight for respondent i , obtained by applying the raking algorithm to the base weights and after iterating the raking/trimming procedure as described above.

UAS survey data sets include final poststratification weights relative to their sample mean. That is, they include

$$w_i^{\text{final}} = \frac{w_i^{\text{post}}}{\frac{1}{N_w} \sum_{j=1}^{N_w} w_j^{\text{post}}}$$

for the N_w respondents who receive a weight, and $w_i^{\text{final}} = 0$ for the N_{nw} respondents who do not. Hence, relative final poststratification weights sum to the size of the sample of respondents who receive a weight in each survey, N_w , and average to 1 within that sample.

5.4 Default weights

By default, all UAS surveys are weighted using the following set of raking factors (as defined in Table 3): gender, race-ethnicity, age, education, and Census region. For surveys that include respondents from Native American batches, the Native American indicator is added to the default set of raking factors.

We have carried out extensive testing and concluded that raking weights produced by this combination of factors perform well across different dimensions. In particular, they exhibit moderate variability, thereby leading to better precision of weighted estimators, and lead to a satisfactory match of the distributions of variables not used as raking factors, thereby improving overall representativeness. Our Monte Carlo studies have shown that these desirable properties are robust to sample sizes ranging from 500 to 2,000 respondents, an interval that includes most of the UAS surveys.

Table 4 shows the distributions of raking factors in the entire UAS sample as of September 2019. In column 1, we report unweighted percentages, in column 2 the percentages after applying base weights, and in column 3 the percentages after applying final weights. By construction, these match population benchmarks, which are taken from the 2018 CPS. The unweighted sample size (excluding respondents from batch 4 and non-Native Americans recruited through Native American batches) is 7,787 and the effective sample sizes using base weights and final poststratification weights are 3,893 (51% of 7,787) and 3,660 (47%), respectively.

Table 5 shows the distributions of some of the variables described in Table 3 that were not used as raking factors. In column 1, we report unweighted percentages, in column 2 the percentages after applying the final weights, and in column 3 the percentages in the population (taken from the CPS-ASEC 2018 for all variables, except for the urban indicator, whose benchmark is taken from the 2017 ACS).

Completed surveys with default weights are available for download on the UAS webpage. Each survey data set contains the following variables generated by the weighting procedure: base weight, imputation flag, CPS year used, final weight.

5.5 Custom weights

On request, we provide custom weights to users. These can be obtained by choosing which socio-demographic variables, among those listed in Table 3, should be used by the raking algorithm to generate poststratification weights.

Raking can then be performed on one-way marginals, by matching population distributions of single socio-demographic variables, such as gender or education, as well as on two-way marginals,

Table 4: Unweighted and weighted distributions of the raking factors (%)

Variable	Unweighted	Weighted with	
		Base weights	Final weights ^a
Male	41.58	44.11	48.36
<i>Race-ethnicity</i>			
White	63.55	73.65	63.39
Black	7.97	9.58	11.93
Other	10.47	7.83	8.51
Hispanic	18.01	8.94	16.17
Native American	7.71	5.21	2.50
<i>Age (years)</i>			
18–39	32.39	28.35	38.02
40–49	18.63	19.07	16.12
50–59	19.43	20.22	17.15
60+	29.55	32.36	28.71
<i>Education</i>			
High school or less	23.21	19.44	39.67
Some college	37.62	34.83	28.09
Bachelor or more	39.16	45.74	32.25
<i>Census region</i>			
Northeast	9.98	14.64	17.68
Midwest	21.80	27.88	20.76
South	25.28	36.38	37.86
West	42.94	21.09	23.70

Notes. Entire UAS sample as of September 2019 ($N = 7,787$). Population benchmarks from CPS-ASEC 2018.

^a These match their population counterparts by construction.

by matching the distributions of interaction variables, such as gender \times education. The preferred set of raking factors may feature both single and interaction variables, such as, for instance, race-ethnicity and gender \times education. The use of two-way marginals corrects for discrepancies between distributions referring to specific subgroups that would not be accounted for by using one-way marginals alone. The customized set is limited to a maximum of 6 variables (single variables, interaction variables or a combination of both). Restrictions on the number and type of raking factors are imposed to ensure convergence of the algorithm and to reduce weight variability.

Table 5: Unweighted and weighted distributions of selected variables that were not used in the raking (%)

Variable	Unweighted	Weighted with final weights	Population
<i>Household size</i>			
1 member	15.91	15.14	14.74
2 members	41.39	40.48	34.78
3 or 4 members	31.54	32.08	35.70
5 or more members	11.16	12.29	14.78
<i>Household income</i>			
<\$30,000	25.80	27.71	22.49
\$30,000–\$59,999	25.67	25.94	26.89
\$60,000–\$99,999	23.44	22.64	23.77
≥\$100,000	25.08	23.71	26.85
<i>Urbanicity</i>			
Urban	39.82	38.17	38.40
Rural or mixed	60.18	61.83	61.60
<i>Work status</i>			
Working	61.85	62.11	59.59
Retired	19.79	18.64	18.71
Other	18.36	19.25	21.70
<i>Employment type</i>			
Government	18.42	16.14	13.76
Private	68.03	71.23	76.19
Self-employed	13.55	12.63	10.05

Notes. Entire UAS sample as of September 2019 ($N = 7,787$). Population benchmarks from CPS-ASEC 2018 and ACS 2017 for urban indicator only.

Appendix

A Empirical background for our sampling algorithm

As stated in the main text, we developed our adaptive sampling algorithm as a result of our experiences in the American Life Panel (ALP). In the ALP, we observed that selective nonresponse and attrition had led to discrepancies between the distribution of certain characteristics in the panel and the corresponding distribution in the population, for example, in the CPS. Specifically, the ALP had an underrepresentation of low-income individuals. Therefore, a specific “vulnerable populations” sample was drawn that targeted low-income neighborhoods. However, this resulted in an overrepresentation of individuals in urban areas. This led to the question of whether we could find a more sophisticated way to draw additional samples, such that observed discrepancies in certain variables could be reduced without introducing discrepancies in other variables.

When formalizing the problem theoretically, we decided to frame it as a comparison of discrete distributions, which often is reasonable in practice and allows for more straightforward computations, for example, computing the variance of the weights. Practically, we realized that we needed to base the sampling on zip code characteristics, and thus a two-stage sampling procedure in which zip codes were drawn first and addresses within the drawn zip codes second.

For comparing discrete distributions, the Pearson chi-square presented in section 2 is the natural metric. The goal of reducing this chi-square by drawing additional samples then involves a number of choices:

- How many zip codes will be drawn;
- How many addresses will be drawn per zip code; this could depend on characteristics of the zip code, especially its population;
- A function mapping the characteristics of the zip code into the probability of drawing the zip code.

As to the latter, we write this in two parts: (i) a function $w(\cdot)$ such that $w_k = w(D_k)$ is a measure of the desirability of drawing zip code k , where D_k was defined in section 2; and (ii) a rescaling of these into probabilities of sampling: $p_k = w_k / \sum_j w_j$.

Using the existing ALP as the basis, we simulated various choices. The results presented here were computed in the Spring of 2013, using the distribution of the ALP at that time and zip code distributions from the most recent ACS data at the time. We presented this at the Panel Survey Methods Workshop in Ann Arbor, MI in May 2014 (Angrisani, Kapteyn, Meijer, & Saw, 2014).

It turns out that D_k does not vary greatly across zip codes, and hence the same holds for the probabilities if the mapping does not have much curvature. An example of a function we tried is

$$w_k = \frac{1}{1 + cD_k}$$

for some parameter c , which led to sampling probabilities that did not vary much across zip codes. We then looked at transforming the D_k into quantiles Q_k and choosing a transformation that translates quantiles into probabilities, specifically first computing the “utility” function

$$U_k = -\frac{Q_k^{1-a}}{1-a}$$

and then computing $w_k = b + U_k$ for some parameter b . The utility function U_k is the negative of the well-known Constant Relative Risk Aversion (CRRA) utility function, with risk aversion parameter a . From micro-economic studies, we know that this function allows for a wide range of curvatures indexed by a . The minus sign maps higher quantiles (and thus higher discrepancies D_k) into lower utilities.

The parameter b is then a translation that makes all the desirabilities w_k positive. Moreover, we can choose this such that the ratio between the highest and lowest value, and thus the ratio between the highest and lowest sampling probability, is some prespecified amount R .

In our simulations, we also looked at the following methods:

- Equal weights ($w_k = 1$) and thus equal probabilities of sampling the zip codes.
- Probability proportional to size of the zip code, $w_k = N_k$, the zip code population size.
- Pick the zip code with the smallest D_k .

The latter is guaranteed to reduce the discrepancy the fastest, but has the drawback that it does not result in a probability sample, and to the extent that zip codes are different along dimensions that are not included in the distributions that are compared, this may lead to drawing very nonrepresentative samples. Hence, we included this more as a benchmark (lower bound) than as a realistic possibility.

Figure 1 shows the results from a limited simulation that we performed to get some insight into whether the method is likely to work. It shows the evolution of three criteria as more samples are drawn:

- The ratio of the largest to the smallest weight (based on cell-based poststratification weights);
- The design effect (loss of effective sample size compared to simple random sampling);
- The Pearson chi-square D (to measure the difference between sample and population distribution).

The graphs on the left assume $n = 10$ addresses will be drawn per zip code, the ones on the right assume $n = 50$. From these results, we concluded that the algorithm with $a = 1.5$ and $R = 100$ is likely to perform well. As discussed in the main text, when implementing this, we started with $n = 40$, using $n = 10$ and without replacement sampling in LA County, and later switched to proportional-to-size samples per zip code for the nationally representative and California batches.

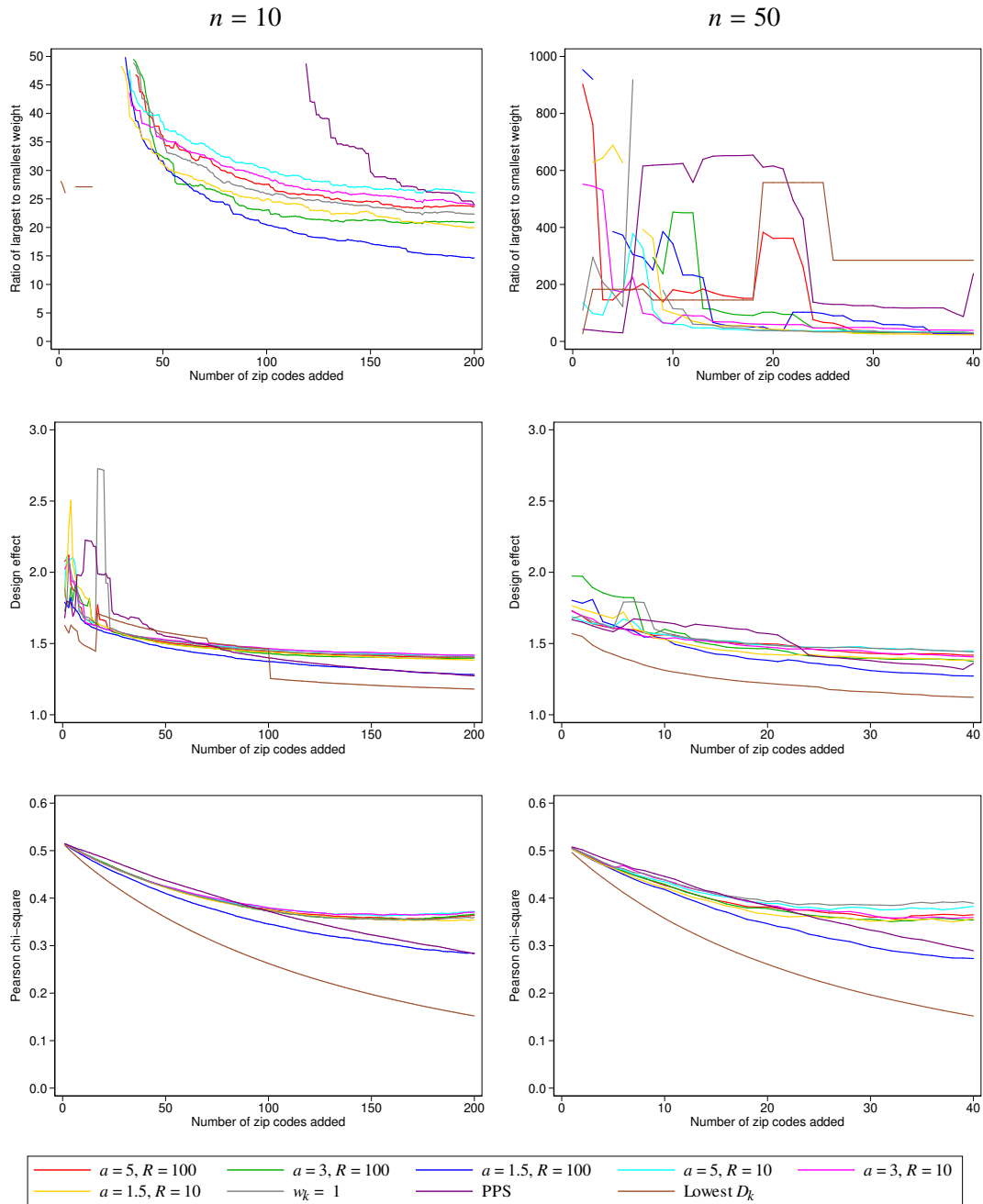


Figure 1: Simulation results for different variants of the sampling algorithm (n is the sample size per zip code)

B Effective sample size metric for the Native American batches

For the nationally or regionally representative batches starting with batch 5, we have used an adaptive sampling method with the Pearson chi-square metric as the measure of fit, as discussed in the main text. For the Native American batches, which were sampled without taking the earlier nationally representative batch into account, we used an effective sample size metric instead. This appendix describes this metric. It is a slightly edited version of a note written in January 2014.

B.1 Deriving an ESS metric

A drawback of the Pearson chi-square metric (1) is that it is undefined if $g_j = 0$ for some j . In many cases, this will not be a problem, because g typically is a known population distribution and values that do not occur in the population will not occur in the sample either, and these cells will simply be excluded. However, if the aim is to sample from a subpopulation that can not be identified a priori, the sample may contain individuals that are not in the subpopulation of interest and thus the desired fraction of individuals with this value is zero.

Our motivation is our intention to draw a sample of American Indians or Alaska Natives (AIANs).⁵ Less than 2% of the U.S. population consists of AIANs, so if we draw a simple random sample from the U.S. population, we obtain a sample with 98% of the individuals having an undesired characteristic. However, we are able to select zip codes from which to draw individuals and we know the racial distribution in these zip codes from the American Community Survey (ACS) 2007–2011 5-year file. Thus, we can give zip codes with higher fractions of AIANs a higher probability of being in the sample, thereby increasing the hit rate. We would also like to obtain a sample that is well-balanced in terms of region (Census division), urbanicity (urban, rural, mixed), and whether the zip code is part of an American Indian area (AI area) or not. The latter are mostly reservations. We stratify by the AI area indicator, but complete stratification by the other characteristics is not practical, because the number of strata ($9 \times 3 \times 2 = 54$, including AI area but excluding individual AIAN status) would become too large compared to the intended sample size (300).

Thus, in each of the two strata, we have 54 cells (including individual AIAN status), but the target distribution has zeros in half of these cells (non-AIAN race). We could modify the Pearson metric (1), for example, by replacing $g_j = 0$ by $g_j = 1$. However, this still makes the Pearson statistic a function of the square of the fraction in the non-AIAN cells, which is unattractive. Non-AIANs will be dropped from the sample, so it seems most realistic to assume that each non-AIAN contributes a certain fixed cost c_0 without having any benefit, and thus the metric should be a function of the total (sum) of the number of non-AIANs, and not of a sum of squares over irrelevant cells. Moreover, a discrepancy in the distribution among AIANs has very different consequences than a discrepancy in the distribution among non-AIANs, which is not reflected in

⁵At the time of drawing the zip codes for the Native American batches, we defined Native American as AIAN. Later, we adopted a broader definition that also includes Native Hawaiian and other Pacific Islanders.

the Pearson metric.

Therefore, here we develop a different metric, which is more suitable for sampling the NA batches, but can be used more generally as well. The metric is based on the concept of the *effective sample size* (ESS). First, suppose we have drawn a simple random sample of size n from a population of interest and we are interested in the mean μ of a variable x . The estimator is the sample mean \bar{x} . If x is independently and identically distributed (i.i.d.) with variance σ^2 , the variance of the sample mean is $V_{\text{SRS}} = \sigma^2/n$. Now suppose that we are in a different situation where, for example because of sampling with unequal probability or clustering, The variance of \bar{x} is V . Then the effective sample size n_{eff} is the sample size that would be required to obtain the variance V from a simple random sample, that is, it is such that $V = \sigma^2/n_{\text{eff}}$. Hence, $n_{\text{eff}} = \sigma^2/V$. For defining our desirability metric, we proceed by assuming x is still i.i.d. but the sample is drawn with unequal probabilities, resulting in the distribution of y (the demographics and other characteristics of the zip codes) being potentially different from the population distribution. To correct this, sampling weights, denoted by w_i for the i -th observation, are constructed. The (weighted) sample mean is

$$\bar{x} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} = \sum_{i=1}^n \tilde{w}_i x_i,$$

where $\tilde{w}_i = w_i/W$ and $W = \sum_{i=1}^n w_i$ is the sum of the weights. Hence, the variance of \bar{x} is

$$V = \sum_{i=1}^n \tilde{w}_i^2 \sigma^2 = \frac{\frac{1}{n} \sum_{i=1}^n w_i^2}{\left(\frac{1}{n} \sum_{i=1}^n w_i\right)^2} \times \sigma^2/n.$$

It follows that the ESS is

$$n_{\text{eff}} = n \frac{\left(\frac{1}{n} \sum_{i=1}^n w_i\right)^2}{\frac{1}{n} \sum_{i=1}^n w_i^2}.$$

For these computations, we assume that the weights are constant within each of the 54 cells of region \times urbanicity \times AI area, which form the categories j of the distributions f_j , g_j , and h_{kj} . Thus, $w_j = g_j/f_j$, where as before g_j is the population fraction in category j of y and f_j is the corresponding sample fraction, and the weight w_j is applied to each observation in category j . There are $n_j = n \cdot f_j$ such observations. Note that

$$\sum_{i=1}^n w_i = \sum_{j=1}^J n_j w_j = n \sum_{j=1}^J g_j = n$$

(where $J = 54$ in our case). Consequently, the ESS is

$$n_{\text{eff}} = \frac{n}{\frac{1}{n} \sum_{i=1}^n w_i^2} = \frac{n}{\sum_{j=1}^J g_j^2 / f_j}. \quad (2)$$

As in the setup in the section 2, suppose we have an existing dataset with effective sample size $n_{\text{eff},0}$ and we are considering which zip code to sample. If zip code k is sampled, n_{1k} additional observations are included, resulting in a new ESS of $n_{\text{eff},1k}$. We can then define the benefits of drawing this sample as the increase in ESS:

$$b_k = n_{\text{eff},1k} - n_{\text{eff},0}. \quad (3)$$

The costs of this amount to c_k , which typically would be a linear function of n_{1k} , but the cost function in our case is slightly more complex (see below). A natural metric for the desirability of zip code k is then the benefit-cost ratio $D_k = b_k/c_k$, that is, increase in effective sample size per dollar.

B.1.1 Dealing with zeros

As argued above, one of the main drawbacks of the Pearson criterion, and the primary reason for developing the ESS criterion, was that it could not deal with zeros in the target distribution g . Unfortunately, the ESS criterion suffers from an analogous problem: the ESS (2) becomes undefined (or, more precisely, zero) if there is a cell j with $f_j = 0$. If the corresponding target is also zero ($g_j = 0$), we can replace g_j^2/f_j by zero, but if this is not the case, this does not give a good impression of the desirability of the sample. If region is one of the characteristics in y , this means that the criterion is uninformative until all regions are filled by at least one zip code. This is an inconvenient problem, especially if the number of “regions” is large (as it is in our case, being defined as Census division \times urbanicity, 27 cells within both strata) and the sample size small, in particular the number of zip codes. One solution would be to use raking weights in such situations, which does not require that all cells be filled, although still all marginals must be nonzero. While we do use raking in the computation of our poststratification weights, this is computationally impractical as part of the sampling algorithm. A practical, though admittedly ad-hoc, approach is to prefill all cells with positive numbers, for example, one respondent in each cell, or a probability mass of g_j in cell j with an initial sample size of $n_0 = 1$. Once the cells are filled with actual respondents, this is only a slight distortion that does not materially affect the selection of new zip codes. However, in the initial stages, the selection of zip codes may be sensitive to exactly how the cells are prefilled. We implemented the following algorithm (within both the AI area and non-AI area strata separately):

1. Set $n_0 = 1$ and $f_j = g_j$ for all j .
2. Draw 10 zip codes at random (and update f here and in the following).
3. Draw 10 zip codes using the adaptive algorithm with the ESS criterion, only taking the urbanicity distribution into account ($J = 3$).
4. Draw 10 zip codes using the adaptive algorithm with the ESS criterion, only taking the region distribution into account ($J = 9$).

5. Draw the remaining zip codes using the adaptive algorithm with the ESS criterion, taking the joint distribution of region and urbanicity into account ($J = 27$).

Note that a similar ad-hoc adaptation to the Pearson criterion would remain sensitive to the specific choice even in later stages of the sampling, because the denominators g_j are the fractions from the target distribution, which stay the same throughout the whole process, whereas in this algorithm, the culprit f_j is updated in each step, and the sensitivity to the initialization quickly disappears.

B.2 Cost function

In the Native American batches, the cost function consists of a number of components: (a) some fixed costs c_0 that do not depend on the sampling and thus can be ignored; (b) a cost c_1 per person who is invited to participate in the initial short screening survey (e.g., postage; small monetary gift); (c) a cost c_{2P} per person who returns the paper survey (e.g., data entry; payment for answering the survey); (d) a cost c_{2I} per person who answers the initial survey online (e.g., data cleaning, but this can be close to zero; payment for answering the survey); (e) a cost c_3 per person who is eligible for entering the panel (cost of trying to get in contact); (f) a cost c_{4I} per person who enters the panel and already has internet access (e.g., setup of accounts, personal details, generating de-identified user IDs); (g) a cost c_{4N} per person who enters the panel and does not have internet access already (ditto, plus cost of a tablet and broadband subscription). More costs, such as maintaining a support line, are involved, but this gives a good impression. Suppose m_k individuals are initially invited from zip code k . From external information (ACS), we know that a fraction e_k of these will be eligible (AIANs). A fraction r_P of initial invitees returns the paper questionnaire and a fraction r_I answers the questionnaire online; a fraction q_I of the eligible individuals enters and has internet and a fraction q_N enters and does not have internet. In the absence of any prior information, we assume that these fractions except e_k do not vary across zip codes or across demographics y . Then the costs are

$$\begin{aligned} c_k &= m_k \{c_1 + r_P c_{2P} + r_I c_{2I} + (r_P + r_I) e_k c_3 + (r_P + r_I) e_k (q_I c_{4I} + q_N c_{4N})\} \\ &= m_k (A + B e_k), \end{aligned} \tag{4}$$

where the constants $A = c_1 + r_P c_{2P} + r_I c_{2I}$ and $B = (r_P + r_I)(c_3 + q_I c_{4I} + q_N c_{4N})$ do not depend on k . It is likely that q_I would be higher for those who entered the initial survey online and lower for those who returned it on paper, which would change the meaning of A and B but not the form of the expression. In our case, we expect differences between AI areas and non-AI areas. Based on informed guesses, we used $A = \$23.492$, $B = \$38.52$ for non-AI areas, and $B = \$133.56$ for AI areas.

C Sampling implementation details by batch

In the following, “NR” means batches that are intended to be nationally representative, “NA” means Native American batches, “LAC” means Los Angeles County batches, and “CA” means

California batches. “ACS” is the American Community Survey of the United States Census Bureau (<https://www.census.gov/programs-surveys/acs/>).

C.1 Batch 1 (ASDE NR)

The first batch consisted of a random sample of about 10,000 individuals from the database of the commercial vendor ASDE. ASDE claimed that they had about 95% of the U.S. population over 18 in their database, so we consider this a simple random sample from the U.S. population over 18.

C.2 Batches 2 and 3 (NA1 & NA2)

These two batches aimed at establishing a Native American panel. This was also obtained from ASDE, but using a different method. For each of these batches, we submitted 421 zip codes to ASDE, who then subsetted their database to only those zip codes and then drew a random sample of size 10,000 from the resulting database, of which we used a random 5,000. We drew the zip codes from the 2007–2011 5-year file from the ACS, excluding zip codes with less than 5% Native Americans and zip codes with fewer than 100 residents. The 421 zip codes were drawn from two strata: 50 were drawn from AIAN areas (which are mostly reservations) and the remaining 371 from non-AIAN areas. These two samples were drawn independently of each other. The population of zip codes we drew from contained 544 zip codes from AIAN areas and 1052 zip codes from non-AIAN areas. Hence, $100/544 = 18\%$ of the AIAN area zip codes satisfying the restrictions were drawn, and $2 \times 371/1052 = 71\%$ of the non-AIAN area zip codes satisfying the restrictions were drawn.

A consequence of this way of sampling individuals is that (in expectation) the number of individuals drawn from a zip code is proportional to the population of the zip code. Hence, the effect of drawing zip code k is drawing $n_{1k} \propto N_k$ individuals from zip code k instead of drawing a fixed number n_1 as for batches 5–12.

We used an *effective sample size* (ESS) based criterion to define desirability of a zip code. See Appendix B for a discussion and derivation. Because ESS increases with sample size, and sample size was proportional to zip code population size, if this would be directly used as the criterion, we would primarily draw large zip codes (more precisely, given that only Native Americans would be eligible, zip codes with large absolute populations of Native Americans). But because both total sample size and the number of zip codes was determined beforehand, this would not be a useful criterion by itself. Therefore, we used the ratio of the increase in ESS due to adding zip code k to an estimate of the cost of adding zip code k as our desirability criterion. That is, we used

$$D_k = \frac{\text{ESS}(\hat{f}_k, g) - \text{ESS}(f, g)}{c_k} \quad (5)$$

where f was the predicted distribution before adding zip code k and \hat{f}_k the predicted distribution after adding zip code k . The cost function we used was

$$c_k = N_k(A + Be_k) \quad (6)$$

where $A = 23.492$, $B = 38.52$ for non-AIAN areas, $B = 133.56$ for AIAN areas, N_k is the total population age 18 and over in the zip code, and e_k is the fraction Native Americans in the zip code.

The characteristics we used for defining the distributions f , g , and h_k were Census Division (9 categories) and urbanicity (3 categories: urban, rural, mixed) of the zip code. These were completely cross-classified, resulting in a one-dimensional distribution with $J = 27$ categories. Because these characteristics are defined at the zip code level, for any given zip code, $h_{kj} = 1$ for exactly one j and $h_{kj} = 0$ for all other j .

The ESS criterion is undefined for $f_j = 0$. Therefore, we used an initialization in which we started with one hypothetical respondent and $f_j = g_j$, then drew a small number zip codes, adding more information in steps, until the full distribution was used. See Appendix B for details.

The algorithm was a variant of the algorithm described in the main text, with D_k as in (5). This was then transformed into sampling probabilities of the zip codes along the lines of section 2.4:

1. Compute the quantile Q_k of D_k . That is, the zip code with the smallest value of D_k gets $Q_k = 1/K$, the next gets $Q_k = 2/K$, and so forth, and the zip code with highest D_k gets $Q_k = 1$.
2. Because higher values of D_k are more desirable, as opposed to section 2.4, compute the transformed quantiles $q_k = K + 1 - Q_k$. Note that the intention was to compute $q_k = 1/K + 1 - Q_k$, i.e., $q_k = 1/K, \dots, K/K$ with $1/K$ being most desirable. Because of this error, we actually used $q_k = K, K + 1/K, \dots, K + 1 - 1/K$ with K being most desirable. We discuss the consequence of this below.
3. Compute the utility function

$$U_k = -\frac{q_k^{1-a}}{1-a},$$

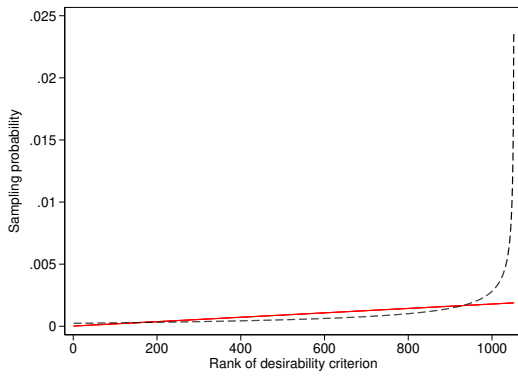
where $a = 1.5$ is the tuning parameter.

4. Compute $w_k = b + U_k$, where $b = (U_{\max} - RU_{\min})/(R - 1)$ was chosen such that the ratio between the highest and lowest value of w_k is equal to $R = 100$.
5. Compute $p_k = w_k / \sum_{\ell=1}^K w_{\ell}$. The probability of drawing zip code k is p_k .
6. Draw a zip code according to the probabilities p_k .
7. Update the distribution f and panel size n .
8. Drop the drawn zip code from the population to draw from and repeat.

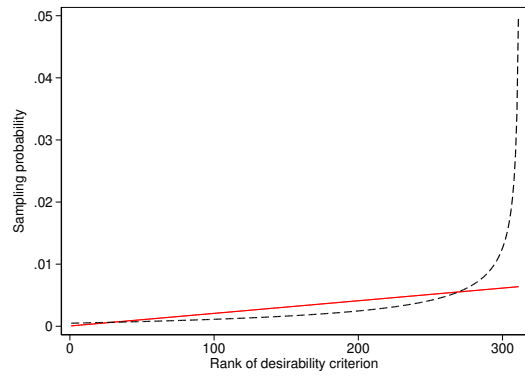
The consequence of the error with computing the quantiles q_k is that the utility function had less curvature than intended. As illustrated in Figure 2, this implied that the most desirable zip codes were drawn with much less probability than intended, whereas zip codes with average desirability had higher probabilities. Overall, the sampling is closer to a simple random sampling than intended.

Figure 2: Intended and actual probabilities of drawing zip codes for various batches

Native American batches (2 & 3)

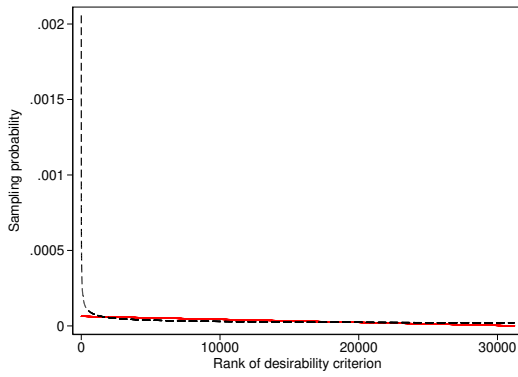


First draw ($K = 1052$)



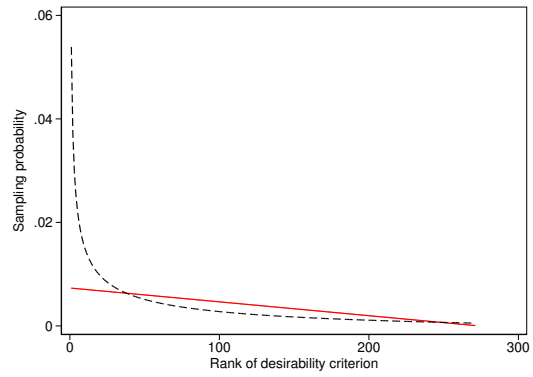
Last draw ($K = 311$)

Batch 5 (MSG1 NR)

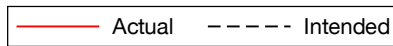


First draw ($K = 31,175$)

Batch 13 (LA County 2)



$K = 271$



Furthermore, due to a miscommunication, the definition of “Native American” used in the sampling was narrower than intended. In the drawing of zip codes, the definition used was “American Indian or Alaska Native” (AIAN; alone or in combination), whereas for the purpose of defining eligibility of this subpanel, Native Hawaiians and Pacific Islanders (NHPI; alone or in combination) are also included. All respondents to the initial survey who were AIAN or NHPI were eligible to enter the panel and respondents who were not AIAN or NHPI were ineligible. Another mishap allowed some individuals of other races to enter the panel. These were retained, but get a weight of zero. Household members of Native Americans are also eligible for the panel, as usual, even if they are not Native Americans themselves.

At the time of drawing the zip codes for batch 3, the number of panel members in batch 2 was very low, and therefore we made no attempts to take the results from drawing batch 2 into account. Hence, we simply reran essentially the same program, except that we excluded the zip codes from batch 2 and generated a new random seed.

C.3 Batch 4 (LAC1)

This batch aimed at sampling families with young children in LA County. This was drawn as a simple random sample of about 2000 individuals from a list of names and addresses provided by the State of California of individuals having given birth in LA County in 2009–2012 within a set of zip codes around restaurants participating in a healthy menu options project. However, we allowed anyone at the address (specifically, new occupants) to answer the survey and enroll in the panel. Because of the highly specific nature of the subsample, we do not provide weights for members recruited within this batch.

C.4 Batch 5 (MSG1 NR)

This batch was the first one in which we used the adaptive sampling algorithm as described in section 2, comparing the then-current distribution in the UAS with external targets, and the first one for which we used MSG as our provider. We used the March 2015 CPS base data to estimate the target distribution g and the 2009–2013 5-year ACS file at the zip code tabulation area (zcta) level to estimate the distributions of the zip codes h_k . For the distribution f in the UAS, we used only panel members from the nationally representative batch 1. The characteristics we used for comparing distributions were sex \times age \times race/ethnicity, sex \times age \times marital status, and sex \times age \times education (see section 2 for the categories used). For the anticipated response rates (appendix D), we only used data from batch 1, with the distribution of the invites (γ_{mj}) equal to the national distribution (g_{mj}) and overall response rate of $\rho = 0.20$. The latter was an informed guess rather than an empirical estimate. The algorithm used for drawing this batch had the same bug as for the Native American batches, so that the probabilities of sampling the most desirable zip codes were much lower than intended, as illustrated in the bottom-left graph of Figure 2.

We excluded zip codes with fewer than 100 residents age 18+. We drew 50 zip codes and 40 addresses per zip code, for a total sample of 2000 addresses.

C.5 Batches 6–12 (MSG2–MSG8 NR)

Batches 6–12 were drawn in the same way as batch 5, with some small differences. In all these batches, zip codes drawn in previous batches (including the Native American batches) were excluded from consideration, so that we drew zip codes without replacement within and across batches.

In batch 6, we used an updated (estimated) overall response rate from batch 1 of $\rho = 1913/10,000$ as the anticipated total response rate; 100 zip codes were drawn for a total sample of 4000 addresses. The actual number of invites was lower (about 3700) because the drawn zip codes (from the ACS) included addresses in Puerto Rico and other zip codes such as college campuses for which the Delivery Sequence File did not contain any residential addresses.

In batch 7, an anticipated overall response rate of $2149/12,000$ was used, based on the number of panel members and invitees from batches 1 and 5; 63 zip codes were drawn for a total sample of slightly over 2500 addresses, although again the actual number of invites was a little lower because of a drawn zip code in Puerto Rico.

From batch 8 onward, we used the estimated distribution of the invites from previous nationally representative batches to estimate the γ_{mj} . This was based on the distributions (from the ACS) of the zip codes that were actually drawn in previous batches, in addition to the national distribution for batch 1. We also used the cumulative number of invites and the (then-)current panel size to estimate the overall response rate. Batch 8 again sampled 100 zip codes for a sample of 4000 addresses.

Batch 9 was drawn in the same way as batch 8, except that 110 zip codes were drawn instead of 100. Because in previous batches we had had some problems with zip codes that did not have any addresses in MSG’s files, the 10 additional zip codes were to be used as backups in case some of the first 100 could not be used. Thus, the target number of addresses was still 4000. Batches 10–12 were drawn in the same way as batch 9.

C.6 Batch 13 (LAC2)

This batch was drawn in Los Angeles County using the adaptive algorithm as in batches 5–12, but with several differences. Because the number of zip codes (more precisely: ZCTAs) from LA County in the ACS file is not very large (roughly 300), we decided to draw zip codes with replacement, and to draw fewer addresses (10) per zip code per iteration. However, zip codes drawn in previous batches were still excluded.

The algorithm used for drawing this batch still had the same bug as for the previous batches, so that the probabilities of sampling the most desirable zip codes were much lower than intended, as illustrated in the bottom-right graph of Figure 2. From this batch onward, we added the (joint) distribution of zip code population size (quintiles derived from the ACS) and urbanicity (rural, mixed, urban; based on Census Bureau data) to the target distribution g , panel distribution f , and zip code distributions h_k . This was done because we had noticed an overrepresentation of small and rural zip codes in previous batches. Related to this, from this batch onward, the target

distribution g was obtained as an aggregate over zip codes from the ACS file (restricted to LA County for this batch).

The target number of addresses was 2000, so $2000/10 = 200$ iterations of the algorithm were performed; 135 zip codes were drawn, with several of them multiple times (in different iterations), resulting in 10–50 addresses per zip code.

C.7 Batches 14, 18, 19 (LAC3–LAC5)

These batches were also drawn in LA County. The algorithm used was similar to the algorithm for batch 13, but we made a few changes.

First, we discovered and corrected the bug that had given us zip code sampling probabilities that were less curved with desirability than intended (Figure 2). Second, zip codes that had been drawn before (including in non-LA County specific batches) were not excluded, so that sampling was with replacement both within and across batches. Third, whereas previous batches sampled from zip codes with at least 100 *residents* according to the ACS, we now sampled from zip codes with at least 100 *households*. This excludes college campuses and institutions for which MSG had no addresses anyway. Fourth, the ACS version used for this was updated to the 2011–2015 (2013–2017 for batches 18 and 19) 5-year ACS file at the zip code tabulation area (zcta) level. Fifth, the algorithm was adapted again to include a “proportional to size” factor that increases the probability that large zip codes will be drawn. That is, the probability of drawing zip code k was $p_k = N_k w_k / \sum_{\ell=1}^K N_{\ell} w_{\ell}$, where N_k is the zip code population size (residents age 18+ in batch 14; households in batches 18 and 19). Combined with the sampling with replacement, this has a tendency to also increase sample size within large zip codes relative to small ones.

In batch 14, the target number of respondents was 1000, so $1000/10 = 100$ iterations were done. In batches 18 and 19, the target number of respondents was 5000, so $5000/10 = 500$ iterations were done in each of these.

C.8 Batches 15 & 16 (CA1, CA2)

These batches were drawn in California. The algorithm used was basically the one used for the nationally representative batches (5–12), so sampling was without replacement. However, the sampling included the corrections to the algorithm introduced in the LA County batches. The target distribution and zip code distributions were again obtained from the 2011–2015 5-year ACS file, and sampling was restricted to zip codes with at least 100 households. Instead of drawing zip codes with a proportional-to-size factor, the number of addresses drawn was proportional to size,⁶ with a target average of 40. That is, in the algorithm, the addition of zip code k was assumed to lead to an addition of $\hat{n}_k = 40(N_k/\bar{N})$, where N_k is the number of households in the zip code and \bar{N} is the average number of households in all eligible zip codes (i.e., zip codes in California with at least 100 households). The target number of addresses was 5000 (10,000 for batch 16), and the

⁶This was done after discussions with Jon Krosnick about this issue.

algorithm continued until this amount was reached. Because \hat{n}_k is typically noninteger, the final number of addresses per zip code was a rounded version of this: rounded up to the nearest integer with probability $\hat{n}_k - \lfloor \hat{n}_k \rfloor$ and rounded down to the nearest integer with probability $1 - (\hat{n}_k - \lfloor \hat{n}_k \rfloor)$. For example, if $\hat{n}_k = 13.7$, the number of addresses is 14 with 70% chance and 13 with 30% chance.

C.9 Batch 17 (MSG9 NR)

This batch used the same algorithm as batches 15 and 16, applied to the U.S. as a whole instead of just California. Zip code distributions were obtained from the 2012–2016 5-year ACS file, and in addition to the earlier innovations, this excluded the Puerto Rico zip codes, preventing the issues we encountered in batches 5–12. The target number of households was 5000, with the number of addresses per zip code proportional to the number of households and a target average of 40 addresses per zip.

D Anticipated response rates

As described in section 2, the adaptive sampling algorithm takes anticipated response rates r_{mj} into account in determining the desirability of each zip code. Let the distribution of sampled individuals be given by the proportions γ_{mj} , and the total number of sampled addresses be N . Then, in the absence of nonresponse, the expected number of individuals with category j of characteristic m is $N\gamma_{mj}$. If the total size of the panel is n , the number of individuals with category j of characteristic m in the panel is $n f_{mj}$, and hence an estimate of r_{mj} is $r_{mj} = n f_{mj} / (N \gamma_{mj})$.

However, given the large number of categories and the relatively small sample sizes, especially for the earlier batches, these estimates may be very low, or even zero, with sparse cells. Moreover, regardless of whether they are low, these estimates may be imprecise. We do not want to rely too strongly on such imprecise or extreme estimates, and therefore we trim the estimates. Specifically, let the overall response rate be $\rho = n/N$ and compute the initial estimates $r_{1mj} = \rho f_{mj} / \gamma_{mj}$. Define the bounds $Q_L = 1/3$ and $Q_H = 3$. Then we winsorize the estimates according to

$$r_{2mj} = \begin{cases} \rho Q_L & \text{if } f_{mj} / \gamma_{mj} < Q_L; \\ r_{1mj} & \text{if } Q_L \leq f_{mj} / \gamma_{mj} \leq Q_H; \\ \rho Q_H & \text{if } f_{mj} / \gamma_{mj} > Q_H. \end{cases}$$

After this winsorizing, the overall estimated response rate may not be ρ anymore, so we rescale them such that the overall response rate is correct again. However, simple rescaling may lead to some response rates being greater than 1. Therefore we compress them using the estimate

$$r_{3mj} = \alpha + \beta r_{2mj}$$

such that

$$\sum_{m=1}^M \sum_{j=1}^{J_m} r_{3mj} \gamma_{mj} = \rho,$$

so $\beta = (\rho - \alpha)/S$, where

$$S = \sum_{m=1}^M \sum_{j=1}^{J_m} r_{2mj} \gamma_{mj}.$$

If $\alpha = 0$ gives all estimates $r_{3mj} \leq 1$, we choose $\alpha = 0$ and $\beta = \rho/S$. Otherwise, we impose the restriction that $\max_{m,j} r_{3mj} = 1$, so $\alpha + \beta r_{\max} = 1$, where $r_{\max} = \max_{m,j} r_{2mj}$. This leads to the solution $\alpha = (\rho r_{\max} - S)/(r_{\max} - S)$ and $\beta = (1 - \rho)/(r_{\max} - S)$.

The proportions γ_{mj} are estimated as follows. Let \mathcal{K}_b be the set of zip codes sampled in batch b , and let N_{kb} be the number of addresses sampled in zip code $k \in \mathcal{K}_b$ in batch b . Furthermore, let $h_{kmj(b)}$ be the fraction in zip code k that has category j of characteristic m , as derived from the ACS file at the time batch b was sampled. Because the ACS file is updated annually, this varies with b . When sampling batch b' , we estimate γ_{mj} using data from batches in the set \mathcal{B} . This is a subset of $\{1, \dots, b' - 1\}$: some very recent batches may be excluded because the recruitment has not progressed enough to estimate response rates, and some batches from different sampling frames may be excluded. (The corresponding panel members in the UAS are also excluded from n and f_{mj} , and the total number of sampled addresses is also evaluated for this subset.) Now,

$$\gamma_{mj} = \frac{\sum_{b \in \mathcal{B}} \sum_{k \in \mathcal{K}_b} N_{kb} h_{kmj(b)}}{\sum_{b \in \mathcal{B}} \sum_{k \in \mathcal{K}_b} N_{kb}} = \frac{1}{N_{b'}} \sum_{b \in \mathcal{B}} \sum_{k \in \mathcal{K}_b} N_{kb} h_{kmj(b)},$$

where $N_{b'}$ is the total number of sampled addresses in the batches from the set \mathcal{B} . Batch 1 was a simple random sample from the population, and thus we treat this batch as a sample from a single “zip code” with distribution $h_{kmj(1)} = g_{mj(1)}$, that is, the distribution for the U.S. as a whole.

E Estimating frame-specific probabilities

In this appendix, we describe how we estimate the frame-specific probability $\pi_{hk|s}$ of household h in zip code k being invited to participate in the UAS, as part of computing the base weights.

E.1 Frame 1: Nationally representative list ($s = 1$; batch 1)

Batch 1 was a simple random sample from the database of the vendor ASDE, who stated that they included over 95% of the U.S. population. Hence, we treat this as a simple random sample from the U.S. population (of 18 and over). As mentioned above, for this frame, we consider the whole U.S. as a single zip code, with $\pi_{k|s=1} = 1$ and $\pi_{h|k,s=1} = n/N$, where $n = 10,000$ is the number of households invited and $N = 117,715,016$ is the number of households in the U.S. population. Hence, $\pi_{hk|s=1} = \pi_{h|k,s=1} \pi_{k|s=1} = 10,000/117,715,016 = 0.00008495$.

E.2 Frame 2: Native American zips ($s = 2$; batches 2 & 3)

As discussed above, the Native American batches were drawn in two stages, with stage 1 being a stratified sample of zip codes and stage 2 being (approximately) a simple random sample from all the households in the set of zip codes drawn in stage 1.

Let d_k be a dummy variable indicating whether zip code k was drawn in the sample, and let $\mathcal{D} = \{d_1, \dots, d_K\}$ indicate the drawn sample, where K is the total number of zip codes from which the sample is drawn ($K = 1596$ for the Native American batches). Then the probability that household h in zip code k is invited conditional on the set of zip codes indicated by the set \mathcal{D} , is

$$\pi_{hk|\mathcal{D},s=2} = \frac{d_k n}{N_{\mathcal{D}}}, \quad (7)$$

where

$$N_{\mathcal{D}} = \sum_{\ell=1}^K d_{\ell} N_{\ell}$$

is the total number of households in the sampled zip codes, N_{ℓ} is the number of households in the population of zip code ℓ , and n is the number of invited households (5,000 for batch 2 and 10,000 for batches 2 and 3 jointly).

Thus, the probability that household h in zip code k is invited is

$$\pi_{hk|s=2} = \mathbb{E} \pi_{hk|\mathcal{D},s=2} = \sum_{q=1}^Q \pi_{\mathcal{D}_q} \pi_{hk|\mathcal{D}=\mathcal{D}_q,s=2}, \quad (8)$$

where \mathcal{D}_q is a particular feasible sample of zip codes, $q = 1, \dots, Q$, and Q is the number of feasible samples of zip codes. For batches 2 and 3 jointly,

$$Q = \frac{544!}{100! 444!} \frac{1052!}{742! 310!} = 6.3 \times 10^{386}.$$

This number is larger than the largest double precision number that Stata can represent and thus—even aside from the difficulty of estimating the probabilities $\pi_{\mathcal{D}_q}$ —it is computationally infeasible to compute the sum (8). However, it can be accurately estimated in a much simpler way. Because

$$m_N = \frac{1}{K_{\mathcal{D}}} N_{\mathcal{D}} = \frac{1}{K_{\mathcal{D}}} \sum_{\ell=1}^K d_{\ell} N_{\ell} = \bar{N}$$

is a sample mean, with $K_{\mathcal{D}} = \sum_{\ell=1}^K d_{\ell}$ the (nonrandom) number of sampled zip codes, m_N is a consistent estimator of the population mean $\mu_N = \mathbb{E}(N_{\ell})$. Therefore, (7) has the same asymptotic distribution as $d_k n / (K_{\mathcal{D}} \mu_N)$ and $\pi_{hk|s=2} \approx \mathbb{E}(d_k) n / (K_{\mathcal{D}} \mu_N) = \pi_{k|s=2} n / (K_{\mathcal{D}} \mu_N)$. The estimate

of $K_{\mathcal{D}}\mu_N$ is of course $K_{\mathcal{D}}m_N = N_{\mathcal{D}}$,⁷ where we use the 2012–2016 5-year average ACS file to obtain the most accurate estimates of N_{ℓ} for the sampling period (2014), and n is a known quantity (5,000 or 10,000). Consequently, the problem is reduced to estimating the probability $\pi_{k|s=2}$ that zip code k is drawn.

In drawing the zip codes, we used a variant of our adaptive algorithm, which used a combination of Census division, urbanicity, zip code population size, and fraction AIANs in the population to determine sampling probabilities, the latter two computed from the 2007–2011 5-year average ACS file. Zip codes were drawn independently in two strata: American Indian areas and other zip codes. Hence, we estimate $\pi_{k|s=2}$ by estimating a logit model for being sampled, separately in these two strata, using dummies for Census division and urbanicity, plus the logarithms of zip code population size and fraction AIANs.

E.3 Frame 3: Los Angeles County young mothers ($s = 3$; batch 4)

As mentioned above, we are unable to compute sampling probabilities for this frame, and thus exclude this batch from the weighting.

E.4 Frame 4: Nationally representative 2-stage, fixed ($s = 4$; batches 5–12)

Batches 5–12 used our adaptive algorithm to draw zip codes in the first stage, and sampled a fixed number (40) of addresses per zip code in the second stage. Consequently, the probability of a household being invited in this frame is $\pi_{hk|s=4} = \pi_{h|k,s=4}\pi_{k|s=4}$ with $\pi_{h|k,s=4} = n_k/N_k = 40/N_k$, $n_k = 40$ being the number of households being invited per zip code and N_k being the number of addresses (households) in the population in zip code k .

An overview of the adaptive algorithm for drawing zip codes in this frame is given in section 2. For each new zip code drawn in a batch, an estimate of the distribution of certain individual characteristics in the UAS up to the previous drawn zip code is compared to the corresponding distribution in the ACS. Zip codes that, when drawn, would move the UAS distribution more in the direction of the ACS distribution have a higher probability of being drawn. The characteristics used are demographics (sex, age, race/ethnicity, marital status, education, and interactions thereof). In later batches, additional zip code level characteristics were added: population size, urbanicity, and region (Los Angeles County, other California, non-California). For a while, a programming error caused less desirable zip codes to be drawn with a higher probability than more desirable zip codes. Furthermore, not all zip codes were included in the database of our vendor for these batches (MSG). Specifically, zip codes from Puerto Rico are not in their database and zip codes for institutions like college campuses, prisons, and hospitals are not in their database. Realizing this, we excluded Puerto Rico in later batches, but we were unable to pre-exclude the other zip

⁷Note that we use this same estimate for households from all feasible zip codes, whether or not they were actually drawn in the sample, because μ_N does not depend on the actual sample.

codes. Therefore, we often drew a few backup zip codes as well, to be used as replacements of the rejected zip codes.

For constructing the base weights, we treat all these variations as random and estimate $\pi_{k|s=4}$ as a combined probability from all these batches, where we code both the rejected zip codes (except Puerto Rico, which we exclude from all computations) and the backup zip codes as being drawn. We estimate the probability using a logit model, with demographic distributions as regressors (fraction male, fractions of racial/ethnic categories, fractions in age categories, fraction not married, fraction in education categories), in addition to dummies for Census Region (augmented with Los Angeles County and other California), population quintile, and urbanicity.

E.5 Frame 5: Los Angeles County 2-stage ($s = 5$; batches 13, 14, 18, & 19)

Batches 13 and 14 used a modified version of the adaptive sampling algorithm. Rather than drawing zip codes without replacement, these drew zip codes with replacement, but with a smaller number of addresses per zip code draw, 10 instead of 40. (The addresses were still drawn without replacement.) Batch 13 excluded zip codes drawn before as part of the nationally representative batches, but batch 14 did not do this. Batch 14 also modified selection probabilities proportional to population size, so that larger zip codes had a proportionally higher probability of being drawn. We ignore these differences for computing the weights and include all LA County zip codes in the population from which zip codes were drawn.

The sampling method for batches 13 and 14 can then be described as follows:

1. Set $b =$ batch number, renumbered such that $b = 1, \dots, B$ ($B = 2$) for the batches of interest.
2. Choose the target number n_b of households to sample and compute the number of iterations $K_b = n_b/10$.
3. In iteration t , compute the probabilities p_{kbt} that zip code k is drawn as a function of the UAS panel composition (in LA County) immediately before drawing batch b and the projected UAS sample composition after iteration $t - 1$.
4. Using the probabilities p_{kbt} , draw one zip code and define the indicator variable $d_{kbt} = 1$ if zip code k was drawn in iteration t of batch b and $d_{kbt} = 0$ otherwise.
5. Do steps 3 and 4 K_b times. We drew zip codes with replacement, so every zip code was eligible in every iteration, regardless of whether it had been drawn in a previous iteration.
6. After completing the K_b iterations, compute $d_{kb} = \sum_{t=1}^{K_b} d_{kbt}$, the number of times zip code k was drawn in batch b and $n_{kb} = 10d_{kb}$, the target number of households to be drawn from zip code k in batch b .
7. Send the list of drawn zip codes and their target numbers of households, i.e., the pairs (k, n_{kb}) , to our vendor MSG. MSG then drew n_{kb} addresses (without replacement) at random from zip code k in the USPS Delivery Sequence File.

In order to estimate the probability $\pi_{hk|s=5}$ that household h in zip code k (from LA County) is invited, we make a few simplifying assumptions.

- The zip code sampling probability p_{kbt} does not depend on b and t . Hence, we can write it as $\pi_{k|s=5}$.
- The population of each zip code k does not change over the period of interest. The population size (number of households/addresses) of zip code k is N_k .

These are not strictly true, but they hopefully have little effect on the weights and it is hard to imagine making useful progress without them.

Probability of inviting a household

It now follows that the probability that household h from zip code k is drawn, conditional on zip code k being drawn and conditional on the values of n_{kb} , is

$$\pi_{h|k, \{n_{kb}\}} = n_k / N_k,$$

where $n_k = \sum_{b=1}^B n_{kb}$. Note that it is implicitly assumed that $n_k < N_k$. We can extend this to the case of the zip code k not being drawn ($n_k = 0$), in which case $\pi_{h|k, \{n_{kb}\}} = 0$.

Let $d_k = \sum_{b=1}^B d_{kb}$ be the total number of times zip code k has been drawn and let $M = \sum_{b=1}^B K_b$ be the total number of draws at the zip code level. Then under the simplifying assumptions stated, d_k follows a binomial distribution with parameters M and $\pi_{k|s=5}$:

$$\Pr(d_k = j) = \binom{M}{j} \pi_{k|s=5}^j (1 - \pi_{k|s=5})^{M-j}, \quad j = 0, \dots, M.$$

Combining the two components and noting that $n_k = 10d_k$, we obtain the marginal probability that household h in zip code k is drawn as

$$\begin{aligned} \pi_{hk|s=5} &= \sum_{j=0}^M \binom{M}{j} \pi_{k|s=5}^j (1 - \pi_{k|s=5})^{M-j} \frac{10j}{N_k} \\ &= \frac{10}{N_k} \sum_{j=1}^M \frac{M!}{(j-1)! (M-j)!} \pi_{k|s=5}^j (1 - \pi_{k|s=5})^{M-j} \\ &= \frac{10M\pi_{k|s=5}}{N_k} \sum_{\ell=0}^{M-1} \frac{(M-1)!}{\ell! (M-1-\ell)!} \pi_{k|s=5}^\ell (1 - \pi_{k|s=5})^{M-1-\ell} \\ &= \frac{10M\pi_{k|s=5}}{N_k} \\ &= \frac{10\mu_k}{N_k}, \end{aligned}$$

where $\mu_k = M\pi_{k|s=5}$ is the expected number of times zip code k is drawn.

Estimating μ_k

Let x_k be a vector of characteristics of zip code k . We include demographic composition (sex, age, race/ethnicity, marital status, education), urbanicity, and population quintile in this. We then estimate $\pi_{k|s=5}$ as a logit probability

$$\pi_{k|s=5} = \frac{\exp(x'_k \beta)}{1 + \exp(x'_k \beta)}.$$

The parameters β can be estimated by writing out the loglikelihood and estimating the model by maximum likelihood. In Stata, we do this using the `glm` command, with a command structure like

```
glm devar varlist , family(binomial  $M$ ) link(logit)
```

where *devar* is the name of the variable containing d_k , *varlist* is the list of variables in x_k , and M is the total number of draws introduced above. After this, `predict` computes the estimate of μ_k .

E.6 Frame 6: California ($s = 6$; batches 15 & 16)

The California batches were drawn with a two-stage method that is similar to the algorithm for drawing the nationally representative two-stage batches (frame 4). In particular, sampling of zip codes was without replacement. Only zip codes from California were included. Zip codes from Los Angeles County were allowed to be drawn, but an indicator for LA County was added to the set of distributional characteristics, so LA County zip codes had a lower probability of being drawn, because the UAS already overrepresented those on account of the LA County batches.

In the second stage, the number of households invited per zip code was proportional to population size (number of individuals age 18+), with an a priori average of 40 households per zip code if every zip code had the same probability of being selected.⁸ This was then slightly scaled to make the total number of invites equal to the target if the last drawn zip had pushed it over the target, and then non-integer numbers of invites were probabilistically rounded up or down to obtain an integer number with the same expectation. These are, however, fairly minor and inconsequential details, so for computing the probability that a household was drawn, we assume that the number of invites (if drawn) for zip code k is a given real number $n_k = 40(I_k/\mu_I)$, where I_k is the number of individuals in the population in zip code k and μ_I is the unweighted mean of I_k across all zip codes in California as computed from the 2012–2016 5-year average ACS file. Then the probability that household h from zip code k is drawn, conditional on zip code k being drawn, is $\pi_{h|k,s=6} = n_k/N_k$, where N_k is the number of households in zip code k .

Consequently, the probability of a household being invited in this frame is $\pi_{hk|s=6} = \pi_{h|k,s=6}\pi_{k|s=6}$ with $\pi_{h|k,s=6} = n_k/N_k$. We estimate the probability $\pi_{k|s=6}$ that zip code k is drawn in the same way as with frame 4, i.e., with a logit model with the same regressors, though restricted to California zip codes ($\pi_{k|s=6} = 0$ for zip codes outside California).

⁸It turns out that larger zip codes initially had a higher probability of being drawn, so that the average number of invites per zip was about 70 in batch 15. It was about 45 in batch 16.

E.7 Frame 7: Nationally representative 2-stage, proportional to size ($s = 7$; batch 17)

This frame used the same algorithm as the California frame (frame 6), except that this included zip codes nationwide instead of just California, and that the number of invites was based on the number of households in the zip code instead of the number of individuals, that is, $n_k = 40(N_k/\mu_N)$, where μ_N is the mean number of households per zip code. With these adjustments, the probability of a household being invited in this frame is $\pi_{hk|s=7} = \pi_{h|k,s=7}\pi_{k|s=7}$ with $\pi_{h|k,s=7} = n_k/N_k = 40/\mu_N$. Again, we estimate the probability $\pi_{k|s=7}$ that zip code k is drawn with a logit model with the same regressors as in frames 4 and 6.

References

- Alattar, L., Messel, M., & Rogofski, D. (2018). An introduction to the Understanding America Study internet panel. *Social Security Bulletin*, 78(2), 13–28. Retrieved from <https://www.ssa.gov/policy/docs/ssb/>
- Angrisani, M., Kapteyn, A., Meijer, E., & Saw, H. W. (2014, May 20). *Recruiting an additional sample for an existing panel*. Poster presented at the Panel Survey Methods Workshop, Ann Arbor, MI.
- Battaglia, M. P., Hoaglin, D. C., & Frankel, M. R. (2009). Practical considerations in raking survey data. *Survey Practice*, 2(5). doi: 10.29115/SP-2009-0019
- Lohr, S. L. (2009). Multiple-frame surveys. *Handbook of Statistics*, 29A, 71–88. doi: 10.1016/S0169-7161(08)00004-7
- Tourangeau, R., Brick, J. M., Lohr, S., & Li, J. (2017). Adaptive and responsive survey designs: A review and assessment. *Journal of the Royal Statistical Society A*, 180, 203–223.
- Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples*. New York: Springer.