

Managing Respondent Burden for a Panel using Permanent Random Number Sampling

N. Ganesh¹, Vicki Pineau¹, J. Michael Dennis¹

¹NORC at the University of Chicago, 4350 East-West Highway, Suite 800, Bethesda, MD 20814

Abstract

Multi-client survey platforms such as probability or non-probability sample panels may inadvertently place a high respondent burden on a select number of panelists. When selecting samples from a panel for client studies, some panelists might be selected much more frequently; furthermore, some panelists may rarely get selected for client studies. The use of permanent random number sampling reduces the problem of unequal respondent burden by “equally distributing” client survey assignments across the entire panel. In this paper, we discuss how we implemented permanent random number sampling for selecting panelists for client studies conducted using NORC’s AmeriSpeak® Panel. We also present results from a simulation study under the permanent random number sampling approach that compares two methods of within household sample selection; we show via simulation that for households with two or more recruited panelists that a probability proportional to size approach performs better in terms of managing respondent burden when compared to random selection of a panelist within a household.

Key Words: AmeriSpeak Panel, permanent random number sampling, panel sample, web survey

1. Introduction

Funded and operated by NORC at the University of Chicago, AmeriSpeak® is a probability-based panel sample designed to be representative of the U.S. household population. Randomly selected U.S. households are sampled with a known, non-zero probability of selection from the NORC National Frame, and then contacted by U.S. mail, telephone interviewers, overnight express mailers, and field interviewers (face-to-face). AmeriSpeak panelists participate in NORC studies or studies conducted by NORC on behalf of NORC’s clients.

The sample frame for the AmeriSpeak is the NORC National Frame, an area probability sample frame constructed by NORC providing sample coverage of 97 percent of U.S. households. The NORC National Frame itself contains almost 3 million households, including over 80,000 rural households added through in-person listing of households that were not recorded on the USPS Delivery Sequence File (see Pedlow and Zhao, 2016).

Once the sample is selected from the National Frame, AmeriSpeak Panel sample recruitment is a two-stage process: initial recruitment using less expensive methods and then non-response follow-up using personal interviewers. For the initial recruitment, sample addresses are invited to join AmeriSpeak by visiting the panel website

AmeriSpeak.org or by telephone (in-bound/outbound). As of July 2017, the AmeriSpeak Panel weighted AAPOR 3 response rate was 33.5% (Montgomery, Dennis, and Ganesh, 2017). For further details on AmeriSpeak, please see Dennis (2017) and <http://amerispeak.norc.org/about-amerispeak/panel-design/>.

For an address that was sampled for AmeriSpeak recruitment and for which a specific adult in the household is recruited into the panel, all other adults in that household are nominated for recruitment by the first recruited adult in the household. Subsequently, NORC attempts to recruit other eligible adults from the recruited household using contact information provided by the first recruited adult. A household is considered to be recruited if at least one adult in the household was recruited into the AmeriSpeak Panel. A panelist is considered active if the panelist has not been withdrawn from the AmeriSpeak Panel; similarly, a household is considered active if at least one panelist in the household has been previously recruited and is currently active.

On average, AmeriSpeak panelists are selected for 2-4 client surveys per month. Given the cost associated with recruiting a panelist, there is a need to balance the number of surveys that a panelist is asked to take in order to manage response rates and panel attrition. Some panelists are selected for more surveys than the average; in particular, hard-to-survey subgroups such as minorities, young adults, and low socio-economic groups are sampled more frequently for studies.

In this paper, we discuss using permanent random number sampling (Ohlsson, 1992 and 1995) to manage respondent burden when selecting samples from the AmeriSpeak Panel for client studies. In Section 2, we provide some background on how we implement permanent random number sampling for the AmeriSpeak Panel. In Section 3, we provide details on the sampling requirements for a typical client study conducted using the panel. In Section 4, we discuss a simulation study and provide a summary of the results along with a comparison of two different within household selection methods. Finally, in Section 5, we provide some concluding remarks.

2. Permanent Random Number Sampling

Permanent random number (PRN) sampling provides an efficient method to control sample overlap when selecting multiple samples from the same sampling frame. Ohlsson (1992 and 1995) describes PRN sampling in the context of business surveys, to minimize respondent burden for businesses by using sample rotation. Using a similar PRN sampling method as described in Ohlsson (1992), controlling sample overlap of all client surveys selected from the AmeriSpeak Panel effectively reduces the number of surveys that are assigned to a particular panelist, thereby reducing respondent burden for panel members. Under PRN sampling, every recruited AmeriSpeak panelist is independently assigned a permanent random number between 0 and 1 (generated from a uniform distribution), and this PRN value is associated with the panelist throughout their panel tenure. To select a client sample of size n from the panel, the panelists are ordered (in increasing order) based on their PRN values, and then a start point between 0 and 1 is selected (as explained below), and from that start point the first n panelists are selected as the sample for the given study. If there is an insufficient number of panelists prior to hitting the maximum PRN value of 1, the sample is continued, wrapping around the PRN value of 0, and then selection of panelists is continued until n panelists have been selected.

The start point for PRN sampling is not randomly chosen; instead, after each sample selection for a client survey, we store the PRN value associated with the “last” selected panelist for the client survey. The start point for the “next” study is the first PRN value after the last stored PRN value for the previous study. For example, if the PRN value of the last selected panelist in the previous study is 0.456, then the start point for the “next” study is the first PRN value after 0.456 (say it is 0.462). From the start point of 0.462, we select the first n panelists as the sample for the “next” study. Note that PRN sampling is simple random sampling without replacement (SRSWOR) of size n from a population of size N .

In practice, when selecting a sample for a client survey from the panel, instead of selecting a SRSWOR from the panel as explained above, we stratify the panel based on various demographic characteristics creating sampling strata (see Section 3), apply PRN in each sampling strata, thus selecting a stratified random sample from the panel. When selecting the stratified random sample, we use as the start point for each sampling stratum, the first PRN value after the PRN value associated with the last selected panelist from the previous sample selection. After each sample selection, we store the PRN value associated with the last selected panelist within each stratum. These stored PRN values for each stratum are then used to inform as to what PRN value to use as the start point for each stratum when selecting a sample for the next study.

3. Within-Panel Sampling for Client Studies

When selecting a sample from the AmeriSpeak Panel for a client study, we attempt to select the sample such that on an unweighted basis, the expected distribution of completes for the study matches the corresponding target population distribution by race/Hispanic ethnicity, age, gender, and education. In order to achieve this, the standard AmeriSpeak approach is that we define 48 sampling strata by cross-classifying the panel using the following demographic variables:

1. Race/Hispanic ethnicity (3 categories: Hispanic, non-Hispanic Black, non-Hispanic All Other)
2. Age (4 categories: 18-34 years, 35-49 years, 50-64 years, 65+ years)
3. Gender (2 categories: male, female)
4. Education (2 categories: some college or less, 4-year college graduate or above)¹

For most surveys, these 48 sampling strata are used for selecting a representative sample of the target population. As explained in Section 2, we use PRN sampling within each of the sampling strata. The required total number of completes for the study is proportionately allocated to each stratum based on the population distribution using the most recent Current Population Survey (CPS). The required number of completes for each stratum is inflated by the reciprocal of the expected survey completion rate (which varies by stratum and is estimated using prior surveys) to obtain the required sample size by stratum. The sample size n_k^* for sampling stratum k is calculated as

$$n_k^* = \frac{n^* p_k}{r_k} \quad (1)$$

¹ Additional demographic variables (such as household income, housing tenure) and/or geographic variables (such as Census Division) could be used as sampling strata under the PRN approach.

where n^* is the required total number of completes, p_k is the population proportion in stratum k as estimated using CPS, and r_k is the expected survey completion rate (based on prior surveys) for stratum k .

In addition to accounting for expected response rates in the selected sample within each sampling strata, there are two further complications addressed with the AmeriSpeak PRN approach:

1. Some studies require a subset of panelists; not all studies are general population studies of adults 18+ years of age; and
2. We generally require that only one panelist per household be selected for a given client study.

In regards to surveys that target a specific subpopulation and where only a subset of all panelists are eligible, ineligible panelists are excluded prior to implementing the PRN sampling approach described in Section 2. That is, for example, if a study targets married males between ages 18 and 55, then we subset the panel to active panelists who are eligible for the study (marital status is collected during panel recruitment), and then select the required sample in each stratum using an identical PRN sampling approach as explained in Section 2 and using (1) to determine the appropriate sample sizes by sampling strata. Note that after sample selection, we store the PRN value associated with the last selected panelist in each stratum. In some strata, no sample would have been selected (for example, when the target population is married males between ages 18 and 55, no sample would be selected in the strata associated with females or adults 65+ years), and for these strata, we retain the last PRN value from the prior study.

The second complication associated with sampling from the AmeriSpeak Panel is that we generally select at most one panelist per household. We select at most one panelist per household to reduce the within household clustering effect. Thus, when using the PRN sampling approach, we first limit the AmeriSpeak Panel to one panelist per household, and then implement the PRN sampling approach discussed in Section 2. Our original approach for selecting one panelist per household was to randomly select one panelist in households with two or more active and recruited panelists (**Method 1: Random Within Household Selection**). However, this resulted in panelists in households with two or more active and recruited panelists being rarely selected for client studies. Thus, we modified our within household selection mechanism (**Method 2: Probability Proportional to Size Within Household Selection**) to:

- (a) Randomly select a panelist if two or more panelists would have been selected from the same household under the PRN approach with no restriction on selecting multiple panelists per household, or
- (b) If a household contains two or more panelists and none of them would have been selected under the PRN approach with no restriction on selecting multiple panelists per household, then select a panelist proportional to the “distance” between the PRN value of the panelist and the PRN value of the last panelist that would have been selected if there was no restriction on selecting multiple panelists per household.

For additional details on the Probability Proportional to Size Within Household Selection approach, please see below:

1. Compute *PRN_SAMP* as follows

- A. If the PRN value for a panelist is smaller than the start point then define $PRN_SAMP = 1 + PRN$.
 - B. Else if the PRN value for the panelist is larger than the start point, then define $PRN_SAMP = PRN$.
2. For stratum k , let $LAST_PRN_SAMP_k$ denote the PRN_SAMP value for the last panelist in the given sampling stratum that would have been sampled assuming we select panelists without restricting to one panelist per household.
 - A. For each stratum k , select all panelists with $PRN_SAMP \leq LAST_PRN_SAMP_k$; denote this set by C'_1 .
 - B. If C'_1 contains two or more panelists in a given household, then randomly select one of the panelist in such a household; denote this set by C_1 .
 - C. Let C'_2 denote the set of panelists after excluding all panelists in households associated with panelists C_1 . That is, C'_2 does not include any panelist that resides in a household associated with panelists C_1 .
 - D. If C'_2 contains two or more panelists in a given household, then randomly select one of the panelists from such a household with probability proportional to $\frac{1}{PRN_SAMP - LAST_PRN_SAMP_k}$; denote this set by C_2 .
 - E. The set C of panelists eligible for sampling for the new study is then defined as $C = C_1 \cup C_2$.

4. Simulation Study

In order to evaluate the PRN sampling approach and to compare the two within household selection methods, we conducted a simulation study. The target population for the simulation study was adults 18+ year of age. Similar to a typical AmeriSpeak client study, the objective was to achieve 1,000 completes assuming a survey completion rate that we typically expect for a study with a 10-14 day field period. The simulation was replicated 500 times; that is, the 500 replications is like conducting 500 general population studies of adults 18+ years of age using the AmeriSpeak Panel. The objective of the simulation was to evaluate the number of survey assignments (among those 500 simulated surveys) by sampling stratum, number of recruited panelists in the household (one recruited panelist in the household vs. two or more recruited panelists in the household), and the two within household selection methods. For a given within household selection method, after each of the 500 replications, the last PRN value was stored for each sampling stratum. Then, for each sampling stratum, the first PRN value after the last stored PRN value (from the prior simulation replicate) was used as the start point for sample selection for the new simulation replicate.

Table 1 provides the distribution of the “wait time” until a panelist is selected for a client survey, where “wait time” refers to the number of surveys (or number of simulations) prior to a panelist being selected for a client study. For example, a wait time of 3.5 indicates that on average the given panelist would need to wait for 3.5 general population client surveys prior to being sampled for one such survey.

- Variation in wait time reflects the longer (or shorter) wait time for panelists in sampling strata that have higher (or lower) survey completion rates and/or specific sampling strata being over (or under) represented in the panel. For example, if

there is one recruited panelist in the household, for the sampling stratum “Hispanic, 18-34, male, Some college or less”, the median wait time is 3.3 or 3.7 surveys (depending on the within household selection method), but for the sampling stratum “NH Black, 50-64, female, college graduate or higher”, the median wait time is 18.9 or 20.0 (see Table 2).

- For households with one recruited panelist, both within household selection methods have a similar wait time. For example, the median wait time under Random Within Household Selection (Method 1) and Probability Proportional to Size Within Household Selection (Method 2) respectively is 7.7 and 8.3 (see Table 1).
- For households with two or more recruited panelists, Probability Proportional to Size Within Household Selection has a shorter wait time compared to Random Within Household Selection (see Table 2). Specifically, for households with two or more recruited panelists, when comparing the wait times under Random Within Household Selection to Probability Proportional to Size Within Household Selection, the median wait time was reduced by 55% (14.7 vs. 8.1).

Table 2 provides the median (across all panelists in a given sampling stratum) wait time until a panelist is sampled for a client study by sampling stratum, within household selection method, and number of recruited panelists in household.

- For both methods, wait times are longer for panelists in households with two or more recruited panelists compared to panelists in households with one recruited panelist.
- For Random Within Household Selection, wait times for panelists in households with two or more recruited panelists is almost twice as long as panelists in households with one recruited panelist; a similar comparison for Probability Proportional to Size Within Household Selection, indicates that the wait times for panelists in households with two or more panelists is only ~10% longer when compared to panelists in households with one recruited panelist.
- Wait times are typically shorter for sampling strata associated with Hispanic or non-Hispanic Black or young adults or males or an education attainment of some college or less; this is a result of additional these sampling strata having a lower survey completion rate compared to sampling strata associated with non-Hispanic White or older adults or college graduate or higher. Another reason for the shorter wait time for Hispanic or non-Hispanic Black or young adults is the additional oversampling of these groups in some client studies.

Table 1: Distribution of wait time until a panelist is selected for a client survey by within household sampling method and number of recruited panelists in household.

	<i>1 panelist in household</i>		<i>2+ panelists in household</i>	
	<i>Random Within Household Selection</i>	<i>Probability Proportional to Size Within Household Selection</i>	<i>Random Within Household Selection</i>	<i>Probability Proportional to Size Within Household Selection</i>
Minimum	3.2	3.4	5.4	3.5
25th percentile	6.2	6.7	11.6	6.8
Median	7.7	8.3	14.7	8.1
75th percentile	10.0	10.4	20.0	10.9
Maximum	19.2	20.0	125.0	29.4

Table 2: Median wait time until a panelist is sampled for a client study by sampling stratum, within household selection method, and number of recruited panelists in household.

<i>Sampling Stratum**</i>	<i>Random Within Household Selection</i>		<i>Probability Proportional to Size Within Household Selection</i>	
	<i>1</i>	<i>2+</i>	<i>1</i>	<i>2+</i>
	<i>panelist in HH*</i>	<i>panelists in HH*</i>	<i>panelist in HH*</i>	<i>panelists in HH*</i>
Hispanic, 18-34, male, Some college or less	3.3	6.8	3.7	4.1
Hispanic, 18-34, female, Some college or less	5.8	12.2	6.3	6.9
Hispanic, 18-34, male, college graduate or higher	5.8	12.8	6.6	7.0
Hispanic, 18-34, female, college graduate or higher	7.9	16.1	8.5	9.3
Hispanic, 35-49, male, Some college or less	3.2	6.7	3.4	3.7
Hispanic, 35-49, female, Some college or less	6.7	14.3	7.1	8.1
Hispanic, 35-49, male, college graduate or higher	6.8	13.5	7.6	7.9
Hispanic, 35-49, female, college graduate or higher	9.3	18.2	9.8	10.6
Hispanic, 50-64, male, Some college or less	3.9	7.8	4.2	4.5
Hispanic, 50-64, female, Some college or less	5.4	11.4	5.9	6.3
Hispanic, 50-64, male, college graduate or higher	8.5	16.7	8.9	9.3
Hispanic, 50-64, female, college graduate or higher	10.9	23.3	11.6	12.5
Hispanic, 65+, male, Some college or less	5.3	11.4	5.7	6.4
Hispanic, 65+, female, Some college or less	5.0	10.0	5.4	5.9
Hispanic, 65+, male, college graduate or higher	16.7	35.9	17.0	17.9
Hispanic, 65+, female, college graduate or higher	11.6	23.3	12.5	13.5
NH Black, 18-34, male, Some college or less	4.2	8.6	4.9	5.2
NH Black, 18-34, female, Some college or less	7.4	14.5	8.1	8.8
NH Black, 18-34, male, college graduate or higher	5.7	11.9	6.4	6.8
NH Black, 18-34, female, college graduate or higher	9.3	19.2	10.2	10.6
NH Black, 35-49, male, Some college or less	6.5	13.5	7.1	7.6
NH Black, 35-49, female, Some college or less	10.6	20.8	11.4	12.2
NH Black, 35-49, male, college graduate or higher	7.4	14.3	7.8	8.1
NH Black, 35-49, female, college graduate or higher	12.8	26.3	13.5	14.3
NH Black, 50-64, male, Some college or less	9.6	19.2	10.4	10.9
NH Black, 50-64, female, Some college or less	13.9	27.8	15.2	16.1
NH Black, 50-64, male, college graduate or higher	13.2	26.3	13.9	14.3
NH Black, 50-64, female, college graduate or higher	18.9	38.5	20.0	20.8
NH Black, 65+, male, Some college or less	7.8	15.2	8.3	8.8
NH Black, 65+, female, Some college or less	11.1	21.7	11.6	12.5
NH Black, 65+, male, college graduate or higher	16.7	34.5	17.9	17.9
NH Black, 65+, female, college graduate or higher	17.9	35.7	18.5	20.0
NH All Other, 18-34, male, Some college or less	4.1	8.3	4.6	4.9
NH All Other, 18-34, female, Some college or less	6.6	13.2	7.0	7.7

Table 2: Median wait time until a panelist is sampled for a client study by sampling stratum, within household selection method, and number of recruited panelists in household.

<i>Sampling Stratum**</i>	<i>Random Within Household Selection</i>		<i>Probability Proportional to Size Within Household Selection</i>	
	<i>1 panelist in HH*</i>	<i>2+ panelists in HH*</i>	<i>1 panelist in HH*</i>	<i>2+ panelists in HH*</i>
	NH All Other, 18-34, male, college graduate or higher	6.0	12.5	6.6
NH All Other, 18-34, female, college graduate or higher	7.8	15.6	8.5	9.3
NH All Other, 35-49, male, Some college or less	4.8	9.6	5.3	5.6
NH All Other, 35-49, female, Some college or less	9.1	18.5	9.6	10.6
NH All Other, 35-49, male, college graduate or higher	7.0	14.3	7.6	7.9
NH All Other, 35-49, female, college graduate or higher	10.0	20.8	10.6	11.4
NH All Other, 50-64, male, Some college or less	6.2	12.2	6.7	7.0
NH All Other, 50-64, female, Some college or less	9.3	19.2	9.8	10.6
NH All Other, 50-64, male, college graduate or higher	9.1	17.9	9.8	10.2
NH All Other, 50-64, female, college graduate or higher	11.1	21.7	11.6	12.5
NH All Other, 65+, male, Some college or less	6.3	12.5	6.8	7.1
NH All Other, 65+, female, Some college or less	6.8	13.9	7.2	7.7
NH All Other, 65+, male, college graduate or higher	14.3	27.8	15.2	16.1
NH All Other, 65+, female, college graduate or higher	13.5	26.3	14.3	14.7

*HH refers to household

**NH refers to non-Hispanic

5. Conclusion

When selecting client samples from the AmeriSpeak Panel, permanent random number sampling provides an efficient way of managing respondent burden. In general, panelists who are non-Hispanic All Other or older or female or have an educational attainment of 4-year college degree or higher are selected for fewer studies given their higher survey completion rates when compared to panelists who are Hispanic or non-Hispanic Black or younger or male or have an education attainment of some college or less. In our simulation study, for households with two or more recruited and active panelists, Random Within Household Selection resulted in panelists being selected for fewer surveys; Probability proportional to Size Within Household Selection was found to be a better within-household selection method as it helps reduce over-use of some panelists (who live in households with just one recruited panelist) and under-use of other panelists (who live in households with two or more recruited panelists). For households with two or more recruited panelists, when comparing the wait times under Random Within Household Selection to Probability Proportional to Size Within Household Selection, the median wait time was reduced by 55% (14.7 vs. 8.1).

References

- Dennis, J.M. 2017. Technical overview of the AmeriSpeak Panel.
http://d3qi0qp55mx5f5.cloudfront.net/amerispeak/i/research/AmeriSpeak_Technical_Overview_2017_05_09.pdf?mtime=1494625611
- Montgomery, R., Dennis, J.M., and Ganesh, N. 2016. Response rate calculation methodology for recruitment of a two-phase probability-based panel: The case of AmeriSpeak.
http://d3qi0qp55mx5f5.cloudfront.net/amerispeak/i/research/WhitePaper_ResponseRateCalculation_AmeriSpeak_2016.pdf?mtime=1478878716
- Ohlsson, E. 1992. The system for co-ordination of samples from the business register at Statistics Sweden. R&D Report, Statistics Sweden.
- Ohlsson, E. 1995. Coordination of samples using permanent random numbers. In Business Survey Methods. Wiley.
- Pedlow, S. and Zhao, J. Bias reduction through rural coverage for the AmeriSpeak Panel. Presented at AAPOR 2016, Austin, TX.
http://d3qi0qp55mx5f5.cloudfront.net/amerispeak/i/research/AAPOR_Pedlow_2016_05_15.pdf?mtime=1472056688