Comparing Two Common Approaches to Within-household Sampling:

A Field Experiment in Costa Rica

Noam Lupu[1], J. Daniel Montalvo[1], Mitchell A. Seligson[1], Elizabeth J. Zechmeister[1], and Kirill Zhirkov[2]

[1] Vanderbilt University

[2] University of Virginia

Author Note

Dr. Mitchell (Mitch) A. Seligson passed away on June 1, 2024. Mitch directly contributed to the research design and data collection components of this paper and supported the development of early versions of the manuscript. We acknowledge Mitch's contributions to this manuscript, alongside his many contributions of the field of public opinion research. Authors are listed alphabetically. We are grateful to Claire Evans, Georgina Pizzolitto, Bryce Williams-Tuggle, Carole Wilson, and Adam Wolsky for their research assistance.

Abstract

We test the notion that a quasi-probabilistic method of selecting individuals within households (last birthday, LB) draws in a different sample compared to a non-probabilistic approach that selects respondents according to known parameters on age and gender (frequency matching, FM). With data from an original field experiment, we evaluate fieldwork efficiency (time and completed cases), economy (cost), success in recruiting a representative sample, and differences across a set of attitudinal and behavioral measures. We find that the FM approach performs better on efficiency and cost and achieves a comparable sample; importantly, this comparability extends across measures of personality traits and public opinion. With appropriate caveats, we conclude that researchers' choice of selection methods should be guided by both theoretical benefits and practical tradeoffs.

Keywords: developing countries, frequency matching, last birthday, respondent selection, sampling, survey design

1. Introduction

In probability-based general population surveys of individuals, researchers often need a method to select respondents from within the household. In theory, a pure probability-based approach (e.g., random selection from household registers) should be carried through to this final stage. In practice, other common techniques for within-household selection include probabilistic (typically, Kish) and quasi-probabilistic approaches (typically, last birthday; Koch 2018). Among these, quasi-probability approaches are often preferred because they limit interviewer discretion while simultaneously reducing the burden associated with methods, like the Kish method, that begin by soliciting a list of household members (Binson, Canchola, and Catania 2000; Gaziano 2005). Yet, for general population opinion surveys in developing contexts, the modal approach for selecting within the household is non-probabilistic (Lupu and Michelitch 2018). In theory, non-probabilistic approaches to within-household selection can lead to significant biases if people who are first or most willing to respond, or who are selected at the interviewers' discretion, are systematically different (e.g., more cooperative and engaged) compared to the rest of the population. Yet to our knowledge there has been very little systematic testing of these theoretical expectations in practice (for an exception, see Guignard et al. 2013). We contribute to this ongoing conversation with an empirical study that asks the following question: what differences emerge when a quasi-probability versus non-probabilistic approach is implemented in the field for the last stage of within-household selection?

Our data comes from a large-scale experiment designed to compare a last-birthday approach to a non-probabilistic within-household approach that selects individuals according to known population frequencies of gender and age. We have decided to focus on these two methods to maximize the relevance of the project: these are the two most frequently used approaches in practice in opinion surveys in developing contexts. Major comparative opinion projects in developing contexts are exemplars: the Afrobarometer alternates selecting men and women within households, a non-probabilistic method designed to match to known binary gender frequencies (Afrobarometer, n.d.); the Arab Barometer's approach is similar, though sometimes uses last-birthday selection (Arab Barometer, n.d.); and the AmericasBarometer uses a non-probabilistic approach in which interviewers select potential respondents within the household based on known age and gender frequencies (AmericasBarometer, n.d.).

We carried out the experiment in the Greater Metropolitan Area of San José, the capital of Costa Rica. Within pairs of census segments that constitute the secondary sampling unit in the same survey, interviewers were instructed to apply in one segment a quasi-probabilistic, last birthday method of within-household selection, with up to ten recontacts. In the other segment, recruitment followed a gender and age frequency matching approach without recontact. We document differences in fieldwork efficiency (time and completed cases), economy (cost), and success in recruiting a representative sample. Our key dependent variables measure two sets of selected attitudes and behaviors. The first is a set of personality measures, relevant across a range of disciplines. The second is a set of attitudes and behaviors that are associated with stable and efficient democratic governance. The former are related to cross-disciplinary concerns about different "types" of individuals being recruited conditional on the approach, while the latter measures allow us to explore a set of variables relevant to comparative public opinion research.

Our study makes three core contributions. First, while scholars have considered the implications of various probability and quasi-probability approaches on survey outcomes (Battaglia et al. 2008; Binson, Canchola, and Catania 2000; Oldendick et al. 1988; O'Rourke and Blair 1983; Salmon and Nichols 1983), to our knowledge we are the first to extend this body of research to a comparative assessment involving a frequency matching approach. Second, while scholars have speculated that non-probabilistic approaches to within household selection of respondents may produce samples biased toward certain traits (Brick 2011; Clark and Steel 2007; Koch 2018), our study subjects those notions to a test, and findings no substantive difference in personality traits or in a set of variables relevant to work on democratic public opinion. Third, we raise awareness among scholars regarding different methods – and potential outcomes – for within-household selection that are common in surveys in developing contexts. By extension, by addressing design choices and caveats, we contribute to developing a research agenda on survey methods in developing countries.

2. Selection of Individuals within Households

Probabilistic methods are widely considered the gold standard for within-household selection of respondents in household surveys. For situations in which accurate household registers are unavailable, a popular approach involves enumerating all individuals living in a household and then randomly selecting a person from that list (Kish 1949). The drawbacks of this method are its length and intrusiveness: it increases the time necessary to complete an interview and the probability that individuals refuse to cooperate because of its complexity or concerns about sharing household details (Battaglia et al. 2008; Gaziano 2005; Jabkowski 2017). In addition, the approach is burdensome to interviewers and thus may lead to suboptimal outcomes (Binson, Canchola, and Catania 2000).

To overcome these problems, methodologists have developed "quasi-probability" approaches that maintain the arbitrary nature of respondent selection without enumerating all the individuals within the household (see Gaziano 2008). Examples are the next- and last-birthday methods: the first person contacted is asked to name the individual with the next upcoming or the most recent (last) birthday, who is automatically selected (Salmon and Nichols 1983). Survey methodology research shows that, compared to the enumeration approach, next- and last-birthday methods yield little to no differences in substantive responses (Oldendick et al. 1988), while they simultaneously decrease survey length and intrusiveness (O'Rourke and Blair 1983), result in fewer dropouts (Binson, Canchola, and Catania 2000), and may produce response rates that are higher than more intrusive probabilistic approaches and similar to non-probabilistic selection methods (Battaglia et al. 2008; Salmon and Nichols 1983).

Another large class of methods for within-household selection is not based on probabilistic or quasi-probabilistic techniques. Instead, individuals are chosen by target demographic characteristics, usually age and gender population frequencies derived from the census. There are many different selection methods within this class (Bryant 1975; Hagan and Collier 1983; Paisley and Parker 1965), often designed to compensate for underrepresentation of certain demographic groups in probabilistic sampling approaches (Jabkowski 2017; Yan et al. 2009). These methods trade off the advantages of probabilistic selection within households for lower fieldwork times, more overall efficiency in fieldwork (given elimination of recontacts), higher response rates, and less need for calibration weighting. There are potentially other benefits, such as the need to match the genders of interviewer and respondent in certain contexts, which makes random selection impractical (Le et al. 2014).

Underrepresentation by age, gender, and other demographic characteristics, the main problem that such selection methods are intended to solve, can indeed be substantial depending on context, mode, and study (Christiansen et al. 2014; Gaziano 2005; Groves and Couper 1998). In face-to-face surveys in developing countries, younger men can be underrepresented in studies that use probabilistic methods for within-household selection. This likely stems from non-contact (vs. non-cooperation); young men are comparatively less likely to be home (Groves and Couper 1998). Conversely, women may be comparatively more likely to be at home due to gender norms regarding responsibilities for managing the household and children within it; in developing countries, women – including among older cohorts – are comparatively more engaged in caregiving activities (Lupica 2024).

It is widely theorized that non-probabilistic approaches for the selection of individuals can generate samples that are biased toward respondents with certain personality traits. Researchers argue, for instance, that non-probabilistic approaches end up selecting more cooperative individuals who are more interested in the survey topic (Brick 2011; Clark and Steel 2007; Koch 2018). Yet, the field lacks studies that test whether these intuitions hold empirically. In many situations, the individual recruited may be the same regardless of the method – that is, it may often be the case that the person with the last birthday who is willing to consent to an interview is also the one who is present and willing to engage at the time of contact in a non-probabilistic approach. Such an outcome may be due to chance (with that probability higher in smaller households) or due to error (on the potential for errors in enumerating household members and/or recalling the most recent birthday, see Battaglia et al. 2008; Martin 1999; Tourangeau et al. 1997). In brief, while in theory two approaches may yield different samples, in practice certain outcomes may be quite similar.

3. Outcome Variables

We identify a range of outcomes to assess when comparing the quasi-probabilistic last birthday versus the non-probabilistic frequency matching method:

1. *Efficiency and economy.* Our priors are that, compared to a last birthday method, the frequency matching method is more efficient in fieldwork time, has a lower per-interview

cost, and has higher response rates. Some research shows that probabilistic methods for selecting respondents within households do not always perform better on other dimensions; for example, non-probabilistic methods can offer substantial cost savings (Cumming 1990; Czaja et al. 1982; Olson, Stange, and Smyth 2014).

2. *Representativeness of sample.* We also anticipate that, compared to a last birthday method, the frequency matching method produces a sample closer to the census-based population on age, gender, education, and wealth – variables available in most census data and often used for both sample description and analysis. Even with multiple recontacts, samples using probabilistic approaches to within-household selection can end up imbalanced on demographic traits such as gender (Jabkowski 2017).
3. *Personality indicators.* As discussed in the prior section, scholarship suggests that compared to individuals selected using a last birthday method, respondents selected using frequency matching may differ in personality traits. We consider the Big Five traits: openness to experience, conscientiousness, extraversion, agreeableness, and emotional stability. These traits are recognized in psychology and political science research as core dimensions on which individuals' personalities vary (see discussion in Gerber et al. 2011) and survey researchers have connected personality traits to willingness to engage in research studies (Marcus and Schütz 2005; Loennqvist et al. 2007).
4. *Political attitudes and behaviors*. Given little prior theory, we investigate but do not establish a priori hypotheses for measures of voting behavior, political ideology, support for democracy, political trust, political participation, and political tolerance. We selected these variables because they commonly appear in public opinion and political behavior studies that are generated using large-scale comparative opinion surveys that use non-

probabilistic methods for within-household selection. Our set of outcome variables does not cover the full range that might be affected by method of within-household selection; one example is household roles (that is who takes on tasks such as opening the door for strangers, shopping, housekeeping, etc.; see Olson and Smyth 2017).

4. Data and Method

4.1. Study Design

Our objective is to provide one test of whether, in practice, a face-to-face survey yields different results when the within-household selection of respondents uses a quasi-probabilistic approach (last birthday) versus a non-probabilistic approach (frequency matching). While in theory it could be beneficial to include a pure probability approach within the study design, we opted to focus our study on two commonly applied approaches in general population survey research in developing contexts (Lupu and Michelitch 2018). We implemented a large-scale field experiment as part of a survey study in Costa Rica, a country in Central America that belongs to developing and middle-income groups according to global indicators. Average household size in Costa Rica in 2011 (the most recent census) was 3.46, which is typical for a middle-income country. We selected Costa Rica because it has detailed census information and the authors have years of experience and an extensive network of research partners in the country.

The design is an area probability survey that targeted a sample of 900 adults within the voting age population of the Greater Metropolitan Area of San José (GMASJ), administered in September–October 2018. As is standard in survey research, the sample excludes people in institutionalized settings like boarding schools, hospitals, military barracks, and alike. We focus on the GMASJ to increase control over the design and implementation of the experiment while decreasing the costs associated with the study. The study was approved by the institutional review board at our institution.

The core of the experiment is random assignment of census segments to one of two within-household sampling methods: (1) a frequency matching approach in which individuals are selected to match known distributions on age and gender, and (2) a quasi-probabilistic approach in which the individual with the most recent birthday is selected. The first approach – the frequency matching (hereafter, FM) approach – is a quota-based approach, but must be distinguished from quota-based approaches that sample from "flow points" (e.g., shopping malls); rather, the sample is probabilistic down to the household, at which point individuals are selected to match gender and age frequencies in the population. The second method – the last birthday (hereafter, LB) approach – selects individuals using a quasi-probabilistic approach and is seen as a reasonable substitute for the gold-standard Kish method (e.g., Gaziano 2005).

As in any fieldwork, these selection methods are bundled with a series of other methodological and fieldwork choices that we describe below. We sought to make choices that were consistent with typical practice in the Latin American region, such that the fieldwork mimicked what each selection method would look like in a typical application. Where there was more room for discretion within typical practice, we made fieldwork choices that would make it more likely to find substantive differences, giving the quasi-probabilistic LB approach the best possible opportunity to outperform the FM approach regarding the nature of the sample. For example, we limited substitutions in the LB method; this is consistent with textbook recommendations and reduces the extent to which the approach mirrors a nonprobabilistic method (e.g., Fowler 2009).

The sampling frame was the list of municipalities, districts, census segments, and maps for the GMASJ from the 2011 census provided by the Central American Population Research Center. GMASJ was the single stratum, and municipalities within it were primary sampling units (PSUs). We selected PSUs using two criteria. First, municipalities with more than 100,000 inhabitants were self-selected; that is, the selection probability for those municipalities equaled one. Under this rule, San José was self-selected; due to its size, it was sub-divided into two PSUs. Second, we selected the remaining municipalities applying probability proportional to estimated size (PPeS). Following this design, we selected eight PSUs within the GMASJ: Alajuelita, Aserrí, Desamparados, Escazú, Goicochea, Montes de Oca, San José, and Tibás.

In the next stage, we selected census segments with PPeS. Segment constituted secondary sampling units (SSUs). To avoid contamination of the experiment at the household level (i.e., houses on the same block falling into both conditions), we assigned the treatment at the segment level. We first determined the FM segments, by selecting at random 75 census segments from the 7,703 segments in the selected municipalities. We fixed the number of selected households within each segment to six. Fixing the number of households within segments allows us to have an evenly distributed number of interviews per segment. For the study, a household is a group of people who eat their meals together. Table 1 presents the data on selected municipalities: their population, the number of segments, and planned interviews.

Households were chosen using systematic selection, with interviewers beginning at the northeast corner of the segment, moving clockwise, skipping one house after each completed interview, and selecting one person per household. In apartment buildings, interviewers identify a starting point and select households on each floor, skipping one unit after each completed interview. If interviewers do not complete the interviews on one floor, they must proceed to the next floor, either upstairs or downstairs.

Table 1. Data on selected municipalities (PSUs)

Municipality (PSU)	Population	Share of GMASJ	Segments (SSUs)	Planned sample
Alajuelita	77,801	0.055	16	96
Aserrí	58,430	0.042	16	96
Desamparados	207,082	0.147	18	108
Escazú	56,733	0.040	16	96
Goicoechea	114,736	0.082	16	96
Montes de Oca	49,008	0.035	16	96
San José	287,619	0.205	36	216
Tibás	64,834	0.046	16	96

Note. The total population of GMASJ is 1,403,963. Segments and planned sample combine the FM and LB conditions. For instance, 16 segments for the district of Alajuelita mean 8 FM segments and 8 LB segments.

To implement the FM approach, interviewers used a matrix combining three age cohorts (18–29, 30–45, 46+) and two gender (male/female) categories to select individuals so that the sample matched known frequencies on these characteristics as established by census data. In completing six interviews per segment, interviewers were free to complete the six-cell matrix in any order and selected households were not revisited. If it was not possible to secure an interview at a given household (due to refusal, ineligibility, or no one at home), the interviewer moved on to the next household without skipping a dwelling. If the interviewer exhausted all dwellings in the selected segment, the protocol permitted them to move to another adjacent segment to complete the six interviews.

We next selected the LB segments, using an approach designed to effectively match these segments to the FM segments. To do this, we identified all the segments that were contiguous to the selected FM segments. Using information from the 2011 census, we sorted these neighboring segments by size and discarded segments with insufficient numbers of households. The decision to remove segments with an insufficient number of households was taken to avoid major obstacles during fieldwork. We then randomly selected one segment from the pool of neighbors for each FM segment, yielding a total of 75 LB segments.

In the LB condition, interviewers were also instructed to complete six interviews in each segment and to designate the northeast corner as the starting point. In these cases, the interviewer approached the first house and asked to interview the adult with the most recent (last) birthday. If an interviewer found no one at home or was asked to visit at a different time (that is, the household was selected but the target respondent within that household was not identified), they were instructed to return to that house and make another contact attempt. As in the FM approach, if successful, the interviewer was instructed to skip one dwelling before approaching the next one. If the selected individual was identified but not available, the interviewer was instructed to revisit the household up to nine more times (for a total of ten attempts). If the selected individual declined or remained unreachable after all the attempts were exhausted, the interviewer moved to the adjacent dwelling unit and repeated the same protocol; at a maximum, interviewers would visit the first 12 dwellings in the segment, if contact at the initially-selected dwelling failed each of the six times, prompting the interviewer to approach the nearest neighbor as the one permitted substitution. This design means that households at the tail of the segment generally had lower probabilities of being included in the final sample. Importantly, this probability would be zero for a dwelling at the tail of a segment with more than 12 dwellings in the LB condition.

Guidance and practices in face-to-face research range from zero to ten callbacks; we opted to allow for the higher end to maximize opportunities to realize the intended sample. We did not specify that callbacks had to vary by period of the week (weekend vs. weekday) because retaining teams in particular segments for an extended period would substantially increase fieldwork costs. Substitutions were more strictly limited to avoid undermining the probabilistic design (Fowler 2009), and we followed recommendations to substitute with a near neighbor (Stopher 2012). If still unsuccessful, no additional substitutions were permitted, meaning that segments with fewer than six completed interviews were possible so long as either both original and substitution attempts resulted in refusals or the maximum number of attempts was exhausted.

A reputable local survey firm carried out the interviews. To the extent possible given work schedules, the same interviewers were assigned to each segment within any given FM–LB pair – that is, an interviewer assigned to a given FM segment was also assigned to the corresponding LB segment. Our review of location markers confirmed that interviewers were following protocols for the selection of households but also revealed an unanticipated deviation. This deviation occurred because of incorrect spatial geolocations for the selected FM segments. Due to a miscommunication with the fieldwork team, those spatial geolocations were catalogued using a coding scheme that matched the prior census (conducted in 2000) rather than the 2011 census. The LB segments were selected using the more recent census and, because of updates in the intervening years, not all these segments were contiguous. This deviation is important to note but is orthogonal to the objective of the study. Assignment was still random. Moreover, we find no significant demographic differences between LB segments that are closer to versus farther away from the originally planned locations.

Interviewers were aware that different approaches were being used but were blind to the study's purpose. All interviewers were trained by the research team, and their office personnel were also trained in a rigorous quality control program. Computer-assisted personal interviewing permitted capturing dozens of quality control markers including location, voice recordings, and time stamps. To ensure interviewers followed all instructions and as part of our standard approach to quality control, every interview was audited by the local team and approximately one quarter were re-audited by the research team; in all, 5.1% of interviews were canceled, mostly due to errors reading questions, and each was immediately replaced with interviews from the same segment with the same protocol as the original.

4.2. Weighting and Variance Estimation

We developed the calibration weights by raking over known population distributions of gender and age categories (18–30, 31–45, 46+). The population distribution was taken from the 2011 census, the most recent one at the time when the study was implemented. There is no weighting in addition to the demographic one: prior to calibration, observations are treated as if coming from simple random sampling, where each unit has an equal probability of selection. We should note that the design is not simple random sampling but effectively a four-stage clustered design with municipalities as the PSUs, census segments as the SSUs, households as the tertiary sampling units (TSUs), and individuals within households as the elementary units of interest.

To account for the sample design, we use Taylor-linearized variance estimation, the most common method of computing standard errors of regression coefficients for complex survey data (Woodruff 1971; Fuller 2009). This method uses the Taylor expansion to obtain a first-order (tangent line) approximation for the estimator, and the variance of the estimator itself is calculated as the variance of this linear approximation. The Taylor series method estimates the overall variance by combining variances among sampling units with stratum-level variance. We use a relatively simple design with no stratification, thus making Taylor linearization a suitable method of variance estimation. Our analysis is done in Stata 19.0 (StataCorp 2025).

5. Comparing Methods of Selection

We assess outcomes across the FM and LB selection methods according to the above-identified sets of dependent variables that address efficiency and economy, representativeness, and the nature (personality, political attitudes, behaviors) of the sampled individuals.

5.1. Efficiency and Economy

We begin by documenting key indicators of survey fieldwork efficiency for the FM and LB selection methods: numbers of completed interviews, response rates, and fieldwork time and cost. The comparisons are presented in Table 2. Importantly, the FM method results in more than twice as many completed interviews as the LB condition: 451 vs. 220. However, the response rates (AAPOR RR1) for the two methods are similar. The higher number of completed cases for the FM method is achieved by making many more interview attempts: permitting interviewers to continue to the next house until they find someone at home substantially increases efficiency in reaching the targeted completions (raw numbers of obtained interviews; response rates reflect the proportion of completed cases out of all attempts). It is necessary to note that an interviewer could make more than one initial attempt on the same household, so that the number of initial attempts was greater than the number of contacted households. In contrast, the LB method does not achieve the realized sample of 450 because our strict deployment of this approach limited substitutions (else it becomes effectively equivalent to the type of nonprobabilistic approach used in the other experimental condition; see Fowler 2009).

Consistent with what prior studies suggest will occur, implementing the LB selection method substantially increased fieldwork time: fieldwork was completed in 29 days for the FM condition and 49 days with the LB method. These differences are even more stark when we calculate time per completed interview: the FM approach (0.06 days per interview) was far more time-efficient than the LB method (0.22 days per interview). Given that the FM method is more commonly deployed, it could be that the team had comparatively more experience with this approach and that translated into some efficiencies. However, we surmise that the difference is likely due to the time interviewers in the LB condition spent on repeated contact attempts.

Table 2. Comparison of response rates and fieldwork time

	FM	LB
Interviews		
Total initial attempts	2,435	1,213
Refusals and breakoffs	435	406
Ineligible	125	1
Unknown eligibility	1,177	481
Other incomplete cases	247	105
Completed cases	451	220
AAPOR RR1	19.6%	18.2%
Fieldwork time (days)		
Total	29	49
Average per completed interview	0.06	0.22

Note. Fieldwork time is defined as the number of days elapsed from the first to the last day of data collection.

Analyzing interview duration confirms this conjecture. We compare the amount of time interviewers spent with respondents (net time) to the amount of time that passed from the first contact attempt until the interview was completed (gross time) accounting for complex survey design. Net interview times are approximately the same across the two conditions: 0.76 hours in the FM condition and 0.72 hours in the LB condition ($\Delta = 0.03$, $p = .189$). However, gross interview times are very different: 0.92 hours in the FM condition and 42.4 hours in the LB condition ($\Delta = 41.4$, $p = .003$). These estimates account for complex survey design. Of course, interviewers in the LB condition could complete other interviews between contact attempts, so these figures may understate their efficiency. Still, these results strongly suggest that the difference in time efficiency between the FM and LB conditions is primarily due to interviewers in the LB condition spending time contacting selected respondents.

Given these results, we also investigate the effects of contact attempts on both efficiency and representativeness in the LB condition. We find that the first contacts were generally successful: 68.2% of all completed interviews needed only one attempt whereas 83.6% needed no more than two. At the same time, imposing limits on the number of contact attempts does not seem to radically increase time efficiency. The gross interview time was 32.9 hours even with only one contact attempt – better than the LB condition average but still substantially worse than the FM condition average.

The result of additional fieldwork time in the LB condition is that survey costs increase. We do not calculate the precise difference in cost between the two methods because we do not have detailed information on the pay scale; however, an estimate provided by the local survey firm that implemented the fieldwork informs us that each interview in the LB condition was over four times more expensive than an interview using the FM approach. Overall, the extent to which the LB approach increases fieldwork costs is likely conditioned by the context and design choices (e.g., costs can potentially be reduced by requiring fewer callbacks).

5.2. Representativeness of sample

How do these two approaches compare in terms of the representativeness of the samples they produce? We compare the LB and FM conditions by contrasting both unweighted and weighted sample means to the available census data. In assessing the samples, we examine self-reported years of formal education and wealth (using a measure based on a battery of items on whether the household possesses certain essentials, a common measure of socioeconomic status in developing contexts; see Filmer and Pritchett 2001). We use five items to form the wealth index: internet access, computer (either a desktop or a laptop), at least one cell phone, car, and flatscreen TV. We calculated the additive total of these indicators, so that wealth ranges from zero (household has none of the items) to five (all items).

Table 3 presents descriptive data on the samples that the FM and LB methods produced. Note that the census data used for this analysis includes only those GMASJ municipalities that were also included in the survey sampling frame. By design, the FM sample is extremely close to the census benchmarks on age and gender without weights. As anticipated by prior research, the LB sample underrepresents men (approximately 42% vs. 48% in the census) and younger people (mean age of 47.6 years vs. about 40 in the census), requiring weights to make the quasi-probabilistic sample data representative of the population with respect to age and gender. These numbers are the same for the originally sampled and substituted units in the LB condition. The samples are close to the census benchmark on education, with or without weights. Also, the LB sample shows somewhat higher dispersions (as measured by observed standard deviations) for all variables except gender.

Table 3. Comparison of the FM and LB samples' representativeness

	Unweighted		Weighted		2011
	FM	LB	FM	LB	Census
Female					
Mean	0.501	0.582	0.522	0.522	0.522
SD	0.501	0.494	0.501	0.494	
Abs. dif. from census	0.021	0.060	< 0.001	< 0.001	
Age					
Mean	40.8	47.6	40.3	41.1	39.8
SD	15.3	20.5	15.3	20.5	
Abs. dif. from census	1.0	7.8	0.5	1.3	
Education					
Mean	10.7	10.7	10.7	10.7	10.3
SD	4.4	4.8	4.4	4.8	
Abs. dif. from census	0.4	0.4	0.4	0.4	
Wealth index					
Mean	3.80	3.60	3.81	3.65	2.92
SD	1.26	1.40	1.26	1.40	
Abs. dif. from census	0.88	0.68	0.89	0.73	

Note. SD = standard deviation.

To explore potential differences in the demographic variables beyond means, we also analyzed observed distributions of age, education, and income in the FM and LB samples as well as in the census (see Figure 1). The distribution of age in the FM sample is very close to the census, whereas all other distributions are substantially different. The patterns of deviations from the census are similar for the two samples. For instance, both FM and LB methods seem to overrepresent individuals with average education levels, even though the resulting mean estimates are almost equal to the one from the census.

[Figure 1 about here]

We note that the households surveyed in both the LB and FM samples are more affluent than the corresponding census data would suggest. We cannot say for sure what may have generated that outcome, but there are two probable causes. One is that it results from less affluent individuals being less willing or available to respond to surveys. This tendency to over-sample individuals higher in socioeconomic status has been documented in the United Kingdom (Sturgis et al. 2018) and in the United States (Kennedy et al. 2018), among others. Another possibility is an overall increase in wealth in the region from 2011 (the census year) to 2018 (when our data were collected). Evidence from LAPOP's AmericasBarometer shows that, on the same 0–5 index, average household wealth in San José increased from 2.38 in 2010 and 2012 to 3.73 in 2018 ($\Delta = 1.35$, $p < .001$). Of course, these are survey data, so we cannot be sure that true increases in wealth are driving the difference between our samples and the 2011 census. Still, when it comes to basic demographics, both the LB (although only with weighting) and FM methods are comparable in their capacity to produce representative samples.

To further extend the analyses related to representativeness, we compare the two samples in terms of household size, number of young children in the house, marital status, and employment status; prior research suggests that such variables may be particularly sensitive to within-household selection methods (Smyth, Olson, and Stange 2019). While none of the differences are significant (see Table 4), estimates suggest that the LB method is somewhat less likely to reach larger households and households with children under 13 than the FM method. However, more statistical power would be needed to capture this difference.

Table 4. Comparison of the FM and LB samples by some household and individual characteristics

	FM	LB	Absolute difference
Household characteristics			
Size	3.95	3.60	0.35
			(0.19)
Children under 13	0.78	0.60	0.19
			(0.11)
Respondent's characteristics			
Partnered	0.47	0.47	< 0.01
			(0.05)
Employed	0.50	0.50	< 0.01
			(0.04)

Note. Weighted estimates. Standard errors in parentheses. None of the estimated differences are statistically significant (all *p*-values greater than .05).

5.3. Personality Indicators, Political Attitudes, and Behaviors

The above results align with expectations that we derive from prior research. Our more novel test is whether the sampled respondents have different personality dispositions, biasing estimates on attitudinal measures. To investigate this, we assess differences in key psychological indicators across the two sampling methods. In measuring personality traits, we draw on the Big Five model that has been shown to impact survey response style and propensity (Hibbing et al. 2019; Valentino et al. 2020). We also compare the samples on political attitudes and behaviors that, according to the literature, are particularly important to the quality and endurance of democratic governance. This decision reflects our own area of expertise, in the field of comparative public opinion, which dictated the contents of the questionnaire.

The first of the political measures is political participation (e.g., Verba, Schlozman, and Brady 1995): self-reported voting and a composite index for other forms of civic engagement; indices are calculated as the simple average of constituent items. Second, symbolic ideology helps citizens organize their political preferences (Knutsen 1997), making it interesting to see whether they differ across the two samples. Third, we estimate differences in political orientations that many believe are conducive to liberal-democratic regimes (Almond and Verba 1963; Inglehart and Welzel 2005; Lipset 1959; Putnam 1993). Specifically, we compare support for democracy (Mattes 2018), political trust (Levi and Stoker 2000), and political tolerance (Sullivan and Transue 1999). Following the common practice in comparative political research, including in Latin America (e.g., Morris and Klesner 2010), we calculate the trust index using only explicitly political or government institutions. See Table 5 for the descriptions of the items.

Table 5. Descriptions of the personality and political variables

Variable	Description
Big Five personality traits	Ten-item personality inventory (TIPI; Gosling, Rentfrow, and Swann 2003)
Voted last election	Respondent voted in the last presidential election
Political participation index	During the last campaign, respondent (1) put up a party flag, (2) used a sticker, (3) did door-knocking, (4) donated to a candidate, (5) attended a political meeting
Ideology	Scale from 1 = *Left* to 10 = *Right*
Support for democracy	Respondent agrees that democracy is better than other forms of government
Political trust index	Respondent trusts (1) the Congress, (2) the police, (3) political parties, (4) the President, (5) the Supreme Court, (6) the Fourth Chamber [of the Supreme Court], (7) the local government, (8) elections
Political tolerance index	Respondent thinks that political dissidents should be allowed to (1) vote, (2) conduct peaceful demonstrations, (3) run for public office, (4) appear on television

Table 6 presents the results, with all the variables normalized to the same 0–100 scale for ease of comparison. These estimates use calibration weights for both FM and LB samples to account for the demographic skew in the latter. Overall, we find no significant differences between the two samples on any of the analyzed indicators; these nulls effects persist when we control for interviewer effects (i.e., when interviewer IDs are added to the regression models as indicator variables). Conclusions based on this analysis should take into consideration that the size of the standard errors is affected by the sample size. Yet, even still, the differences are extremely small in terms of magnitudes: the largest absolute difference found, the one on the Big Five trait conscientiousness, is only −2.06 units within the possible range from −100 to 100.

Table 6. Comparison of personality traits and political variables

	FM	LB	Absolute difference
Big Five personality traits			
Openness to experience	75.0	75.1	0.14
			(1.99)
Conscientiousness	79.4	77.4	2.06
			(1.57)
Extraversion	73.6	73.6	0.04
			(1.84)
Agreeableness	67.5	69.1	1.60
			(2.55)
Emotional stability	65.2	64.2	1.02
			(1.95)
Political variables			
Participation			
Voted last election	72.9	71.2	1.76
			(5.01)
Political participation index	10.5	9.5	0.96
			(1.46)
Ideology (right)	48.7	49.6	0.98
			(2.88)
Pro-democratic values			
Support for democracy	74.4	73.6	0.83
			(2.66)
Political trust index	47.2	47.6	0.33
			(1.55)
Political tolerance index	54.6	54.4	0.24
			(2.84)

Note. Weighted estimates. Standard errors in parentheses. All variables normalized to 0–100 scale. None of the estimated differences are statistically significant (all *p*-values greater than .05).

6. Conclusion

Area probability samples using probabilistic or quasi-probabilistic approaches to selecting individuals within households are difficult to draw. In addition to the omnipresent challenge of high nonresponse rates, such studies run up against the lack of registers with which to select and pre-contact households. They also confront elevated expenses and security risk associated with maintaining teams in the same location long enough to make numerous in-person recontacts. It is critical that survey practitioners and data users take these constraints seriously. When researchers insist on practices that are optimal in theory, but excessively challenging in practice, the actual methods applied in the field may be inconsistent and irreproducible if enumeration teams deviate from protocols and, worse, obscure their work (AAPOR/WAPOR 2021).

Most survey researchers working on major social science projects in developing contexts recognize this challenge, and many use some form of matching as a practical recourse. Yet, this choice has been something of a leap of faith, with no prior study systematically putting the question to a test: does this deviation from the theoretical ideal impact outcomes? Our study has been designed to offer one lens through which we can directly address this question. Our conclusions are based on this one study and its design choices; thus, our study's results are not definitive but, rather, a contribution to what we hope will be ongoing efforts to uncover the empirical challenges and tradeoffs that accompany different sampling approaches used in practice across general population surveys in developing contexts.

Practically speaking, no survey sample is perfect. For instance, both samples have relatively low response rates (just under 20%). In that sense, samples drawn in Costa Rica seem to be subject to the same types of challenges present in surveys around the world (Curtin, Presser, and Singer 2005; Singer 2006; Smith 1995; Williams and Brick 2018). And yet, our results show that the FM approach is not only more efficient in time and money but also produces a sample that is comparable to what one can gather when applying a quasi-probabilistic approach to within-household selection in combination with weights. Importantly, similarities between the FM and LB samples in our experiment could have been driven by their adjacent geographic locations and commonalities in individuals who ultimately agreed to participate.

However, if these factors are indeed more important than the household selection method, it only reinforces our main conclusion: for practical purposes of survey research in developing countries, the FM approach performs comparatively well. Importantly, despite the greater capacity for interviewer discretion and for self-selection into the survey by more cooperative individuals, our results reveal no such biases, at least on the variables we observe, across the two selection methods.

Our experiment is only one study, deploying one set of protocols, and cannot support the conclusion that such deviation from probabilistic or quasi-probabilistic practices will always yield similar outcomes. Further, our design is not without limitations. First, we did not include a true probabilistic method to select respondents within households, such as the Kish grid. Such methods are rarely used in developing contexts because they are known to increase fieldwork time and reduce cooperation (Battaglia et al. 2008; Binson et al. 2000; O'Rourke and Blair 1983; Salmon and Nichols 1983). This means we cannot know whether a probabilistic selection method might have yielded different results.

Second, our selection mechanism at the household level was not perfectly random. Substitution of non-responding households was allowed, meaning that the size of the gross sample was not defined in advance. Practically, the final gross sample in the FM condition turned out to be twice as large as the gross sample in the LB condition. Therefore, differences between the two conditions cannot unambiguously be attributed to within-household selection methods alone. Implementing our experiment more perfectly would have meant constructing a sample of households and then, randomly assigning each household in this predefined sample to either the FM or LB condition. However, implementing this design in a country like Costa Rica would have been prohibitively costly. Such an experiment would also have deviated substantially from the practices employed by field researchers and harmed the ability to compare costs, which was a relevant part of the experiment.

Third, the probability of selecting a census segment for the LB sample in our study was conditional on what segment was selected for the FM sample. While this design deviates from perfect random assignment on the segment level, we believe there were good practical reasons for implementing it. The chosen design allowed us to maintain reasonable survey costs while maximizing the closeness between FM and LB segments. As a result, any differences between the two samples can be attributed to the household selection methods rather than to differences between census segments located too far from one another.

Even with these limitations in mind, we draw two conclusions. First, it is critical that both theory and practice guide survey design. While theory can tell which methods are preferable in the abstract, the tradeoffs presented in practice must also direct choices in the field. Second, in practice, matching approaches to within-household sampling can yield shorter fieldwork times, lower costs, and samples comparable to those generated via quasi-probabilistic methods. It is these gains that have attracted many to the approach, leaving open the question of whether these gains come at the cost of very different types of individuals recruited into the survey that would be achieved via a more costly quasi-probability within-household selection method. We underscore that this is not the definitive answer, but we nonetheless offer as one data point this observation: across a deployment of the two methods that have been designed to permit as much opportunity for personality and related attitudinal differences to emerge, we find scant evidence of differences across the realized samples. While researchers in developing contexts ought to strive toward the gold standard, they may not leave all that much behind when accepting a practical solution to genuine resource constraints.

References

AAPOR/WAPOR. 2021. "Task Force Report on Quality in Comparative Surveys." https://wapor.org/resources/aapor-wapor-task-force-report-on-quality-in-comparative-surveys/

Afrobarometer. n.d. "Sampling." Accessed September 26, 2025. https://www.afrobarometer.org/surveys-and-methods/sampling/

AmericasBarometer. n.d. "Sample Design." Accessed September 26, 2025. https://www.vanderbilt.edu/lapop/survey-designs.php

Arab Barometer. n.d. "Methodology." Accessed September 26, 2025. https://www.arabbarometer.org/survey-data/methodology/

Almond, Gabriel A., and Sidney Verba. 1963. *The Civic Culture: Political Attitudes and Democracy in Five Nations*. Princeton University Press.

Battaglia, Michael P., Michael W. Link, Martin R. Frankel, Larry Osborn, and Ali H. Mokdad. 2008. "An Evaluation of Respondent Selection Methods for Household Mail Surveys." *Public Opinion Quarterly* 72 (3): 459–69.

Binson, Diane, Jesse A. Canchola, and Joseph A. Catania. 2000. "Random Selection in a National Telephone Survey: A Comparison of the Kish, Next-Birthday, and Last-Birthday Methods." *Journal of Official Statistics* 16 (1): 53–59.

Brick, J. Michael. 2011. "The Future of Survey Sampling." *Public Opinion Quarterly* 75 (5): 872–88.

Bryant, Barbara E. 1975. "Respondent Selection in a Time of Changing Household Composition." *Journal of Marketing Research* 12 (2): 129–35.

Christensen, Anne Illemann, Ola Ekholm, Charlotte Gluemer, and Knud Juel. 2014. "Effect of Survey Mode on Response Patterns: Comparison of Face-to-Face and Self-Administered Modes in Health Surveys." *European Journal of Public Health* 24 (2): 327–32.

Clark, Robert G., and David G. Steel. 2007. "Sampling within Households in Household Surveys." *Journal of the Royal Statistical Society. Series A: Statistics in Society* 170 (1): 63–82.

Cumming, Robert Graham. 1990. "Is Probabilistic Sampling Always Better? A Comparison of Results from a Quota and a Probabilistic Sample Survey." *Community Health Studies* 14 (2): 132–37.

Curtin, Richard, Stanley Presser, and Eleanor Singer. 2005. "Changes in Telephone Survey Nonresponses over the Past Quarter Century." *Public Opinion Quarterly* 69 (1): 87–98.

Czaja, Ronald, Johnny Blair, and Jutta P. Sebestik. 1982. "Respondent Selection in a Telephone Survey: A Comparison of Three Techniques." *Journal of Marketing Research* 19 (3): 381–85.

Filmer, Deon, and Lant H. Pritchett. 2001. "Estimating Wealth Effect without Expenditure Data or Tears: An Application to Educational Enrollments in States of India." *Demography, 38*(1): 115–32.

Fowler, Floyd J., Jr. 2009. *Survey Research Methods*. 4th ed. Sage.

Fuller, Wayne A. 2009. *Sampling Statistics*. Wiley.

Gaziano, Cecilie. 2005. "Comparative Analysis of Within-Household Respondent Selection Techniques." *Public Opinion Quarterly* 69 (1): 124–57.

Gaziano, Cecilie. 2008. "Last-Birthday Selection." In *Encyclopedia of Survey Research Methods*, edited by Paul J. Lavrakas, 417–18. Sage.

Gerber, Alan S., Gregory A. Huber, David Doherty, and Conor M. Dowling. 2011. "The Big Five Personality Traits in the Political Arena." *Annual Review of Political Science* 14: 265–87.

Gosling, Samuel D., Peter J. Rentfrow, and William B. Swann Jr. 2003. "A Very Brief Measure of the Big Five Personality Domains." *Journal of Research in Personality* 37 (6): 504–28.

Guignard, Romain, Jean-Louis Wilquin, Jean-Baptiste Richard, and François Beck. 2013. "Tobacco Smoking Surveillance: Is Quota Sampling an Efficient Tool for Monitoring National Trends? A Comparison with a Random Cross-Sectional Survey." *PLOS One* 8 (10): e78372. https://doi.org/10.1371/journal.pone.0078372

Hagan, Dan E., and Charlotte Meier Collier. 1983. "Must Respondent Selection Procedures for Telephone Surveys Be Invasive?" *Public Opinion Quarterly* 47 (4) 547–56.

Hibbing, Matthew V., Matthew Cawvey, Raman Deol, Andrew J. Bloeser, and Jeffery J. Mondak. 2019. "The Relationship between Personality and Response Patterns on Public Opinion Surveys: The Big Five, Extreme Response Style, and Acquiescence Response Style." *International Journal of Public Opinion Research* 31 (1): 161–77.

Inglehart, Ronald, and Christian Welzel. 2005. *Modernization, Cultural Change, and Democracy: The Human Development Sequence*. Cambridge University Press.

Jabkowski, Piotr. 2017. "A Meta-analysis of Within-Household Selection Impact on Survey Outcome Rates, Demographic Representation and Sample Quality in the European Social Survey." *Research & Methods* 26 (1): 31–60.

Kennedy, Courtney, Mark Blumenthal, Scott Clement, Joshua D Clinton, Claire Durand, Charles Franklin, Kyley McGeeney, Lee Miringoff, Kristen Olson, Douglas Rivers, Lydia Saad, G Evans Witt, and Christopher Wlezien. 2018. "An Evaluation of 2016 Election Polls in the United States." *Public Opinion Quarterly* 82 (1): 1–33.

Kish, Leslie. 1949. "A Procedure for Objective Respondent Selection within the Household." *Journal of the American Statistical Association* 44 (247): 380–87.

Kish, Leslie. 1967. *Survey Sampling*. 2nd ed. Wiley.

Knutsen, Oddbjørn. 1997. "The Partisan and the Value-Based Component of Left–Right Self-Placement: A Comparative Study." *International Political Science Review* 18 (2): 191–225.

Koch, Achim. 2018. "Within-Household Selection of Respondents." In *Advances in Comparative Survey Methodology: Multinational, Multiregional, and Multicultural Contexts*, edited by Timothy P. Johnson, Beth-Ellen Pennell, Ineke A. L. Stoop, and Brita Dorer, 139–57. Wiley.

Le, Kien T., J. Michael Brick, Abdoulaye Diop, and Darwish Alemadi. 2013. "Within-Household Sampling Conditioning on Household Size." *International Journal of Public Opinion Research* 25 (1): 108–18.

Le, Kien Trung, Abdoulaye Diop, Jill Wittrock, Darwish Al-Emadi, and Elmogiera Elawad. 2014. "Gender Prespecified Sampling for Cost Control." *International Journal of Public Opinion Research* 26 (4): 441–52.

Levi, Margaret, and Laura Stoker. 2000. "Political Trust and Trustworthiness." *Annual Review of Political Science* 3: 475–507.

Lipset, Seymour Martin. 1959. “Some Social Requisites of Democracy: Economic Development and Political Legitimacy.” *American Political Science Review* 53 (1): 69–105.

Loennqvist, Jan-Erik, Sampo Paunonen, Markku Verkasalo, Sointu Leikas, Annamari Tuulio-Henriksson, and Jouko Loennqvist. 2007. “Personality Characteristics of Research Volunteers.” *European Journal of Personality* 21 (8): 1017–30.

Lupica, Carina. 2024. “Women Caregivers: Pillars of an Aging Society.” *IADB Blog.* https://blogs.iadb.org/igualdad/en/women-caregivers-pillars-of-an-aging-society/

Lupu, Noam, and Kristin Michelitch. 2018. “Advances in Survey Methods for the Developing World.” *Annual Review of Political Science* 21: 195–214.

Marcus, Bernd, and Astrid Schütz. 2005. "Who are the People Reluctant to Participate in Research? Personality Correlates of Four Different Types of Nonresponse as Inferred from Self-and Observer Ratings." *Journal of personality* 73 (4): 959-984.

Martin, Elizabeth. 1999. “Who Knows Who Lives Here? Within-Household Disagreements as a Source of Survey Coverage Error.” *Public Opinion Quarterly* 63 (2): 220–36.

Mattes, Robert. 2018. “Support for Democracy.” In *Oxford Research Encyclopedia of Politics*. https://doi.org/10.1093/acrefore/9780190228637.013.622

McBurney, Peter. 1988. “On Transferring Statistical Techniques across Cultures: The Kish Grid.” *Current Anthropology* 29 (2): 323–25.

Morris, Stephen D., and Joseph L. Klesner. 2010. “Corruption and Trust: Theoretical Considerations and Evidence from Mexico.” *Comparative Political Studies* 43 (10): 1258–85.

Nemeth, Renata. 2004. “Representativeness Problems Inherent in Address-Based Sampling and a Modification of the Leslie Kish Grid.” *Bulletin of Sociological Methodology* 83: 43–60.

Oldendick, Robert W., George F. Bishop, Susan B. Sorenson, and Alfred J. Tuchfarber. 1988. “A Comparison of the Kish and Last Birthday Methods of Respondent Selection in Telephone Surveys.” *Journal of Official Statistics* 4 (4): 307–18.

Olson, Kristen, Mathew Stange, and Jolene Smyth. 2014. “Assessing Within-Household Selection Methods in Household Mail Surveys.” *Public Opinion Quarterly* 78 (3): 656–78.

Olson, Kristen, and Jolene D. Smyth. 2017. "Within-household Selection in Mail Surveys: Explicit Questions are Better than Cover Letter Instructions." *Public Opinion Quarterly* 81 (3): 688-713.

O'Rourke, Diane, and Johnny Blair. 1983. “Improving Random Respondent Selection in Telephone Surveys.” *Journal of Marketing Research* 20 (4): 428–32.

Paisley, William J., and Edwin B. Parker. 1965. “A Computer-Generated Sampling Table for Selecting Respondents within Households.” *Public Opinion Quarterly* 29 (3): 431–36.

Putnam, Robert D. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton University Press.

Salmon, Charles T., and John Spicer Nichols. 1983. “The Next-Birthday Method of Respondent Selection.” *Public Opinion Quarterly* 47 (2): 270–76.

Singer, Eleanor. 2006. “Introduction: Nonresponse Bias in Household Surveys.” *Public Opinion Quarterly* 70 (5): 637–45.

Smith, Tom W. 1995. “Trends in Non-response Rates.” *International Journal of Public Opinion Research* 7(2): 157–71.

Smyth, Jolene D., Kristen Olson, and Mathew Stange. 2019. “Within-Household Selection Methods: A Critical Review and Experimental Examination.” In *Experimental Methods in Survey Research: Techniques that Combine Random Sampling with Random Assignment*, edited by Paul Lavrakas, Michael Traugott, Courtney Kennedy, Allyson Holbrook, Edith de Leeuw, and Brady West, 23–45. Wiley.

StataCorp. 2025. *Stata Statistical Software: Release 18*. StataCorp LLC.

Stopher, Peter. 2012. *Collecting, Managing, and Assessing Data Using Sample Surveys.* Cambridge University Press.

Sturgis, Patrick, Jouni Kuha, Nick Baker, Mario Callegaro, Stephen Fisher, Jane Green, Will Jennings, Benjamin E. Lauderdale, and Patten Smith. 2018. “An Assessment of the Causes of the Errors in the 2015 UK General Election Opinion Polls.” *Journal of the Royal Statistical Society. Series A: Statistics in Society* 181 (3): 757–81.

Sullivan, John L., and John E. Transue. 1999. “The Psychological Underpinnings of Democracy: A Selective Review of Research on Political Tolerance, Interpersonal Trust, and Social Capital.” *Annual Review of Psychology* 50: 625–50.

Tourangeau, Roger, Gary Shapiro, Anne Kearney, and Lawrence Ernst. 1997. “Who Lives Here? Survey Undercoverage and Household Roster Questions.” *Journal of Official Statistics* 13 (1): 1–18.

Valentino, Nicholas A., Kirill Zhirkov, D. Sunshine Hillygus, and Brian Guay. 2020. “The Consequences of Personality Biases in Online Panels for Measuring Public Opinion.” *Public Opinion Quarterly* 84 (2): 446–68.

Verba, Sidney, Kay Lehman Schlozman, and Henry E. Brady. 1995. *Voice and Equality: Civic Voluntarism in American Politics*. Harvard University Press.

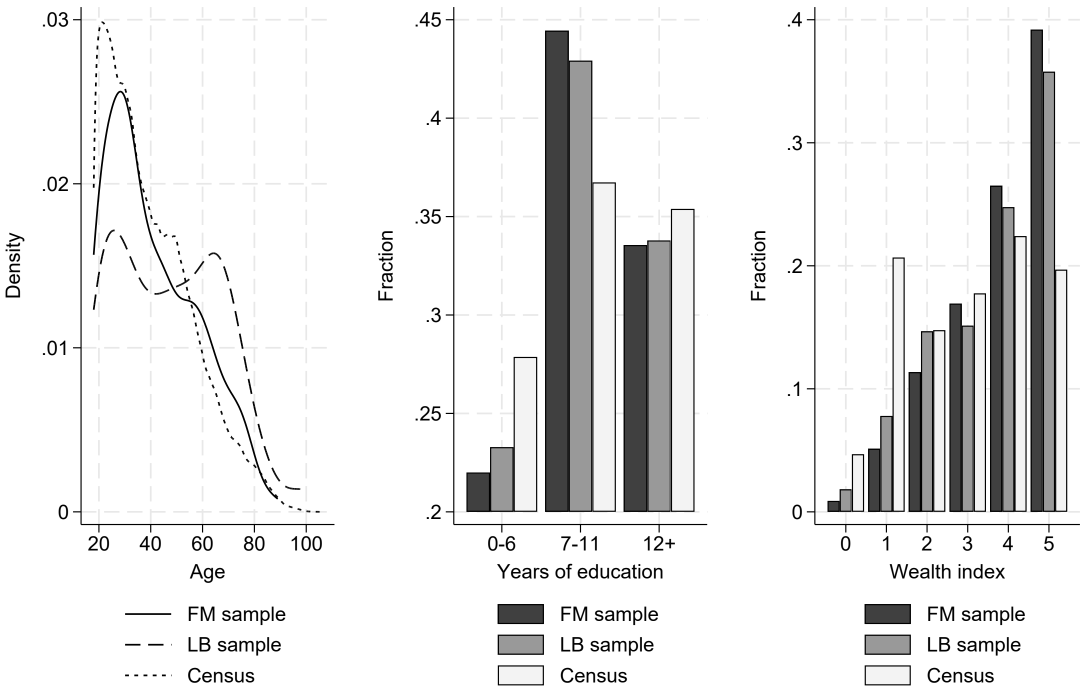
Williams, Douglas, and J. Michael Brick. 2018. "Trends in U.S. Face-To-Face Household Survey Nonresponse and Level of Effort." *Journal of Survey Statistics and Methodology* 6 (2): 186–211.

Woodruff, Ralph S. 1971. "A Simple Method for Approximating the Variance of a Complicated Estimate." *Journal of the American Statistical Association* 66 (334): 411–14.

Yan, Ting, Roger Tourangeau, and Rose McAloon. 2009. A Meta-analysis of Within-Household Respondent Selection Methods. *Proceedings of the Survey Research Methods Section 17, American Statistical Association*. http://www.asasrms.org/Proceedings/y2009/Files/400064.pdf

---- Figure 1.pdf ----

Figure 1. Distributions of age, education, and wealth in the FM and LB samples vs. the census



Density
.03
.02
.01
0
20
40
60
80
100
Age
FM sample
LB sample
Census
Fraction
.45
.4
.35
.3
.25
.2
0-6
7-11
12+
Years of education
FM sample
LB sample
Census
Fraction
.4
.3
.2
.1
0
0
1
2
3
4
5
Wealth index
FM sample
LB sample
Census