

# Comparing Methods for Sampling Individuals within Households:

## A Field Experiment in Costa Rica<sup>\*</sup>

Noam Lupu  
Vanderbilt University  
noam.lupu@vanderbilt.edu

J. Daniel Montalvo  
Vanderbilt University  
daniel.montalvo@vanderbilt.edu

Mitchell A. Seligson  
Vanderbilt University  
mitchell.a.seligson@vanderbilt.edu

Elizabeth J. Zechmeister  
Vanderbilt University  
liz.zechmeister@vanderbilt.edu

Kirill Zhirkov  
Vanderbilt University  
kirill.zhirkov@vanderbilt.edu

January 11, 2021

---

<sup>\*</sup> 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

The survey sampling ideal mandates probabilistic or quasi-probabilistic selection methods at all stages of a multistage area sample. In theory, this approach should yield superior samples. In practice, various constraints lead many probability-based surveys to apply a non-probabilistic approach at the final stage: the within-household selection of individuals. We report on an original field experiment that directly compares a quasi-probabilistic selection method (*last birthday*) to a nonprobabilistic approach that selects respondents according to known parameters on age and gender (*frequency matching*). We evaluate three types of outcomes: efficiency and economy of fieldwork, success in recruiting a representative sample, and differences in estimates of public opinion. We find that the frequency-matching approach performed better on the first dimension and no different than the last birthday method on the second and third. We conclude that selection methods that appear suboptimal in theory may yield better outcomes in practice.

In conventional public opinion surveys, researchers follow a probabilistic approach down to the selection of households, whether via a physical address (face-to-face, mail) or telephone (e.g., RDD).<sup>1</sup> At that point, researchers need a method to select survey respondents from within the household.<sup>2</sup> That decision is important as it shapes subsequent estimates (Ziniel 2008).<sup>3</sup> In theory, a probability-based approach should be carried through to this final stage. This is because meaningful sampling errors can be calculated when each eligible person in each household has a known and nonzero probability of being included in the sample. In addition, arbitrary (nonrandom) selection of respondents within households can lead to significant biases if people who are first or most willing to respond are systematically different (e.g., more cooperative) compared to the rest of the population.

Probability-based selection of individuals within households sampled for a survey is a well-established norm, in theory. In practice, however, probabilistic sampling at the household level may not yield better outcomes (for a review, see Gaziano 2005). This is particularly likely when executing probability-based selection confronts practical constraints. These constraints often surface in face-to-face surveys fielded in developing country contexts: in addition to the standard challenge of high nonresponse rates, these projects face hurdles such as a lack of pre-existing registers of names and addresses and the cost of maintaining teams in rural hamlets to permit sufficient recontact attempts. For these reasons, many major survey projects in developing contexts apply nonprobabilistic methods in the selection of individuals within a household (Lupu and Michelitch 2018). Nevertheless, scholars have yet to compare probabilistic and

---

<sup>1</sup> As an exception, many online studies target individuals without consideration of their household.

<sup>2</sup> Individuals are key in public opinion studies, to the degree that researchers are more interested in the behaviors and attitudes of individuals (e.g., their vote choice) than in household-level variables (e.g., aggregate income).

<sup>3</sup> There are approaches in which more than one individual per household is interviewed (e.g., Clark and Steel 2007), but in this paper, we focus specifically on methods to select a single respondent within a household.

nonprobabilistic methods within face-to-face research using a strict experimental approach: random assignment of the method of within-household selection.<sup>4</sup>

To address this gap, we designed a largescale field experiment to directly compare probabilistic within-household selection of individuals to a nonprobabilistic within-household approach that selected individuals according to known population frequencies of gender and age. We carried out the experiment in the Greater Metropolitan Area of San José, the capital of Costa Rica. Within pairs of segments in the same survey, interviewers were instructed to apply in one segment a quasi-probabilistic, last-birthday method of selection within the household, with up to twenty recontacts. In the other segment, interviewers recruited participants applying a gender and age frequency-match approach without recontact.

We assess the results along three core dimensions: efficiency and economy of fieldwork, success in recruiting a representative sample, and differences in the public opinion estimates derived from each approach. In our selection of public opinion variables for the comparison, we focus specifically on two types: set of personality measures and a set of behaviors and orientations that, according to comparative public opinion research, are associated with stable and efficient democratic governance—that is, attitudes that are particularly important in a developing country context. After assessing the three dimensions, we conclude that the frequency matching method seems to be the optimal choice in a developing country context.

Our results highlight the interplay between the theoretical foundations of survey methods and the distortions that challenges in the real world can introduce. We demonstrate that these

---

<sup>4</sup> Guignard et al. (2013) compare random selection to quota selection of respondents across two phone surveys, but the sampling frames differ across the two studies.

distortions can be highly consequential for fieldwork efficiency and data quality, while methods that appear suboptimal in theory can yield better outcomes in practice.

### **Selection of Individuals within Households**

Probabilistic methods are widely considered the gold standard for within-household selection of respondents in household surveys. Kish (1949) pioneered an approach that involves, first, enumerating all the individuals living in a given household and, second, randomly selecting an interviewee from that list. Usually, the first adult contacted in the household is asked to provide this information and then an interviewer attempts to recruit the selected individual. The drawbacks of the Kish method are its length and intrusiveness: it increases both the time necessary to complete an interview and the probability that the gateway respondent refuses to cooperate because of the complexity of the process and concerns about sharing details about the household's composition (Battagila et al. 2008; Gaziano 2005; Jabkowski 2017; Yan 2009).

To overcome these problems, methodologists have developed approaches that maintain the arbitrary nature of respondent selection without enumerating all the individuals within the household.<sup>5</sup> Examples of this approach are the next- and last-birthday methods: the first person contacted is asked to name the individual with the upcoming or the most recent (last) birthday, who is automatically selected as the interviewee (Salmon and Nichols 1983). Technically, these are quasi-probabilistic rather than true probabilistic methods since births are distributed non-randomly and non-uniformly throughout the year (Gaziano 2008). Survey-methodological research shows that, compared to the Kish enumeration approach, next- and last-birthday methods yield little to no differences in substantive responses (Oldendick et al. 1988), while they

---

<sup>5</sup> There are also within-household selection methods that combine probabilistic and quasi-probabilistic elements (e.g., Le et al. 2013; Rizzo, Brick, and Park 2004).

simultaneously decrease survey length and intrusiveness (O'Rourke and Blair 1983), and result in fewer dropouts (Binson, Canchola, and Catania 2000). There is also evidence that these quasi-probabilistic methods produce response rates that are higher than more intrusive probabilistic approaches and similar to nonprobabilistic selection of any adult member of the household (Battaglia et al. 2008; Salmon and Nichols 1983). Of the practitioners fielding face-to-face surveys in developing contexts who were surveyed by Lupu and Michelitch (2018), only 8% reported using a full Kish table, whereas 22% reported using the last-birthday method.

Another large class of methods to select respondents within households is not based on probabilistic or quasi-probabilistic techniques. Instead, individuals are chosen on the basis of target demographic characteristics, usually age and gender population frequencies derived from the census. There are many different selection methods within this class (e.g., Bryant 1975; Hagan and Collier 1983; Paisley and Parker 1965), often designed to compensate for underrepresentation of young males in most probabilistic sampling approaches. These methods trade off the theorized advantages of randomly selecting respondents within households for lower fieldwork times,<sup>6</sup> higher cooperation rates, and less need for post-calibration weighting. In their survey of practitioners, Lupu and Michelitch (2018) find that 59% reported using a frequency-match approach for within-household selection of respondents.

In theory, the use of a non-probabilistic approach for the selection of individuals can generate samples that are biased toward respondents with certain personality traits. Researchers speculate, 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, there is a dearth of evidence that tests whether these theoretical intuitions transfer to

---

<sup>6</sup> And generally, more efficient fieldwork efforts, given that recontacts are eliminated.

empirical realities. If anything, existing research provides evidence that probabilistic methods for selecting respondents within households do not always perform better, and suggests that non-probabilistic methods can offer substantial cost savings with little to no added sampling bias (Cumming 1990; Czaja, Blair, and Sebestik 1982; Olson, Stange, and Smyth 2014).

Moreover, field research reveals that the theoretically superior probabilistic and quasi-probabilistic selection methods have their own implementation challenges. One of them is related to enumeration of household members: respondents often have trouble completing this task accurately, which can cause up to a third of respondents to be selected in error (Battaglia et al. 2008; Lavrakas, Bauman, and Merkle 1993; Lind, Link, and Oldendick 2000; Martin 1999; Tourangeau et al. 1997). This issue is even more acute in developing contexts where large extended families living in a single household are still common (McBurney 1988).<sup>7</sup> Further, many factors that Kish (1967: 398–401) himself acknowledged, such as larger households or multiple families per dwelling unit, systematically reduce the probability of selecting a poor person versus a rich one when the commonly-used norm of one interview per household is adhered to (see also Nemeth 2004).<sup>8</sup> In addition, even with multiple recontacts, samples drawn in this way can end up with notable imbalances on demographic traits such as gender (Jabkowski 2017).

In other words, both approaches are imperfect when it comes to how they perform in practice, leaving open the question of whether the theoretical advantages of probabilistic and

---

<sup>7</sup> Cultural differences can also complicate implementation of probabilistic methods. For instance, in the Middle East, researchers have to match the genders of interviewer and respondent, making random selection impractical (Le et al. 2014).

<sup>8</sup> Kish (1967: 399) notes that, “the proper fractional representation of each adult is approximated closely with only the eight tables.”

quasi-probabilistic approaches translate into comparatively better outcomes than non-probabilistic methods for within-household selection in the field.

### **Expectations**

This discussion yields a set of expectations regarding outcomes from non-probabilistic (e.g., frequency matching) and quasi-probabilistic (e.g., last-birthday) methods of respondent selection within households:

1. Compared to a probabilistic method of within-household selection, the frequency-match method yields lower fieldwork time, lower per-interview cost, and higher cooperation rates.
2. Compared to a probabilistic method, the frequency-match method produces a sample that is closer to the census-based population on various demographic and socioeconomic status measures.
3. Compared to individuals selected using a probabilistic or quasi-probabilistic method, respondents selected using a frequency-match method will differ in some of their personality traits (e.g., on traits related to cooperativeness).
4. There are few differences between public opinion estimates generated based on data gathered via methods that use frequency-match versus probabilistic methods for within-household selection.

### **Study Design**

Our objective is to test whether, in practice, a face-to-face survey yields different results when the within-household selection of respondents uses a quasi-probabilistic approach versus a frequency-matched non-probabilistic approach. Our focus is on research in developing contexts, where non-probabilistic within-household approaches are common (Lupu and Michelitch 2018).

With that in mind, we implemented a large-scale field experiment as part of a survey study in Costa Rica, a middle-income country in Central America.<sup>9</sup> We selected Costa Rica because it is a country that has detailed census information and a strong capacity for high-quality survey research. In addition, we have years of experience and an extensive network of research partners in the country.

The study was administered in September–October 2018 as an area probability sample of 900 adults within the voting age-population of the Greater Metropolitan Area of San José (GMASJ).<sup>10</sup> The core of the field experiment is random assignment of census segments to one of two *within-household* sampling methods: (1) a frequency-match 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 is a quota-based approach, but must be distinguished from quota-based approaches that sample from “flow points” (e.g., shopping malls); rather, in this study the sample is probabilistic down to the household, at which point individuals are selected to match gender and age frequencies in the population. In what follows, we refer to this as the Frequency Match (FM) approach. The second method, which selects individuals using a quasi-probabilistic approach, is widely considered a reliable substitute for the gold-standard Kish method. In what follows, we refer to this method as the Last Birthday (LB) approach.

The sampling frame was the list of municipalities (*cantones*), districts (*distritos*), census segments, and maps for the metropolitan San José region in Costa Rica from the 2011 census

---

<sup>9</sup> In 2019, the average GDP per capita for the 17 Latin American countries for which the World Bank has data (Venezuelan data is not available) was \$8,105 USD; in Costa Rica this figure was \$12,238.

<sup>10</sup> As is standard in public opinion research, the sample excludes people in institutionalized settings: boarding schools, hospitals, police academies, military barracks, and inmates of the country’s jails. 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 had IRB approval from the authors’ institution.

provided by the Centro Centroamericano de Población. Within the GMASJ, a municipality was a primary sampling unit (PSU). 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. 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.<sup>11</sup>

In the next stage, we selected census segments with PPeS, after an urban/rural stratification. These segments constituted the secondary sampling unit (SSU). 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 SSU level. We determined the FM segments first. To do so, we selected at random 75 census segments from the 7,703 segments in the selected municipalities. We then fixed the number of households within each segment to six.<sup>12</sup> Households were selected using systematic selection, with enumerators beginning at the northeast corner of the segment, moving clockwise, skipping one house after each completed interview and selecting one respondent per household. To implement the frequency match, enumerators used a matrix of three age cohorts (18–29, 30–46, 47+) and gender (male/female) categories to select individuals so that the sample matched known frequencies on these characteristics as established by census data. The selected households were not revisited. If the

---

<sup>11</sup> San José was the only self-selected municipality. Due to its large size, San José was further subdivided into two PSUs.

<sup>12</sup> Fixing the number of households within segments allows us to have an evenly distributed and dispersed number of interviews per segment. For the study, a household is considered to be a group of people who eat their meals together. In apartment buildings, enumerators identify a starting point and select households on each floor, skipping one unit after each completed interview. If enumerators do not complete the interviews on one floor, they must continue on the next floor, either upstairs or downstairs.

enumerator could not secure an interview at a given household (due to refusals, ineligibility, or no one at home), the enumerator moved on to the next household.

We next selected the LB segments, using a probabilistic 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. We sorted these neighboring segments by size and discarded segments with an insufficient number of households from the pool of segment neighbors.<sup>13</sup> We then randomly selected one segment from the pool of neighbors for each FM segment, yielding a total of 75 LB segments. Table S1 in the Supplementary Material shows the number of neighboring segments by district and the final number of segments selected for the LB condition.

In the LB condition, enumerators were again instructed to complete six interviews in each segment and to designate the northeast corner as the starting point. In these cases, the enumerator approached the first house and asked to interview the adult with the most recent (last) birthday, thus establishing the conditions for quasi-probabilistic selection within the household. As in the FM approach, if successful, the enumerator was instructed to skip one residential unit before approaching the next one. If no one was home or the selected individual was not available, the enumerator 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 enumerator moved to the adjacent dwelling unit. The enumerator then repeated the same protocol at this dwelling unit. If still unsuccessful, no additional substitutions were

---

<sup>13</sup> The information on the size of each segment as well as the cartography came from the 2011 census. Segments varied substantially in terms of the number of households: while some segments had as few as four households, others had as many as 286. The decision to remove segments with an insufficient number of households was taken in order to avoid major obstacles during fieldwork. Twenty-one (out of the 214 neighbor segments) had fewer than 10 dwellings and were not included as candidates in the selection of segments for the LB condition.

permitted, meaning that segments with fewer than six completed interviews were possible so long as either both attempts resulted in refusals or the maximum number of attempts (i.e., twenty, counting both original and substitution) was exhausted.<sup>14</sup>

A highly regarded and experienced local survey firm, with whom we had worked several times before, carried out the interviews. In order to minimize interviewer effects, the same enumerators were assigned to each segment within any given FM–LB segment pair.<sup>15</sup> The entire enumeration crew was trained by the research team for fieldwork and their office personnel were also trained by the research team in a rigorous quality control program. The use of Computer Assisted Personal Interviewing (CAPI) permitted the capture of dozens of quality control markers including geo-location, voice recordings, and time stamps.<sup>16</sup> Every interview was audited in near-real time by the local team and approximately one quarter were re-audited by the authors’ research lab.<sup>17</sup>

### **Comparing Methods of Selection**

We start presenting our findings by documenting the key indicators of survey fieldwork efficiency for the FM and LB selection methods: numbers of complete interviews, response

---

<sup>14</sup> If additional substitutions were permitted, the method would depart from the intention of the design and move closer in practice to the non-probabilistic approach.

<sup>15</sup> Our review of location markers confirmed that enumerators were following protocols for the selection of households, but also revealed an unanticipated deviation in the selection of the segment. 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. See Table S2 in Supplementary Material.

<sup>16</sup> Interviewees were informed that parts of the interview would be recorded for quality control purposes.

<sup>17</sup> Based on this quality control process, a small number of interviews were canceled (5.1%), mostly because of errors in the reading of questions; canceled interviews were replaced in real time by sending the team back to the same segments with the assigned sampling protocol.

rates, and fieldwork time and cost. The corresponding comparisons are presented in Table 1. Importantly, the FM method resulted in more than twice as many complete interviews as the LB condition: 451 vs. 220. However, the response rates (AAPOR RR1) for the two methods are similar. In other words, the higher number of completes for the FM method is achieved by making many more interview attempts.<sup>18</sup>

**Table 1.** Comparison of response rates and fieldwork time (days)

	FM	LB
Interviews		
Total initial attempts	2,435	1,213
Refusals and breakoffs	435	406
Other incompletes	1,549	587
Completes	451	220
AAPOR RR1	19.6%	18.2%
Fieldwork time (days)		
Total	29	49
Average per completed interview	0.06	0.22

Consistent with what prior studies suggest will occur, implementing the LB selection method substantially increased fieldwork time. In our study, fieldwork was complete 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) turns out to be far more time-efficient than the LB method (0.22 days per interview), by a factor of nearly four.

The result of this additional fieldwork time is that survey costs also increase roughly in proportion. We do not calculate the precise difference in cost between the two methods because we do not have detailed information on the local firm’s pay scale; however, an estimate provided

---

<sup>18</sup> Interviewers in FM segments made more attempts because they could knock on many doors in order to find a cooperative respondent with matching age and gender. In LB segments, in contrast, interviewers made fewer attempts because they had to try repeatedly to contact the same households.

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.

How do these two approaches compare in terms of the representativeness of the samples they produce? We compare the LB and FM conditions by examining both unweighted and weighted sample means to the available census data. We developed the post-calibration weights by raking over known population distributions of gender, age categories (18-30, 31-45, 46+), education, and wealth for the GMASJ.<sup>19</sup> Education is measured in self-reported years of formal education. The wealth measure was constructed using 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). According to the results of a factor analysis, five items formed the corresponding index: having internet access, a computer (either a desktop or a laptop), at least one cell phone, a car, and an LCD TV.<sup>20</sup> We calculated the additive total of these five indicators, so that the measure can capture values from zero (household has none of the items) to 5 (all of the items).

Table 2 compares the samples that the two sampling methods produced. Note that the census data used for this analysis include 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 post-stratifying the sample. As anticipated by prior research, the LB approach yielded a sample that underrepresents men and younger individuals, requiring post-calibration weights to make the quasi-probabilistic sample data demographically representative of the population. Interestingly, the two conditions yielded

---

<sup>19</sup> These parameters were taken from the 2011 Costa Rica census, which we received from the partner firm in Costa Rica.

<sup>20</sup> See Table S3 in the Supplementary Material for the results of the factor analysis.

samples that are remarkably close to the census benchmark on education, with or without applying weights.

**Table 2.** Comparison of FM and LB sample representativeness

	Unweighted		Weighted		2011 Census
	FM	LB	FM	LB	
Female	0.501 (0.023)	0.582 (0.033)	0.522 (0.008)	0.522 (0.023)	0.522
Age	40.8 (0.82)	47.6 (1.28)	40.3 (0.40)	41.1 (1.73)	39.8
Education	10.7 (0.21)	10.7 (0.33)	10.7 (0.54)	10.7 (0.56)	10.3
Wealth index	3.80 (0.06)	3.60 (0.09)	3.81 (0.07)	3.65 (0.10)	2.92

*Note.* Standard errors in parentheses

It is also the case that households surveyed in both the LB and FM samples were more affluent than the corresponding census data would suggest. Even though we cannot say for sure what generated that outcome, there are two probable causes that are not mutually exclusive. One possibility 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 is a phenomenon that has been documented in the United Kingdom (Sturgis et al. 2018) and United States (AAPOR 2017), among others. Another possibility is an overall increase in wealth in the region from 2011 (when the census was conducted) to 2018 (when our data were collected). For instance, evidence from LAPOP’s AmericasBarometer data shows that, according to the same 0-5 index, average household wealth in the San José province of Costa Rica increased from 2.38 in the 2010 and 2012 rounds to 3.73 in the 2018/2019 round ( $\Delta = 1.35, p < .001$ ). Of course, these are still 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.

Even if the two approaches both yield samples that are demographically similar, theory suggests that the sampled respondents may have different personality dispositions, biasing the estimates on attitudinal measures. To test this expectation, we turn to estimating differences in key psychological and public opinion indicators across the two sampling methods. We start from personality traits, drawing on the Big Five model that has been shown to impact both survey response style and substantive preference estimates (Hibbing et al. 2019; Valentino et al. 2020). We also compare the two samples on political attitudes and behaviors that, according to the literature, are particularly important to the quality and endurance of democratic governance.

The first of them is political participation (e.g., Verba, Schlozman, and Brady 1995), for which we have both self-reported voting and a composite index for other forms of civic engagement.<sup>21</sup> Second, symbolic ideology is an important tool that helps voters organize their issue positions and partisan preferences (Knutsen 1997). Ideological labels in Latin America are known to depend on context (Zechmeister 2006), making it particularly interesting to see whether they differ across the two samples. Third, we estimate differences in political orientations that many scholars, from Lipset (1959) and Almond and Verba (1963) to Putnam (1993) and Inglehart and Welzel (2005), believe to be conducive to liberal-democratic regimes. Specifically, our comparison includes stated support for democracy as the preferred form of

---

<sup>21</sup> The indices are calculated as the simple average of the constituent items. Reliability analyses for the index variables are presented in Table S4 in the Supplementary Material.

government (Mattes 2018), political trust (Levi and Stoker 2000), and political tolerance (Sullivan and Transue 1999).<sup>22</sup> See Table 3 for the short descriptions of the survey items.

**Table 3.** Short 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 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 = <i>Left</i> to 10 = <i>Right</i>
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

*Note.* See Table S5 in the Supplementary Material for the exact formulations of all questions and responses.

Table 4 presents the results, with all the variables normalized to the same 0–100 scale for ease of comparison. These estimates use post-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 public opinion indicators. Note that the differences are also 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.

<sup>22</sup> 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. Reported results do not change if a broader index that includes trust in civic institutions like the media and the Catholic Church is used instead.

**Table 4.** Comparison of key public opinion estimates

	FM	LB	Difference
<i>Big Five personality traits</i>			
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)
<i>Political variables</i>			
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 at the 95% level.

## Conclusion

Area probability samples using probabilistic or quasi-probabilistic approaches to selecting individuals within households are exceptionally difficult to draw in developing contexts. 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, and they confront elevated expenses 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 investigators insist on practices that, in theory are optimal, but fieldworkers are unwilling or unable to comply in practice, the actual methods applied in the field may become inconsistent if enumeration teams deviate from protocols and, worse, obscure their applied methods.

Most survey researchers working on major social science projects in developing contexts recognize this challenge, and use some form of frequency matching as a result. As we noted earlier, Lupu and Michelitch (2018) find that 59% of investigators on such projects report using frequency matching, while very few use a probabilistic or quasi-probabilistic approach for the within-household selection of respondents. Yet, this choice has been something of a leap of faith, with no prior study systematically putting the question to a test: in an experimental design, does this deviation from the theoretical ideal produce subpar outcomes? Our study was designed to address this question directly.

Practically speaking, no survey sample is perfect. For instance, both samples have relatively low response rates (just under 20%). Regardless, samples drawn in Costa Rica seem to be subject to the same types of challenges that are present in surveys around the world (e.g., among many, Curtin, Presser, and Singer 2005; Singer 2006; Smith 1995). And, yet, our results show that a frequency-match 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 post-calibration weights.

We draw two conclusions. The first is that it is critical that both theory and practice guide researchers' choices when they design survey studies. The second is that, in practice, frequency matching approaches to within-household sampling can yield better outcomes, especially in developing countries: shorter fieldwork, lower costs, and samples comparable to those generated

via quasi-probabilistic methods. Randomization within the household can be extremely costly and time inefficient, with no detectable gains over other widely-used alternative field methods.

## References

- AAPOR. 2017. "An Evaluation of 2016 Election Polls in the U.S."  
<https://www.aapor.org/Education-Resources/Reports/An-Evaluation-of-2016-Election-Polls-in-the-U-S.aspx>
- Almond, Gabriel A., and Sidney Verba. 1963. *The Civic Culture: Political Attitudes and Democracy in Five Nations*. Princeton, NJ: 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.
- 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.
- 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, pp. 417–18. Thousand Oaks, CA: Sage.
- 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*. New York: 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.
- 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. New York: John Wiley & Sons.
- 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 (3MC)*, edited by Timothy P. Johnson, Beth-Ellen Pennell, Ineke A. L. Stoop, and Brita Dorer, 139–57. Hoboken, NJ: John Wiley & Sons.
- Lavrakas, Paul J., Sandra L. Bauman, and Daniel M. Merkle. 1993. "The Last-Birthday Selection Method and Within-Unit Coverage Problems." Paper presented at the annual meeting of the American Association for Public Opinion Research, St. Charles, IL.

- 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.
- Lind, Katherine, Michael Link, and Robert Oldendick. 2000. "A Comparison of the Accuracy of the Last Birthday versus the Next Birthday Methods for Random Selection of Household Respondents." Paper presented at the annual meeting of the American Association for Public Opinion Research, Portland, OR.
- Lipset, Seymour Martin. 1959. "Some Social Requisites of Democracy: Economic Development and Political Legitimacy." *American Political Science Review* 53 (1): 69–105.
- Lupu, Noam, and Kristin Michelitch. 2018. "Advances in Survey Methods for the Developing World." *Annual Review of Political Science* 21: 195–214.
- 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." *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.
- 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, NJ: Princeton University Press.
- Rizzo, Louis, J. Michael Brick, and Inho Park. 2004. "A Minimally Intrusive Method for Sampling Persons in Random Digit Dial Surveys." *Public Opinion Quarterly* 68 (2): 267–74.
- Salmon, Charles T., and John Spicer Nichols. 1983. "The Next-Birthday Method of Respondent Selection." *Public Opinion Quarterly* 47 (2): 270–76.

- Singer, Elinor 2006. "Introduction: Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70 (5): 637–45.
- Smith, Tom W. 1995. "Research Notes: Trends in Non-Response Rates." *International Journal of Public Opinion Research* 7 (2): 157–71.
- 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* 181 (3): 757–81.
- Sullivan, J. L. and J. 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*. Published ahead of print. <https://doi.org/10.1093/poq/nfaa026>
- Verba, Sidney, Kay Lehman Schlozman, and Henry E. Brady. 1995. *Voice and Equality: Civic Voluntarism in American Politics*. Cambridge, MA: Harvard University Press.
- Yan, Ting. 2009. "A Meta-Analysis of Within-Household Respondent Selection Methods." Paper presented at the annual meeting of the American Association for Public Opinion Research, Hollywood, FL.

Zechmeister, Elizabeth. 2006. "What's Left and Who's Right? A Q-method Study of Individual and Contextual Influences on the Meaning of Ideological Labels." *Political Behavior* 28 (2): 151–73.

Ziniel, Sonja. 2008. "Within-Unit Selection." In *Encyclopedia of Survey Research Methods*, edited by Paul J. Lavrakas, 962–64. Thousand Oaks, CA: Sage.

## Supplementary Material

**Table S1.** Last Birthday segment selection

Municipality	FM neighboring segments	
	Total	Selected
Alajuelita	24	8
Aserrí	21	8
Desamparados	25	9
Escazú	24	8
Goicoechea	25	8
Montes de Oca	24	8
San José	51	18
Tibás	20	8
Total	214	75

**Table S2.** Comparison of LB segments by closeness originally planned locations

	Closer	Farther	Difference
Female	0.48	0.55	0.07 (0.08)
Age	40.79	41.77	0.98 (2.32)
Education	11.42	9.89	-1.53 (1.13)
Wealth index	3.78	3.45	-0.32 (0.24)

*Note.* Distances between centroids of originally planned and actually surveyed LB segments. Closer = distance less than median. Farther = distance greater than median. Weighted estimates. Standard errors (for difference estimates only) in parentheses. None of the estimated differences are significant on the 95% confidence level.

**Table S3.** Wealth index: results of factor analysis

Variable	Factor loading
Internet	0.88 (0.04)
Computer	0.86 (0.04)
Cellphone	0.71 (0.09)
Car	0.63 (0.05)
LCD TV	0.60 (0.06)

*Note.* Estimator = weighted least squares adjusted for means and variances (WLSMV).  $\chi^2 = 6.45$ ,  $df = 5$ ,  $p = 0.265$ ; RMSEA = 0.021; CFI = 0.998. Geomin rotation. Standard errors in parentheses. All factor loadings significant at the 99.9% confidence level.

**Table S4.** Reliability statistics for multi-item indices

	No. of items	Cronbach's alpha
Political trust	10	0.87
Political participation	5	0.68
Political tolerance	4	0.80

**Table S5.** Wording of survey items [translated from Spanish]

---

*Big Five personality traits*

“Now I am going to read you a series of personality traits that may or may not apply to you. Please use the 1–7 ladder to indicate the extent to which you agree or disagree that these statements apply to you. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.”

Openness to experience

- A person [who is] open to new experiences and intellectual
- An uncreative and unimaginative person (reversed)

Conscientiousness

- A dependable and self-disciplined person
- A disorganized and careless person (reversed)

Extraversion

- A sociable and active person
- A quiet and shy person (reversed)

Agreeableness

- A critical and quarrelsome person (reversed)
- A generous and warm person

Emotional stability

- An anxious and easily upset person (reversed)
- A calm and emotionally stable person

Answers coded from 1 = “Strongly disagree” to 7 = “Strongly agree”

*Voted last election*

“Did you vote in the last presidential elections?”

Answers: 1 = “Yes” and 2 = “No”

*Political participation index*

“During the last campaign, did you...”

1. Put a flag of a political party on your house
2. Stick a political campaign stick on your car or home
3. Visit houses to take voters to the polls
4. Contribute money to help a candidate
5. Attend a political meeting or demonstration

Answers: 1 = “Yes” and 0 = “No”

*Ideology*

“On this card there is a 1–10 scale that goes from left to right. The number one means left and 10 means right. Nowadays, when we speak of political leanings, we talk of those on the left and those on the right. In other words, some people sympathize more with the left and others with the right. According to the meaning that the terms ‘left’ and ‘right’ have for you, and thinking of your own political leanings, where would you place yourself on this scale? Tell me the number.”

Answers coded from 1 = “Left” to 10 = “Right”

*Support for democracy*

“...Democracy may have problems, but it is better than any other form of government. To what extent do you agree or disagree with this statement?”

Answers coded from 1 = “Strongly disagree” to 7 = “Strongly agree”

*Political trust index*

1. To what extent do you trust the National Congress?
2. To what extent do you trust the National Police?
3. To what extent do you trust the political parties?
4. To what extent do you trust the President?
5. To what extent do you trust the Supreme Court?
6. To what extent do you trust the Fourth Chamber [of the Supreme Court]?
7. To what extent do you trust the local or municipal government?
8. To what extent do you trust elections in this country?

Answers coded from 1 = “Not at all” to 7 = “A lot”

*Political tolerance index*

1. There are people who only say bad things about the country form of government, not just the current government but the system of government. How strongly do you approve or disapprove of such people’s right to vote?
2. How strongly do you approve or disapprove that such people be allowed to conduct peaceful demonstrations in order to express their views?
3. Still thinking of those who only say bad things about the country form of government, how strongly do you approve or disapprove of such people being permitted to run for public office?
4. How strongly do you approve or disapprove of such people appearing on television to make speeches?

Answers coded from 1 = “Strongly disapprove” to 7 = “Strongly approve”

---