Using Mixed Methods to Better Sample Migrant Populations: Surveying the Venezuelan Diaspora in Colombia and Peru

ABSTRACT

Reaching migrant and diaspora populations entails distinct methodological hurdles arising from their mobility, clustered settlement patterns, and the frequent absence of comprehensive sampling frames. Adaptive Cluster Sampling (ACS) offers a probabilistic but flexible design for surveying hard-to-reach populations, but implementing ACS requires making methodological decisions that depend on the context. We employ a mixed-methods approach to implement ACS to survey Venezuelan migrants living in Colombia and Peru in 2024. To do so, we use qualitative interviews, administrative records, and small-scale pilot studies to identify settlement clusters, refine definitions of households and respondent eligibility, and set threshold triggers and stopping rules for sample expansion. These initial studies inform our implementation of multistage ACS in two large-scale surveys. Our results demonstrate that the mixed-method formative phase proved indispensable for identifying high-density neighborhoods and adapting fieldwork procedures to local contexts. Even though our two studies sampled the same migrant population, successful cross-national sampling required calibrating ACS parameters to country-specific densities. In Peru, larger census zones served as the optimal sampling unit, whereas in Colombia, smaller city blocks were more effective due to the greater concentration of Venezuelan households. We conclude that mixed-methods approaches can enhance coverage, efficiency, and feasibility in adaptive designs for surveying high-mobility and hard-to-reach populations, particularly in urban settings where traditional probability sampling methods are impractical.

Keywords: Adaptive Cluster Sampling, Migrant Populations, Venezuelan Diaspora, Mixed Methods, Latin America.

The primary challenge survey researchers face when sampling hard-to-reach populations like migrants – who are sparse and typically clustered in specific cities and neighborhoods, often temporarily – is the lack of reliable sampling frames. As a result, standard probability-based designs frequently struggle with incomplete coverage and limited precision, making them inefficient and resource-intensive for highly clustered populations (Kalton 2009; Lynn et al. 2018). Failure to accurately identify these respondents can generate biases that misrepresent the composition and responses of migrant populations, leading to misguided policy and humanitarian decisions (Kühne et al. 2019).

One solution that scholars have proposed is Adaptive Cluster Sampling (ACS), a probabilistic, multi-stage design that begins with a standard cluster sample and adaptively expands to neighboring clusters when predefined conditions are met, thereby maintaining probabilistic inference while minimizing coverage error and preserving cost-benefit efficiency (Thompson 1990; Aga et al. 2023). In practice, ACS begins like a standard cluster sample but expands the sample to nearby clusters whenever a sampled area meets a predefined threshold, continuing this process until no additional eligible clusters are found or a stopping rule is reached. Since the key characteristic of ACS is its adaptability, implementing it also requires making some methodological decisions that are specific to the population of interest and the context. This paper focuses on the definitions of a household and respondent eligibility, the multistage sampling design, and the threshold triggers and stopping rules for sample expansion in an application of the ACS design to the Venezuelan diaspora in Colombia and Peru in 2024.

Large-scale Venezuelan out-migration to destinations throughout the Americas began in 2016 and has continued since, with over six million Venezuelans seeking political freedom, physical safety, and economic opportunities outside Venezuela (Alvarez et al. 2022). This population is characterized by high rates of geographic mobility, irregular living arrangements, and wide variation in income and education levels (Muñoz-Pogossian & Winkler 2023). Colombia and Peru have received the largest populations of Venezuelan migrants, making them ideal cases for assessing how sampling the same migrant group under significantly different fieldwork conditions and local contexts may differentially shape the parameterization and performance of ACS, with consequences for coverage, cost, and fieldwork feasibility.

We implemented a mixed-methods approach to developing the ACS design in each country. First, we combined qualitative interviews and data from administrative records to

identify clusters of Venezuelan households. Formative qualitative research was essential for understanding local settlement patterns, refining eligibility criteria, and guiding the selection of feasible sampling units in contexts where reliable frames were absent. Building on these insights, we then conducted pilot face-to-face household interviews in high-concentration urban areas before scaling up to a large-sample national survey.

Our results demonstrate that sampling the same migrant population across two different Latin American countries demanded substantial adaptation to local contexts. In both settings, a mixed-methods approach proved crucial to informing the ACS design across our two surveys. The field teams had to redefine what constitutes a household, departing from official census definitions, and to screen participants by country of birth rather than nationality. Differences in Venezuelan population density across the two countries required distinct sampling units: city blocks in Colombia and census tracts in Peru. These different population densities also implied markedly different threshold rules for triggering and stopping sampling expansion. We conclude with practical recommendations for researchers and practitioners seeking to sample hard-to-reach migrant populations.

Method and Data

Qualitative Interviews

To determine the optimal sampling units and threshold rules for implementing ACS, we conducted a series of mixed methods approaches combining in-depth qualitative interviews, administrative records, and pilot studies. We first carried out formative qualitative research to inform sampling design and identify areas with significant Venezuelan migrant presence. This involved collecting census and official government data. Next, in order to identify the spatial distribution of Venezuelan households, we conducted three in-depth qualitative interviews with local migrant-serving organizations: *Fundación Juntos se Puede* in Colombia and *VeneICA* and *Servicio Jesuita a Migrantes* in Peru. In order to understand the complexity of Venezuelan migration in the two countries, we also interviewed a member of the research group

Observatorio de Venezuela in Colombia who is an expert on the Venezuela diaspora.[1] These interview subjects were identified based on a broad web search for local organizations and practitioners working with migrants in the two countries.

The semi-structured interviews lasted an hour on average, during which we covered a series of predefined topics. The discussions helped us understand the geographic distribution of Venezuelan migrants in Colombia and the region, gain familiarity with the available records on migration patterns, and garner recommendations for additional data sources to consult. This qualitative data provided the foundation for delineating primary sampling units (PSUs) and secondary sampling units (SSUs) and for assessing where adaptive expansion would be most feasible.

Quantitative Pilot Studies

Researchers have proposed various probabilistic techniques to reach hidden and clustered populations, ranging from screening-based designs and respondent-driven sampling to name-based ('onomastic') approaches (Heckathorn 1997; Kalton 2009; Reichel & Morales 2017).[2] These multiple approaches share the goal of overcoming either (a) the absence of a sampling frame, or (b) the absence of geographic information on where clustered populations reside. At the same time, however, a central challenge with ACS arises from the potential for uncontrolled sample growth (Aga et al. 2023): when the sampling process encounters areas with high densities of eligible respondents, the adaptive mechanism can result in situations where a small number of large clusters dominate the final sample.

The literature offers different approaches to deal with this problem, each with distinct methodological implications, practical considerations, and tradeoffs.[3] We constrain the adaptive process by truncating clusters at the secondary sampling units by default. This truncation approach offers a pragmatic compromise between statistical efficiency and fieldwork feasibility. We adopt this strategy by restricting the addition of sampling points (such as city blocks) to the

[1] While each organization operates in distinct national and local contexts, the conversations across the two countries converged on the centrality of documentation and the economic and social vulnerability of migrants.
[2] See Isernia et al. (2018) for a comprehensive review of different approaches.
[3] When density is high and clusters expand rapidly, practitioners can apply truncation at SSU boundaries to control sample growth. Conversely, when density is low, relaxing truncation or elevating the SSU can reduce empty canvassing and fieldwork costs.

census segment that corresponds to the original sampling location. In practice, when a block meets the expansion threshold, we add only its in-segment neighbors; adjacent blocks that cross the segment boundary are excluded from the network.

The adaptive feature of ACS lies in its ability to continue or stop selection based on whether target households are found within each PSU. We implement the ACS within each previously identified PSU.[4] We initially selected 68 SSUs in Colombia and 58 in Peru, randomly choosing a block at the centroid of the PSU as the starting point for the ACS procedure.[5] Pilot studies were implemented in both countries to evaluate the operational feasibility of ACS. The pilot studies served three main objectives: (1) to assess the feasibility of implementing ACS in dense urban contexts; (2) to define appropriate geographic levels for PSU and SSU; and (3) to refine protocols for triggers to initiate and terminate adaptive expansion.

The pilot fieldwork in Lima, Peru was conducted by the Institute for Peruvian Studies (*Instituto de Estudios Peruanos*, IEP) between April 29 and May 9, 2024, yielding 32 interviews. The pilot study in Bogotá, Colombia was conducted by BIS Consulting between June 19 and June 26, 2024 and resulted in 33 complete interviews.[6] In the first stage, we selected a series of PSUs within cities with probability proportional to size. Definitions of PSUs vary across the two countries reflecting differences in available mapping and census structures. To achieve gender balance, interviewers identified the Venezuelan household member with the most recent birthday and requested to interview that person. If no one recalled whose birthday was most recent, any Venezuelan present at the time was interviewed.

---

[4] Salehi and Seber (1997) propose a multistage ACS design that distinguishes between PSUs and SSUs. Under this framework, adaptive sampling proceeds within a PSU by adding neighboring SSUs to the initial sample, but expansion is halted once the boundaries of the PSU are reached. Su and Quinn (2003) offer a variation of this method by recommending termination of the adaptive process after a fixed number of iterations, typically greater than three. The optimal number of iterations depends on the underlying population density which is initially unknown to researchers.

[5] In the main study, 139 PSUs were selected. This includes 71 districts in Peru and 68 census sectors in Colombia, from which secondary sampling units were drawn and adaptive cluster expansion was implemented. The target sample size for the main study sought 1,008 interviews across three provincial strata in Peru (Lima n = 464, Trujillo n = 304, and Tumbes n = 240). In Colombia, the design also targeted 1,008 interviews, with 252 interviews in each of four city strata (Bogotá, Medellín, Ipiales, and Cúcuta).

[6] In Lima, pilots were conducted in two districts representing different levels of migrant concentration: Chorrillos (high-density, 16 interviews) and Independencia (medium-to-low density, 16 interviews). In Bogotá, pilots were carried out in Suba (high-density, 17 interviews) and Engativá (medium-density, 16 interviews).

Results

Qualitative Interviews with Migrant-Serving Organizations and Experts and Administrative Records

We conducted a mixed methods approach combining in-depth qualitative interviews and quantitative pilot studies. To identify areas with significant Venezuelan migrant presence, we began by collecting census and official government data in both countries. This preliminary mapping of migrant concentration enabled us to conduct the qualitative interviews with an informed understanding of Venezuelans migrants' settlement patterns. The in-depth interviews confirmed the lack of reliable sampling frames and revealed that the spatial patterns of Venezuelan migrants across both countries was clustered in dense urban areas such as Bogotá and Lima. The migrant-serving field teams operating in more peripheral or isolated zones—farther from transportation and formal employment—reported frequently spending substantial time traveling to areas far from the city center without successfully locating Venezuelan migrants. These organizations reported that Venezuelan migrants concentrate in central or semi-central neighborhoods located near transportation hubs, markets, and other commercial corridors.[7] This finding was crucial to minimizing fieldwork costs and time and designing the adaptive cluster rules prior to launching the large-scale survey.

Another key insight from the interviews was the prevalence of dual nationality among Venezuelans residing in Colombia. Many Colombians who hold Venezuelan nationality were not born in Venezuela and, therefore, would not meet the criteria for migrant status within the scope of our study. To ensure comparability across the two contexts of our study, we therefore decided to only interview individuals born in Venezuela, regardless of whether they held Venezuelan nationality.

The interviews also revealed that multiple families of Venezuelan migrants frequently lived within the same dwelling, meaning that we needed to reassess our definition of a household. Traditional survey methodologies often define a household as a group of individuals who "eat most meals from the same pot." However, this definition posed validity concerns when

[7] Still, the organizations interviewed report that these outlying areas house a small share of Venezuelans, many of whom may deliberately avoid visibility due to concerns about their documented status or interactions with local authorities.

applied to the Venezuela migrants because communal dining arrangements are common among migrants. In such contexts, multiple individuals from different families may share meals in communal settings, rendering the conventional definition misleading and inadequate for household sampling. Accordingly, we redefined household eligibility for our sample as Venezuelan-born individuals living in a household who are related by birth, marriage, civil union or adoption. When multiple Venezuelan family units shared the same dwelling, each family unit was treated as a separate household for sampling purposes. Therefore, in dwellings containing more than one eligible Venezuelan household, one respondent was selected from each household.

Finally, our interviews also revealed substantial social stigma towards Venezuelan migrants, due to xenophobic host populations or the precarity caused by their undocumented status. These issues could deter participation in the survey. Prior research suggests that interviewer–respondent similarity in terms of nationality or ethnicity enhances rapport and response accuracy (Tourangeau et al. 2000; Lipps 2010; Adida et al. 2016; West and Blom 2017; Lupu and Michelitch 2018; Miles et al. 2022). Our fieldwork teams therefore prioritized Venezuelan interviewers in order to gain respondents' trust, improve response rates, and reduce the likelihood of social desirability bias.[8]

Quantitative Pilot Surveys

Operational experience from the pilot phase led to several practical decisions.[9] During the early stages of fieldwork in Lima, significant implementation challenges emerged when attempting to use city blocks as the SSU. The adaptive component of the sampling design was activated when the number of identified Venezuelan households within an SSU reached the predetermined threshold. Initially, we used 16 interviews as the benchmark for activation but because the density of Venezuelan respondents within individual blocks was too low to trigger the adaptive mechanism, the ACS procedure frequently terminated prematurely.

[8] Interviewers were recruited by local fieldwork teams in compliance with labor and research regulations in each field site. All recruitment procedures were reviewed to ensure they were lawful, non-coercive, and consistent with institutional ethical approval at Vanderbilt University.

[9] For example, an initial qualitative observation from the IEP team in Peru was that the fieldwork was most effective when conducted after standard work hours, which helped improve gender balance among respondents.

As a result, we had to redefine the SSU as the census zone for ACS implementation in Peru. Census zones facilitated adaptive expansion and simplified logistical coordination. Therefore, PSUs were defined as districts and were selected with probability proportional to migrant population size based on available census data. By contrast, in Bogotá, the distribution of Venezuelan households and local enumeration systems made blocks the most appropriate sampling unit. PSUs were defined as localities, again emphasizing areas of high migrant density.

In Peru, adaptive sampling was initiated when 10 or more Venezuelan residents were identified within a census zone, whereas in Colombia the threshold was set at 3 or more residents within a block. These differences reflect the use of distinct SSUs: census zones in Peru encompass much larger geographic areas and populations than city blocks in Colombia, necessitating a higher activation threshold. These parameters were empirically informed by pilot fieldwork, balancing the need for adequate coverage with resource efficiency and ensuring that adaptive recruitment expanded only in areas of sufficient migrant concentration.

Conclusion

Our experience studying the same hard-to-reach population across two countries underscores the need for applied researchers to employ mixed methods in designing an implementation of ACS. The issues inherent in sampling hard-to-survey populations underscore the value of adaptive sampling methodologies that are cost-effective and feasible. The combination of qualitative interviews and pilot studies allowed us to implement ACS efficiently while ensuring appropriate geographical coverage across two contexts with very different migrant densities. The formative qualitative research was indispensable in identifying geographic units with high and medium densities of Venezuelan migrants, refining both the household and respondent eligibility criteria, and choosing the triggers to initiate and terminate adaptive expansion.

By redefining the sampling unit and adapting the threshold expansion protocol, these changes informed a revised fieldwork protocol designed to increase fieldwork efficiency and respondent contact rates. While the density of the Venezuelan migrant population and observed response rates favored the use of census zones rather than individual blocks as SSUs in Peru, in Colombia, the distribution of Venezuelan households and local enumeration systems made

blocks the most appropriate. To improve cost-effectiveness and reduce fieldwork time without compromising representativeness, we tailored the ACS design to different clustered population structures rather than applying a uniform protocol across different contexts.

We recommend that researchers begin with qualitative interviews of experts and organizations working on the ground with the target population to better understand study design elements including geographic distribution, definitions of households, and eligibility criteria. We also recommend conducting pilot studies that begin at lower-level sampling units—such as city blocks—and conduct an initial sweep of three to five blocks to establish a baseline of target population density and clustering. Fieldwork teams should start with provisional expansion thresholds and escalate to a larger SSU if fewer than predefined eligible households are located across the first sweep—as in our Peru implementation. Researchers can then aggregate to the next-higher SSU and reassess the population clustering before proceeding. This early diagnostic phase will allow researchers to make informed decisions about whether to continue sampling at that level or to shift to higher-level SSUs, optimizing both coverage and resource allocation. Such mixed-methods approaches are crucial for adapting sampling approaches for hard-to-reach populations to the local context.

REFERENCES

Adida, Claire L., Karen E. Ferree, Daniel N. Posner, and Amanda L. Robinson. 2016. "Who's Asking? Interviewer Coethnicity Effects in African Survey Data." *Comparative Political Studies* 49 (12): 1630–60.

Aga, Gizem, David C. Francis, Filip Jolevski, Jose Rodriguez Meza, and James S. Wimpey. 2023. "An Application of Adaptive Cluster Sampling to Surveying Informal Businesses." *Journal of Survey Statistics and Methodology* 11 (5): 1246–66.

Alvarez, Juan, Maria M. Arena, Alexandra Brousseau, M. Hashem Faruqee, Eleonora W. F. Corugedo, Jaime Guajardo, Gustavo Peraza, and Javier Yepez. 2022. *Regional Spillovers from the Venezuelan Crisis: Migration Flows and Their Impact on Latin America and the Caribbean.* Washington, DC: International Monetary Fund.

Heckathorn, Douglas D. 1997. "Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations." *Social Problems* 44 (2): 174–99.

Kalton, Graham. 2009. "Methods for Oversampling Rare Subpopulations in Social Surveys." *Survey Methodology* 35 (2): 125–41.

Kühne, Simon, Jannes Jacobsen, and Martin Kroh. 2019. "Sampling in Times of High Immigration: The Survey Process of the IAB-BAMF-SOEP Survey of Refugees." *Survey Methods: Insights from the Field*.

Kurt, Selin, and Sabine Hans. 2017. "Sampling Immigrants in the Netherlands and Germany." *Comparative Migration Studies* 5 (1).

Lipps, Oliver. 2010. "Interviewer-Respondent Socio-Demographic Matching and Survey Cooperation." *Survey Practice* 3 (4).

Lupu, N. and Michelitch, K., 2018. Advances in survey methods for the developing world. *Annual Review of Political Science*, *21*(1), pp.195-214.

Lynn, Peter, Alita Nandi, Violetta Parutis, and Lucinda Platt. 2018. "Design and Implementation of a High-Quality Probability Sample of Immigrants and Ethnic Minorities: Lessons Learnt." *Demographic Research* 38: 513–548.

Miles, Hannah, Christian Els, and Cholpon Ramizova. 2022. "The Role of Interviewer Ethnicity on Survey Responses: A Case Study in Refugee Camps in Bangladesh." *Survey Practice*, September.

Muñoz-Pogossian, Betilde, and Abigail Winkler. 2023. *The Persistence of the Venezuelan Migrant and Refugee Crisis.* Washington, DC: Center for Strategic and International Studies.

Reichel, David, and Laura Morales. 2017. "Surveying Immigrants without Sampling Frames: Evaluating the Success of Alternative Field Methods." *Comparative Migration Studies* 5 (1): 1.

Thompson, Steven K. 1990. “Adaptive Cluster Sampling.” *Journal of the American Statistical Association* 85 (412): 1050–59.

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response.* Cambridge: Cambridge University Press.

West, Brady T., and Annelies G. Blom. 2017. “Explaining Interviewer Effects: A Research Synthesis.” *Journal of Survey Statistics and Methodology* 5 (2): 175–211.