VANDERBILT UNIVERSITY

Center for Global Democracy

Venezuelan Diaspora in Colombia

Technical Information

Target Pop.	Year	Sample Size	Weighted/Unweighted	Fieldwork dates
Venezuelan Diaspora in Colombia	2024	1,006	Weighted	September 18th – December 7th, 2024

1. Introduction

This technical report outlines the survey methodology used to gauge public opinion among the Venezuelan diaspora in Colombia. Sponsored by the United States Agency for International Development (USAID), this study is part of a broader project aimed at understanding how integration influences Venezuelan migrants' tendencies toward onward migration in Peru and Colombia. With over 7.8 million Venezuelans living abroad, including 4.5 million in these two countries, this study addresses critical gaps in understanding the willingness, knowledge, and capacity of Venezuelans to fully integrate into their host societies.

The selection method used in this study is *Adaptive Cluster Sampling* (ACS), a pseudo-probability sampling technique designed to address the challenges of surveying hard-to-reach populations. These challenges include unreliable sampling frames and the low prevalence of the Venezuelan diaspora within the general population. Unlike non-probability methods such as snowball or venue-based sampling, ACS enables the calculation of selection probabilities, thereby enhancing the rigor and reliability of the sampling process.

ACS begins with a probability-based selection of primary sampling units (PSUs), using available sampling frames such as census data, cartographic resources, and prior studies. PSUs are selected with probability proportional to the size of the Venezuelan population in each unit. In the second stage, a random sample of secondary sampling units (SSUs) is drawn. If a selected SSU contains a member of the target population (i.e., Venezuelan households), neighboring SSUs are added to the sample. This adaptive expansion continues recursively until no additional target households are found, the PSU is fully covered, or the desired sample size is reached.

2. Formative Research and Pilot Testing

Formative research is essential to assess the feasibility of using the ACS with the Venezuelan diaspora. It provides key insights into demographic characteristics, social networks, and access points, ensuring the sampling design is context-sensitive, efficient, and ethically sound.

To conduct this research, LAPOP partnered with Instituto de Estudios Peruanos (IEP) to carry out pilot studies in Suba and Engativa (Bogota, Colombia). IEP subcontracted BIS Consulting to conduct the study from June 26 to 28, 2024, with 33 Venezuelan participants and 2 interviewers. The findings informed the ACS design.

Key Takeaways of the Formative Study in Colombia:

- Use blocks as SSUs in Colombia, as they contain sufficient numbers of Venezuelan households (as opposed to Peru).
- Include at least 3 Venezuelan households in a selected block to expand the sample using ACS.
- Develop a protocol for field teams when approached by the National Police.
- Prioritize female interviewers to foster openness.
- Equip teams with portable chargers due to limited device battery life.
- Provide physical maps to aid navigation and daily planning.
- While respondents were generally open regardless of interviewer nationality, Venezuelan interviewers were preferred for greater empathy.
- Delay fieldwork until after Venezuela's presidential elections (July 28, 2024) to avoid response bias.

3. Sampling Strategy in Colombia

Based on a review of the ACS methodology, the availability of sampling frames, and the results from the formative study, LAPOP developed the following sampling strategy.

a) Sampling Frame

The sampling frame for the study in Colombia is based on data from the 2018 National Population and Housing Census (with a count of people who had been residing in Venezuela 12 months prior to the 2018 census).

b) Stratified, Multi-Stage Cluster Design

The sample for the pseudoprobability survey was selected using a stratified, multi-stage adaptive cluster sampling design. The strata include the districts of Bogotá (n=252) and Medellín (n=252), and the municipalities of Ipiales (n=252) and Cúcuta (n=252).[1] These cities were chosen based on criteria such as the density of the target population and USAID's areas of interest.

c) Primary Sampling Units

In the first selection stage, 68 Primary Sampling Units (PSUs), or census sectors, were identified within the districts and municipalities. The districts of Bogotá, Medellín, and the municipality of Cúcuta each included 21 PSUs, while the municipality of Ipiales included 5 PSUs. These sectors were selected with probability proportional to the size of the Venezuelan population.

d) Secondary Sampling Units

ACS was implemented in the second stage of sampling, within the previously defined PSUs. The process began with the random selection of a city block—referred to as a Secondary Sampling Unit (SSU)—within each PSU. LAPOP initially selected 68 SSUs, one per PSU.

If the initial SSU contained a sufficient number of Venezuelan households, the field team added an adjacent SSU to the sample. This process continued—expanding to neighboring SSUs—until an SSU was reached that did not meet the threshold for household concentration. Based on pilot results from a study in Bogotá, LAPOP defined a "sufficient concentration" as at least 3 Venezuelan households per SSU.

Fieldwork training covered selection protocols, data collection software, ethics, and quality control, and concluded with a certification test. Quality control measures included GPS verification, continuous monitoring, and regular tracking sheets.

The team conducted the SSU integration tour starting with the SSU located directly north of the original one. If this SSU contained a sufficient concentration of Venezuelan households, the selection continued in a clockwise spiral. The process ended when an SSU with fewer than 3 Venezuelan households was encountered or

[1] According to Migración Colombia 2023 and 2018 Census, 37.7% of the Venezuelan population in Colombia resides in these 4 cities.

when all blocks within the PSU had been covered. Geofences were established at the SSU level.

The selection protocol is illustrated in Figure 1, using the Bogotá case as an example. The initial block selected is 0107. If at least 3 Venezuelan households were identified in this block, the adjacent block to the north was added, and additional neighboring blocks were included in a clockwise spiral.

In this example, the order of block visits was: 0108, 0104, 0103, 0106, 0213, 0202, 0110, 0115, 0203, 0206, 0207, 0114, 0101, 0204, 0205, 0210, 0209, 0208, 0211, 0212. This sequence assumes that each block contains at least 3 Venezuelan households.

Figure 1: Protocol for Selecting City Blocks (SSUs) Within Urban Sectors in Colombia

It was not necessary to conduct interviews in all 3 households to proceed to the next block. In fact, the next SSU could have been selected even if no interviews were completed in the previous one, as long as the household threshold was met. The adaptive component of the ACS design allows the selection process to depend on the presence or absence of Venezuelan households within each SSU. As a result, completing fieldwork depended not only on reaching the target sample size per city, but also on whether a block contained the minimum required number of households or had already been fully covered.

In contrast, conventional sampling designs involve selecting all sample points in advance, before any field observations are conducted.

e) Blocks and Households

Within each block, the interviewer began at the northeast corner and proceeded to visit 100% of the households, also moving in a clockwise direction. If a household included multiple Venezuelan households, one individual from each household was interviewed.

f) Respondents

In the third and final stage, individual respondents were selected. To be eligible, participants had to be Venezuelan by birth or naturalization, provided they were not born in Colombia. Additionally, they had to reside in the household and be of legal age.

The study did not use quotas. However, to promote gender balance among respondents, LAPOP implemented a variation of the "last birthday" method. Interviewers asked: "*Can you tell me whether the Venezuelan household member with the most recent birthday is a man or a woman?*" If the answer was "a man," the interviewer requested to speak with a male household member; if "a woman," a female member was requested.

If no one of the selected gender was available at the time, a return visit was scheduled. If, after the return visit, no one of the selected gender was present, the household was not interviewed. If the respondent did not know who had the most recent birthday, any eligible Venezuelan present was interviewed.

g) Sampling Summary and Sampled Population

Component	Peru	Colombia
Strata	Projected sample size: 1,008 • Lima (*n*=464) • Trujillo (*n*=304) • Tumbes (*n*=240) Final sample size: 1,077 • Lima (*n*=507) • Trujillo (*n*=327) • Tumbes (*n*=243)	Projected sample size: 1,008 • Bogota (*n*=252) • Medellin (*n*=252) • Ipiales (*n*=252) • Cúcuta (*n*=252) Final sample size: 1,006 • Bogota (*n*=254) • Medellin (*n*=251) • Ipiales (*n*=252) • Cucuta (*n*=249)
PSUs	Districts • Selected: 71 • Observed: 16	Census sectors • Selected: 68 • Observed: 28
SSUs	Census zones • Selected: 1 per PSU (71) • Observed: 38	City blocks • Selected: 1 per PSU (68) • Observed: 252
ACS Trigger	≥10 Venezuelans in a census zone	≥3 Venezuelans in a block
Stopping Rule	<10 Venezuelans, full PSU coverage, or n=1,008 achieved	<3 Venezuelans, full PSU coverage, or n=1,008 achieved
Household Selection	All households attempted	All households attempted
Respondent Selection	Venezuelans not born in Peru, modified last-birthday method	Venezuelans not born in Colombia, modified last-birthday method

4. Estimation Approach

The weighting process for the 2024 Venezuelan Diaspora Survey in Peru and Colombia was designed to produce nearly unbiased estimates of the Venezuelan population in selected areas of both countries. It followed a multi-step approach to account for selection probabilities, unknown eligibility, and nonresponse.

a) **Design weights** were calculated as the inverse of the selection probabilities across the survey's multi-stage sampling. This included adjustments for the adaptive cluster sampling method, where household selection expanded to neighboring segments if the initial segment met a minimum threshold of Venezuelan households.

b) **Eligibility adjustments** were applied to account for households with undetermined eligibility due to non-contact or refusal to complete the screening.

c) **Nonresponse** adjustments were made to reduce potential bias from differences between respondents and nonrespondents.

d) **Calibration and post-stratification weights** were applied to ensure that the sampling distribution of key variables mirrors the distribution in the frames

Finally, the distribution of the resulting weights was reviewed for extreme values, which were trimmed to minimize undue variability in the survey estimates.

The Stata version of the dataset has been saved with the command **`svyset upm [pw=wt], strata(estratopri)`** needed for estimating parameters that are adjusted for the complex sampling (weights and survey characteristics).

```
In R:
COLOMBIA <- survey::svydesign(ids = COLOMBIA$upm,
                              strata = COLOMBIA$estratopri,
                              weights = COLOMBIA$wt,
                              data = COLOMBIA)
```

5. Response Rates

Rate	Peru	Colombia
RR1	4.1%	13.8%
RR2	4.1%	13.9%
RR3	41.2%	30.1%
RR4	41.5%	30.1%
Total Attempts (excluded quality control attempts)	43834 (57)	8983 (314)
Total eligible respondents	1105	1020
Completed interviews	1077	1006
Partial (insufficient)	10	2
Eligible breakoff	18	12
Total unknown eligibility	25323	6246
Total ineligible	17406	1717
Estimated eligibility rate (CASRO)	6.0%	37.3%

6. Quality Control in Colombia

In the Venezuelan Diaspora Study in Peru, Quality Control was based on FALCON-CATI© (Fieldwork Algorithm for LAPOP's Control over Survey Operations and Norms). It includes, but is not limited to, an interviewer identity monitoring check, time checks, a reading control check, and data fabrication and falsification audits. The system also includes a quality control score that assigns penalties (or demerits) to interviews during the audit. In this system, higher scores indicate more serious errors, and we refuse to accept (that is, we require the cancelation of) low quality interviews.

The local firm audited 100% of the interviews. All interviews were also run through LAPOP's automatic flagging system, and then LAPOP's team manually audited a subset of the interviews. A total of 314 interviews were canceled in Colombia. The other most predominant reasons for canceling an interview were reading issues on the study information sheet, poor reading of multiple questions, and interviewers offering an interpretation of questions.

7. Socio-Demographic Profile of Weighted and Unweighted Datasets

The following tables summarize key socio-demographic characteristics, comparing unweighted and weighted estimates to reflect population distributions.

Variable	Peru (Unweighted)	Peru (Weighted)	Colombia (Unweighted)	Colombia (Weighted)
Location	Lima (47%), Trujillo (30%), Tumbes (23%)	Lima (88%), Trujillo (10%), Tumbes (3%)	Bogotá (25%), Medellín (25%), Ipiales (25%), Cúcuta (25%)	Bogotá (34%), Cúcuta (34%), Ipiales (7%), Medellín (25%)
Gender	Women (61%), Men (39%)	Women (47%), Men (53%)	Women (71%), Men (29%)	Women (52%), Men (48%)
Age	Avg: 34 yrs	Avg: 35 yrs	Avg: 36 yrs	Avg: 35 yrs
Ethnicity	White (30%), Morena (40%), Mestiza (18%), Black (9%), Other (3%)	White (26%), Morena (33%), Mestiza (23%), Black (12%), Other (5%)	White (32%), Morena (44%), Mestiza (17%), Black (5%), Other (1%)	White (32%), Morena (48%), Mestiza (13%), Black (6%), Other (1%)
Marital Status	Single (48%), Married (45%), Other (7%)	Single (49%), Married (44%), Other (7%)	Single (44%), Married (50%), Other (6%)	Single (46%), Married (48%), Other (6%)
Education	None (2%), Primary (8%), Secondary (59%), Superior (31%)	None (2%), Primary (10%), Secondary (56%), Superior (32%)	None (1%), Primary (17%), Secondary (59%), Superior (22%)	None (3%), Primary (20%), Secondary (58%), Superior (19%)
Employment	Employed (50%), Unemployed (24%), Caregiver (8%), Other (17%)	Employed (54%), Unemployed (25%), Caregiver (6%), Other (15%)	Employed (37%), Unemployed (36%), Caregiver (10%), Other (15%)	Employed (43%), Unemployed (37%), Caregiver (9%), Other (11%)

Variable	Peru (Unweighted)	Peru (Weighted)	Colombia (Unweighted)	Colombia (Weighted)
Sector (Employed)	Informal (93%), Formal (7%)	Informal (91%), Formal (9%)	Informal (88%), Formal (12%)	Informal (85%), Formal (15%)
Monthly Income	<S/635 (29%), S/635–980 (24%), S/981–1294 (20%), S/1295–1961 (14%), ≥S/1962 (12%)	<S/635 (28%), S/635–980 (27%), S/981–1294 (21%), S/1295–1961 (14%), ≥S/1962 (10%)	<COP449,708 (38%), COP449,708–721,963 (17%), COP721,964–967,632 (12%), COP967,633–1,459,510 (20%), ≥COP1,459,510 (13%)	<COP449,708 (35%), COP449,708–721,963 (15%), COP721,964–967,632 (12%), COP967,633–1,459,510 (18%), ≥COP1,459,510 (20%)
Household Size	1 (9%), 2 (17%), 3 (25%), 4 (21%), 5 (15%), 6+ (14%)	1 (4%), 2 (15%), 3 (26%), 4 (21%), 5 (20%), 6+ (14%)	1 (11%), 2 (15%), 3 (18%), 4 (24%), 5 (15%), 6+ (17%)	1 (7%), 2 (11%), 3 (15%), 4 (20%), 5 (16%), 6+ (31%)
Occupancy	Renter (91%), Homeowner (3%), Other (6%)	Renter (95%), Homeowner (2%), Other (3%)	Renter (93%), Homeowner (3%), Other (4%)	Renter (90%), Homeowner (6%), Other (4%)

8. Fieldwork Challenges and Lessons Learned

a) **Recruitment and Turnover:** Finding experienced Venezuelan interviewers was difficult, with high turnover (Peru: 18 to 11 interviewers; Colombia: four additional training workshops).

b) **Security**: Thefts and assaults required group deployments; six PSUs in Peru were abandoned due to threats.

c) **Hostility:** Some residents were hostile, and undocumented migrants feared participation due to immigration concerns or landlord restrictions.

d) **Logistical Issues:** Spiral routes were inefficient due to terrain barriers; transportation between PSUs was time-consuming.

e) **Protocol Violations:** In Colombia, inadequate supervision led to issues like interview fabrication and interviews conducted outside geofences.

For additional information on the survey design, contact Daniel Montalvo at daniel.montalvo@vanderbilt.edu.