

Annex A

Technical Description of the Sample

Sample Design¹

In Colombia, the sample was designed to include all non-institutionalized adults (i.e., it excludes people living in jails, schools, hospitals, and on military bases). It is a random stratified sample. The stratification ensures the inclusion of the most important geographic regions in the country: the Pacific, Atlantic, Central, Eastern, Old National Territories (*Antiguos Territorios Nacionales*), and Bogotá. The sample was sub-stratified to include cities with more and with less than 300,000 inhabitants. Finally, the sample was further sub-stratified into urban and rural areas.

We used population projections for 2006 from the 1993 census, the most recent in Colombia.² According to the census, 21.58% of the population lives in the Atlantic region, 17.57% in the Pacific region, 23.63% in the Central region, 18.31% in the Eastern region, 3.28% in the Old National Territories, and 15% in Bogotá.

Comentado [LR21]: two 'periods' in orig

Sample selection was also multistage. The first step was the municipality, then the census sector, followed by the census section, and finally the block, dwelling, and household. We used a system of sex and age quotas to select the respondent inside each household.

We interviewed 1,491 informants. Technically, our sampling error was $\pm 2.54\%$. This means that if we drew repeated samples in Colombia, 95% of them would reflect the views of the population with no less accuracy than $\pm 2.54\%$. Our sample, however, was stratified and clustered. This means that even though the stratification increased the accuracy of the sample, the clusters we used to control fieldwork costs reduced it. Of course, other factors beyond sampling can reduce the accuracy of the sample, including non-response, errors in selecting the respondent, misunderstanding the questions, among others. But in terms of the science of survey sampling, a confidence interval of $\pm 2.54\%$ is very good.

Table A-1 summarizes the standard errors and the design effects for some variables and indices in the survey. The design effects (DEF) indicate the efficiency of a cluster design compared to a simple random design. A DEF of 1 indicates that the variances obtained in both designs are the same, meaning that the cluster design was as efficient as a simple random design. If the DEF is greater than 1, it means that the cluster design had a variance greater than that produced by a simple random design. And if the DEF is less than 1, it means that the variance of the cluster design is even smaller than that produced by the random design.

Comentado [K22]: not capitalized in original

¹ This section and the following were adapted from the 2004 Costa Rica report, "Cultura democrática, seguridad ciudadana y capital social en Costa Rica," by Luis Rosero-Bixby and Jorge Vargas-Cullell.

² The results of the census advanced by the government since 2005 were not available when the sample used in this study was designed.

Table A-1. Standard Errors and Design Effects for Selected Variables

Variable	Mean	Std. Error	DEF
q2 (age)	35.934	.315	.560
ed (level of education)	9.367	.191	2.336
Wealth (wealth index)	3.835	.097	3.767
PSA5 (support for the system)	57.211	.755	1.532
tol (political tolerance)	53.226	1.024	1.614
soct1r (evaluation of the national economy)	39.266	.628	1.242
l1 (self-defined ideological location)	6.244	.082	1.246
m1r (evaluation of the government)	63.998	.799	1.478
exc7r (perception of corruption)	73.423	1.008	1.359

strata=region (estratopri)
upm=sector (colsect)

According to the above table, the cluster design for this survey was very efficient. In fact, with the exception of age, education, and wealth, all the DEF were close to 1. The standard errors for most variables were also very moderate.

Sample Results and Description of the Respondents

The probabilistic design of the sample, as well as the availability of a good sampling framework, are sufficient conditions to expect that the interviewed group is representative of the Colombian population. However, due to the effects of random errors and inevitable distortions in the sampling design, the sample could deviate from the characteristics of the population it represents. It could include biases that should be reported. Table A-2 allows us to answer the question: how representative is the sample of the population? In it, we compare some characteristics of the sample with the 1993 census.³

³ There are no recent projections for most of the indicators we analyze here.

Table A-2. Sample vs. 1993 Census (18 years and older)

Characteristics	1993 Census	Survey in Colombia
(N)	26,735,000	1,491
% of men	49	50
% > 30 years old	62	60
% single	40	32
% married or living with partner	50	59
% with primary education	44	34
% with secondary education	34	46
% with post-secondary education	11	20
% in Atlantic region	22	22
% in Bogotá	15	15
% in Central region	24	24
% in Eastern region	18	18
% in Pacific region	18	17
% in Old National Territories	3	4

We can see that there is some congruity between the sample of this survey and the 1993 census. Some characteristics such as age, sex, and regional residence are virtually identical. There is a slight deviation in the percentages of married and single people. And finally, there is a gap in the three education variables, where the widest is for people with secondary education, which was 34% in the 1993 census and is 46% in the 2006 survey. The explanation for the gaps in marriage and education variables may lie in the fact that the census is almost 13 years old. This is a long enough period for there to be significant changes in these variables. Unfortunately, there are no reliable projections from the 1993 census in terms of education and age groups; the figures from the 2005 census are not available either. There are some projections that include current students, but this makes them a non-comparable statistic with the variables of this survey.

Because the sample is generally representative of the population, it does not need to be weighted. The sample is, therefore, self-weighted.

Table A-3 compares sample characteristics between men and women.

Comentado [LR23]: format changes here from A-3 to A.3. I kept the first for consistency.

Table A-3. Characteristics of the Sample, by Sex

Characteristic	Total	Men	Women
(N)	1,491 (100%)	50%	50%
Average age	37.33	37.55	37.11
% married or living with partner	58.82	55.90	61.74
Family income (Mode) % between 361,000 and 420,000 pesos	26.95	26.48	27.43

We used sex and age quotas to select respondents. Therefore, the percentages of men and women that we have are very close. Their ages are also very similar. There is a small difference in the percentage of people who are married or living with a partner, where women have a slightly higher percentage (61.74%) than men. With regards to family income, there is a higher percentage of women in the 361,000 to 420,000-peso bracket, which is the mode of family income, although the gap is very small.

Technical Description of the Sampling Design

Universe

The survey universe has national coverage of adults living in all six regions of the country: Bogotá, the Atlantic, Pacific, Central, Eastern regions, and the Old National Territories. The universe is also comprised of all adults living in urban and rural areas.

The universe was divided in two sectors: one of cities more than 300,000 residents, and the other of cities with less than 300,000 inhabitants.

Population

The sample was circumscribed to all non-institutionalized adults; in other words, it excludes people living in jails, schools, hospitals, and on military bases. Private households in these areas were included.

Final Unit of Selection

Because the questionnaire included questions not only limited to the respondent but also to other household members, the statistical unit of observation was the household. The respondent could only live in one household.

Because each household belongs to a dwelling, sometimes shared by more than one household (often relatively stable over time), each dwelling was selected as the final unit of selection.

Sampling Method

We chose probabilistic, stratified, multistage sampling with a random selection of the units at each stage. First, the sample was stratified by the size of the municipality (cities with more and with less than 300,000 inhabitants), then by region and then by area (rural and urban).

It is multistage sampling because within each urban area, we started with primary sampling units (sectors), followed by secondary sampling units (sections), then third units (blocks), and then with final sampling units (clusters of dwellings) of 6 to 8 units in urban areas and 10 to 12 in rural areas. In each dwelling, the interviewer selected only one household as the observation unit.

The respondent was selected according to age and sex quotas. In each block, the surveyor had to include at least one man and one woman in the following age groups:

- 18 to 27 years old
- 28 to 40 years old
- Over 40 years old

Each interviewer was assigned one specific block. Once in the block, interviewers listed the first 20 dwellings they encountered. They had instructions to do a minimum of 8 surveys of the 20 dwellings they listed, balancing the sex and age quotas.

We chose the method of selection according to the following considerations:

We needed representative samples at the following levels:

- National
- First stage strata:
 - o Cities with more than 300,000 inhabitants
 - o Cities with less than 300,000 inhabitants
- Second stage strata:
 - o Bogotá
 - o Atlantic region
 - o Pacific region
 - o Eastern region
 - o Central region
 - o Old National Territories

- Third stage strata:
 - o Urban area
 - o Rural area
 - Study domains:
 - o Cities with more than 300,000 inhabitants (obliged selection)
 - o Cities with less than 300,000 inhabitants
- a) For each stage, we calculated margins of error that corresponded to minimum quality standards.
- b) We sought to facilitate the operability of the fieldwork.
- c) We worked with the best and most up-to-date sampling framework available for each municipality (population census, cartography, current housing unit lists, among others).

Sampling Framework

The sampling framework is constituted by the updated cartographic inventory and housing unit lists obtained from the 1993 census. The Centro Nacional de Consultoría obtained the 2003 versions from the Departamento Nacional de Estadística (DANE; National Statistics Department).

Calculations by Strata

The sample is composed of 193 sampling points: 160 urban and 33 rural, distributed over 53 municipalities in 26 out of the 32 departments of Colombia.

Sample Sizes, Confidence Intervals, and Margins of Error

The expected confidence interval for the national sample was 95% with a margin of error of 2.6%, assuming a 50/50 ratio in the dichotomous variables.

Table A-4 shows the margins of error for a 95% confidence interval.

Table A-4. Sample Size and Margins of Error (95% Level of Confidence)

Strata Regions	Sample Size	Margin of Error M.A.S.	% M.P.C.
Atlantic	323	5.45	5.58
Bogotá	231	6.45	6.59
Central	358	5.18	5.29
Eastern	274	5.92	6.05
Pacific	251	6.19	6.32
Old National Territories	54	13.34	13.63
Areas			
Urban	1107	2.96	3.01
Rural	384	5.00	5.11
Country Total	1491	2.54	2.6

Survey Team

The CNC involved its five branches (Bogotá, Cali, Medellín, Barranquilla, and Bucaramanga) to ensure high quality in the least possible time. Due to the country's current security situation, we were advised to remain as little time as possible in most of the areas visited, which complicated the work.

Due to the complexity of the questionnaire, we used our most experienced surveyors, many of whom have more than 15 years of experience.

The CNC used a total of 101 staff members, distributed as follows:

Table A-5. CNC Personnel Who Participated in the Study

Activity	Total personnel
Fieldwork coordinators	5
Supervisors	15
Interviewers	40
Quality supervisors in the field	10
Codifiers	7
Data entry	7
Data verification	7
Subtotal fieldwork and data entry	91
Professional and management personnel	5
Administrative personnel	3
Total survey team	99

Comentado [LR24]: doesn't match table

Table A-6. Population by Region: Urban and Rural Area (2006 Projections)

Region	Urban	Rural	Total
Atlantic	6,912,796 (70.9%)	2,843,894 (29.1%)	9,756,690
Bogotá	7,014,111 (100.0%)	1,517 (0.0%)	7,015,628
Central	7,659,319 (71.2%)	3,093,365 (28.8%)	10,752,684
Eastern	5,207,407 (62.6%)	3,106,566 (37.4%)	8,313,973
Pacific	5,322,949 (66.8%)	2,644,004 (33.2%)	7,966,953
Old National Territories	583,895 (39.6%)	890,830 (60.4%)	1,474,725
Total	32,700,477 (72.2%)	12,580,176 (27.8%)	45,280,653

Table A-7. Size and Distribution of the Sample by Strata (Region)

Region	Urban	Rural	Total
Atlantic			
> 300,000 inhabitants	127		127
< 300,000 inhabitants	112	84	196
Total Atlantic	239	84	323
Bogotá			
> 300,000 inhabitants	231		231
Total Bogotá	231		231
Central			
> 300,000 inhabitants	150		150
< 300,000 inhabitants	112	96	208
Total Central	262	96	358
Eastern			
> 300,000 inhabitants	66		66
< 300,000 inhabitants	112	96	208
Total Eastern	178	96	274
Pacific			
> 300,000 inhabitants	91		91
< 300,000 inhabitants	88	84	160
Total Pacific	179	72	251
Old National Territories			
< 300,000 inhabitants	18	36	54
Total Old National Territories	18	36	54
Total	1,107	384	1,491

Table A-8. Respondents and Primary Sampling Units (PSU) by Strata (Region)

Region	Sample	PSU
Atlantic	323	32
Bogotá	231	34
Central	358	39
Eastern	274	26
Pacific	251	26
Old National Territories	54	6
Total	1,491	163

Final Comments on Survey Fieldwork

About the questionnaire: the questionnaire was longer than previous LAPOP versions, 55 minutes on average according to interviewers, but respondents were generally willing to answer the questions and we had few uncompleted surveys.

About the fieldwork: For security reasons, it was not possible to do rural surveys in one of the municipalities. Additionally, since it was not possible to do surveys in Chalan (Sucre) in 2004 and 2005, this municipality was replaced in the 2006 survey.

Despite the fact that some respondents were located in areas where the illegal armed groups have a strong presence, there were no reports of any kind of pressure on respondents to influence their answers. On the contrary, the interviewers emphasized the freedom of opinion of the people who accepted to take part in the study.

As in previous years, the CNC would like to extend its gratitude to every staff member involved in this study, especially the brave men and women who defied security warnings and assumed great risk to accomplish very good work.