Appendix A: Technical Description of the Sample

Sample Design[1]

The sample in Colombia was designed to include all non-institutionalized adults (i.e., it excludes people living in the country's jails, schools, hospitals and military bases). It is a random stratified sample. The stratification ensures the inclusion of the most important geographic regions in the country: Pacific, Atlantic, Central, Eastern, the Old National Territories (*Antiguos Territorios Nacionales*), and Bogota. 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 2004 population projections for the 1993 Census, the most recent in Colombia. According to the census, 22 percent of the population inhabits the Atlantic region, 18 percent the Pacific, 24 percent the Central, 18 percent the Eastern, three percent in the Old National Territories, and 15 percent in Bogota.

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

We interviewed 1,487 respondents. Technically, our sampling error was +/- 2.6 percent. This means that if we drew repeated samples in Colombia, 95 percent of them would reflect the views of the population with no lesser accuracy than +/- 2.6 percent. Our sample, however, was stratified and clustered. This means that, although we increased the precision of the sample through stratification, the clusters we used to control fieldwork costs somewhat reduced it. Of course, other factors beyond sampling can also reduce the accuracy of the results, including non-response, errors in selecting the respondent, misunderstanding of the questions, among others. But in terms of the science of survey sampling, a confidence interval of +/- 2.6 percent is very good.

Table A-1 summarizes the standard errors and design effects for some variables and indexes in the survey. The design effect (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 clustered design had a greater variance 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.

---

[1] This section and the following were adapted from the Costa Rica 2004 report, "*Democratic Culture, Citizen Security and Social Capital in Costa Rica*," by Luis Rosero-Bixby and Jorge Vargas-Cullell.

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

```
pweight:  <none>                              Number of obs(*) =      1487
Strata:   estrapri                           Number of strata =         6
PSU:      upm                                Number of PSUs   =       165
                                             Population size  =      1487


-------------------------------------------------------------------------
    Mean |   Estimate    Std. Err.   [95% Conf. Interval]       Deff
---------+---------------------------------------------------------------
      q2 |   36.94956    .2429573    36.46972    37.4294     .4218436
      ed |   8.653432    .1987666    8.260869    9.045995    3.111623
  wealth |   4.749159    .1148371    4.522357    4.975962    5.038581
    psa5 |   59.86464    .7575729    58.36844    61.36085    1.803944
     tol |   52.47252    .9958056    50.50581    54.43923    1.871466
  exctot |   .1509817    .0132412    .1248304    .1771331    1.16062
    vb2r |   .6452268    .0141062    .6173672    .6730864    1.283041
    it1r |   62.14689    .9869107    60.19775    64.09604    1.576537
     m1r |   67.37145    .6988318    65.99126    68.75164    1.52385
    ls3r |   78.68557    .781379     77.14235    80.22879    1.431366
    ing4r|   69.97636    .9332012    68.13329    71.81943    1.46954
   coups |   36.07397    1.329585    33.44804    38.69989    1.899064
-------------------------------------------------------------------------
(*) Algunas variables contienen valores de omisión.
```

According to the above table, the cluster design for this survey was very efficient. In fact, with the exception of education, wealth, and tolerance, all the DEF were close to 1. The standard errors for most variables were also very moderate. Table A-2 shows the standard errors and DEF for the variable q2 (age) by cluster (region).

Table A-3, in turn, does the same for the Tolerance index. The DEF, as well as the standard errors, indicate that the cluster design by for the regions was more efficient than a simple random design.

Table A-2 Standard Errors and Design Effects for Age by Region

```
pweight:  <none>                              Number of obs    =      1487
Strata:   estrapri                           Number of strata =         6
PSU:      upm                                Number of PSUs   =       165
                                             Population size  =      1487


-------------------------------------------------------------------------
Mean    Subpop. |   Estimate    Std. Err.   [95% Conf. Interval]       Deff
----------------+--------------------------------------------------------
q2              |
      Atlántica |   36.96129    .4678184    36.03735    37.88523    .3278206
         Bogotá |   37.50862    .6836278    36.15846    38.85878    .501808
        Central |   37.04469    .5383934    35.98137    38.10802    .4674513
       Oriental |   36.21111    .533673     35.15711    37.26511    .4008669
        Pacífica|   36.58935    .4313391    35.73746    37.44125    .2390172
Ant. Tr. Nales  |   39.2963     1.892138    35.55933    43.03326    1.103853
-------------------------------------------------------------------------
```

Table A-3 Standard Errors and Design Effects for Tolerance by Region

```
pweight:  <none>                              Number of obs   =      1459
Strata:   estrapri                           Number of strata =        6
PSU:      upm                                Number of PSUs  =       165
                                             Population size  =      1459


--------------------------------------------------------------------------
Mean    Subpop. |   Estimate   Std. Err.   [95% Conf. Interval]      Deff
--------------+-----------------------------------------------------------
tol             |
     Atlántica  |   57.2374    2.587109    52.12787    62.34693    2.62035
       Bogotá   |   50.87382   1.865198    47.19006    54.55758   1.000855
      Central   |   53.1689    2.169709    48.88373    57.45407   2.065211
      Oriental  |   52.01894   2.075452    47.91993    56.11794   1.531916
      Pacífica  |   49.29745   2.001495     45.3445    53.25039   1.459799
Ant. Tr. Nales  |   44.97863   6.783489    31.58127      58.376   3.524574
--------------------------------------------------------------------------
```

Sample Results and Description of the Respondents

The probabilistic design of the sample, as well as the availability of a good sampling frame, are fair conditions to expect that the interviewed group is representative of the Colombian population. However, due to the effects of random errors and inevitable distortions of the sample design, the sample could deviate from the characteristics of the population it represents. It could include biases that should be reported. Table A-4 allows us to answer the question: how representative is the sample of the population? Below we compare some characteristics of the sample with the 1993 census.[2]

Table A-4 Sample vs. 1993 Census (18 years or older)

Características	Censo de 1993	Encuesta en Colombia
(N)	26,735,000	1,487
% de hombres	49	49
% > 30 años	42	60
% soltero	40	32
% casado o en unión libre	50	59
% con primaria	44	27
% con secundaria	34	52
% con educación post-secundaria	11	17
% en Región Atlántica	22	21
% en Bogotá	15	16
% en Región Central	24	24
% en Región Oriental	18	18
% en Región Pacífica	18	18
% en Antiguos Territorios Nacionales	3	4

---

[2] Projections for more recent years are not available for most indicators analyzed here.

We observe that there is congruity between the sample of this survey and the 1993 Census. Some characteristics such as age, gender, 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, rising from 34 percent in the 1993 census to 52 percent in the 2004 survey. The explanation for the gaps in marriage and education variables may lie in the fact that the census is 12 years old. This is a long enough period for vast changes to occur in these variables. Unfortunately, there are no reliable projections for the education variables of the 1993 census for the adult population. There are some education projections which include current students, but this makes them a non-comparable statistic for this survey.

Because in general, the sample is representative of the population, there is no need to use weights. Therefore, the sample is self-weighted.

Table A-5 compares sample characteristics between men and women.

Table A-5. Sample Characteristics by Gender

Característica	Total	Hombres	Mujeres
(N)	1,487 (100%)	49.2%	50.8%
Edad promedio	36.9	37.3	36.6
% casado o en unión libre	59.5	60.1	58.8
Ingresos familiares (Moda) % entre 181,000 y 360,000 pesos	25.9	25.0	24.0

We used gender and age quotas to select respondents. Therefore, our percentages of men and women are very similar to each other. Their ages are also very similar, differing by only one year. There is a slight difference with respect to the percentage of married or in civil union, where men have a slightly higher percentage (56.7%) than women. In terms of family income, there is a higher percentage of women in the 181,000 to 360,000 pesos bracket, which corresponds to the family income mode, although the margin is very slim (2.6 percentage points).

Because of its 20-year experience as well as its leadership in complex, country-wide operations, the Centro Nacional de Consultoría (CNC), was chosen to develop the sampling process, the field interviews, and data entry, verification, and processing.

Technical Description of the Sample Design

Universe

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

The universe was divided in two sectors: one of cities with more than 300,000 inhabitants, 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 military bases. Private households in these areas were contemplated.

Final Selection Unit

Because the questionnaire included questions not 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 housing unit, sometimes shared by more than one household (often relatively stable over time), each housing unit was selected as the final selection unit.

Sampling Method

We chose the probabilistic, stratified, multistage method with randomized selection of units at each stage. First, the sample is stratified by town size (cities with more and less than 300,000 inhabitants), then by region and area (rural and urban).

It is multistage sampling because within each urban area, it starts with Primary Sampling Units (sectors), followed by Secondary Units (sections), then Third Units (blocks) and Final Sampling Units (clusters of housing units) of 6 to 8 in urban areas and 10 to 12 in rural areas. In each housing unit, the interviewer selected only one household as an Observation Unit.

The repondent was selected according to the age and gender 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 area, interviewers listed the first 20 housing units they encountered. They had instructions to do a minimum of 8 surveys of the 20 housing units listed, balancing the gender and age quotas.

The selection method was chosen according to the following considerations:

We needed Representative samples at the following levels:

Nationally

First Stage Strata:

Cities with more than 300,000 inhabitants
Cities with less than 300,000 inhabitants

Second State Strata:

Bogota
Atlantic Region
Pacific Region
Eastern Region
Central Region
Old National Territories

Third State Strata:

Urban Area
Rural Area

Study Domains:

Cities with more than 300,000 inhabitants (obliged selection)
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 interviews
c) We worked with the best and most up to date sampling frame available for each municipality (population census, cartography, current housing unit listings, among others)

Sampling Frame

The sampling frame 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 Estádistica (DANE; National Statistics Department).

Calculations by Strata

The sample is composed of 138 sampling points: 103 urban and 35 rural, distributed over 53 municipalities in 26 out of the 32 departments of Colombia.

Sample Sizes, Confidence Level, and Margins of Error

The confidence levels anticipated for the national sample was 95 percent, with a margin of error of 2.6 percent, assuming a 50/50 proportion in the dichotomous variables.

The margins of error for a confidence level of 95 percent are:

TAMAÑO DE MUESTRA Y MARGENES DE ERROR
Nivel de Confianza del 95%

Estratos Regiones	Tamaño de Muestra	Margen de error M.A.S.	% M.P.C.
Atlantica	323	5,45	5,58
Bogota	231	6,45	6,59
Central	358	5,18	5,29
Oriental	274	5,92	6,05
Pacifica	263	6,04	6,18
Territorios Nacionales	54	13,34	13,63
Areas			
Urbana	1107	2,96	3,01
Rural	396	4,92	5,03
Total Pais	1503	2,53	2,58

Non-Coverage Adjustment

To ensure the efficiency, sufficiency, and coverage of the sample, we adopted an Non-Coverage Adjustment system which guarantees a maximum level of error and a minimum confidence level. In addition, the system eliminates biases resulting from the replacement of household units.

The system applies to the sample size (n) of each stratum a non-coverage factor defined by:

$N* = (1+t) \, n$

Where:

T = rate of non-coverage

$N*$ = final sample size

We chose an average t of 0.18.

Thus,

$$N* = (1+0.18)\ 1{,}500 = 1{,}770\ \text{adults}$$

By the end of the field interviews, we ended up with a real sample size of 1,487 respondents, 21 less than expected due to law and order issues (see "Final Comments on Survey Fieldwork").

Survey Team

The CNC involved its five branches (Bogota, Cali, Medellin, Barranquilla and Bucaramanga) to ensure a high quality survey 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 areas visited, which complicated the operations.

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

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

Table A-6. Universe, Total Population by Region and Area (rural/urban)

Actividad	Total de personal
Coordinadores de campo	15
Supervisores	8
Encuestadores	45
Supervisores de calidad en campo	12
Codificadores	7
Digitadores	7
Verificadores de datos	7
Subtotal campo y digitación	101
Personal directivo y profesional	5
Personal administrativo	3
Total personal del equipo	**109**

COLOMBIA: POBLACION TOTAL, PROYECCION AÑO 2004

Region	Area		
	Urbano	**Rural**	**Total**
Atlantica	6.912.796	2.843.894	9.756.690
Bogota	7.014.111	1.517	7.015.628
Central	7.659.319	3.093.365	10.752.684
Oriental	5.207.407	3.106.566	8.313.973
Pacifica	5.322.949	2.644.004	7.966.953
Territorios Nacionales	583.895	890.830	1.474.725
Total	32.700.477	12.580.176	45.280.653

Distribucion Porcentual (%)

Region	Area		
	Urbano	**Rural**	**Total**
Atlantica	70,9	29,1	100
Bogota	100,0	0,0	100
Central	71,2	28,8	100
Oriental	62,6	37,4	100
Pacifica	66,8	33,2	100
Territorios Nacionales	39,6	60,4	100
Total	72,2	27,8	100

Table A-7 Size and Distribution of the Sample by Strata
TAMAÑOS Y DISTRIBUCION DE LA MUESTRA POR ESTRATOS
RESUMEN GENERAL

	Urbano	Rural	Total
Atlantica			
+ de 300,000 habitantes	127		127
- de 300,000 habitantes	112	84	196
Total Atlantica	239	84	323
Bogota			
+ de 300,000 habitantes	231		231
Total Bogota	231		231
Central			
+ de 300,000 habitantes	150		150
- de 300,000 habitantes	112	96	208
Total Central	262	96	358
Oriental			
+ de 300,000 habitantes	66		66
- de 300,000 habitantes	112	96	208
Total Oriental	178	96	274
Pacifica			
+ de 300,000 habitantes	91		91
- de 300,000 habitantes	88	84	172
Total Pacifica	179	84	263
Territorios Nacionales			
- de 300,000 habitantes	18	36	54
Total Territorios Nacionales	18	36	54
Total	1.107	396	1.503

RESUMEN GENERAL

	MUESTRA	UPM	PUNTOS DE MUESTREO
Regiones			
Atlantica	323	20	40
Bogota	231	1	33
Central	358	23	46
Oriental	274	20	34
Pacifica	263	16	34
Territorios Nacionales	54	6	6
Total	1.503	86	193

Table A-8. Selection of the Sample by Strata, Area and Municipality

Distribucion de la Muestra ciudades con 300,000 o mas habitantes

Region	Departamento	Municipio	Poblacion año 2004						Puntos de muestreo	Tamaño Punto de muestreo	Muestra			
			Total	Urbana	Rural	% Poblacion Total	% Urbano del Municipo	% Rural del Municipo			Total	Urbana	Rural	% de la Muestra
ATLANTICA	ATLANTICO	BARRANQUILLA	1.359.700	1.356.873	2.827	3,0	99,8	0,2	6	7	42	42		2,8
	ATLANTICO	SOLEDAD	336.190	334.135	2.055	0,7	99,4	0,6	2	6	12	12		0,8
	BOLIVAR	CARTAGENA	1.004.074	927.657	76.417	2,2	92,4	7,6	5	7	35	35		2,3
	CESAR	VALLEDUPAR	371.763	300.550	71.213	0,8	80,8	19,2	2	6	12	12		0,8
	CORDOBA	MONTERIA	343.607	268.345	75.262	0,8	78,1	21,9	2	6	12	12		0,8
	MAGDALENA	SANTA MARTA	434.937	418.630	16.307	1,0	96,3	3,7	2	7	14	14		0,9
SUBTOTAL			**3.850.271**	**3.606.190**	**244.081**	**8,5**	**93,7**	**6,3**	**19**		**127**	**127**		**8,4**
BOGOTA	BOGOTA	BOGOTA	7.029.928	7.014.111	15.817	15,5	99,8	0,2	33	7	231	231		15,4
			7.029.928	**7.014.111**	**15.817**	**15,5**	**99,8**	**0,2**	**33**		**231**	**231**		**15,4**
CENTRAL	ANTIOQUIA	MEDELLIN	2.071.391	1.978.967	92.424	4,6	95,5	4,5	10	7	70	70		4,7
	ANTIOQUIA	BELLO	390.012	383.144	6.868	0,9	98,2	1,8	2	6	12	12		0,8
	CALDAS	MANIZALES	378.965	355.368	23.597	0,8	93,8	6,2	2	7	14	14		0,9
	HUILA	NEIVA	367.811	344.544	23.267	0,8	93,7	6,3	2	6	12	12		0,8
	QUINDIO	ARMENIA	316.301	309.391	6.910	0,7	97,8	2,2	2	6	12	12		0,8
	RISARALDA	PEREIRA	510.739	430.355	80.384	1,1	84,3	15,7	2	8	16	16		1,1
	TOLIMA	IBAGUE	444.460	417.658	26.802	1,0	94,0	6,0	2	7	14	14		0,9
SUBTOTAL			**4.479.679**	**4.219.427**	**260.252**	**9,9**	**657,3**	**42,7**	**22**		**150**	**150**		**10,0**
ORIENTAL	CUNDINAMARCA	SOACHA	310.038	304.343	5.695	0,7	98,2	1,8	2	6	12	12		0,8
	META	VILLAVICENCIO	358.621	314.015	44.606	0,8	87,6	12,4	2	6	12	12		0,8
	NORTE DE SANTANDER	CUCUTA	722.485	702.213	20.272	1,6	97,2	2,8	3	8	24	24		1,6
	SANTANDER	BUCARAMANGA	568.136	562.340	5.796	1,3	99,0	1,0	3	6	18	18		1,2
SUBTOTAL			**1.959.280**	**1.882.911**	**76.369**	**4,3**	**381,9**	**18,1**	**10**		**66**	**66**		**4,4**
PACIFICA	NARIÑO	PASTO	422.350	373.405	48.945	0,9	88,4	11,6	2	7	14	14		0,9
	VALLE	CALI	2.369.696	2.342.141	27.555	5,2	98,8	1,2	11	7	77	77		5,1
SUBTOTAL			**2.792.046**	**2.715.546**	**76.500**	**6,2**	**97,3**	**2,7**	**13**		**91**	**91**		**6,1**
TOTAL			**20.111.204**	**19.438.185**	**673.019**	**44,4**	**96,7**	**3,3**	**97**		**665**	**665**		**44,2**

Geographic Location of the Sample
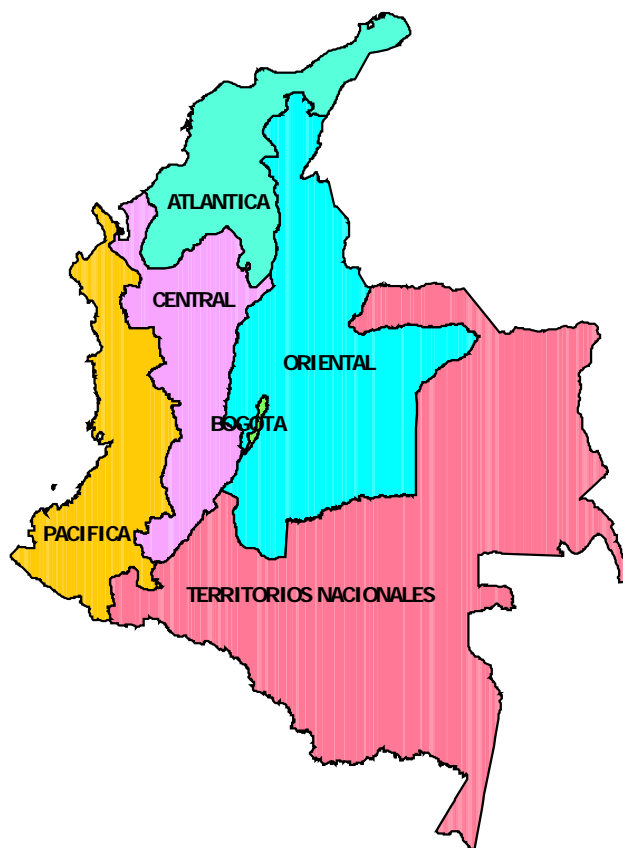
**Cuadro No 1**
Distribucion de la Muestra ciudades con 300,000 o mas habitantes

Region	Departamento	Municipio	Muestra Total
ATLANTICA	ATLANTICO	BARRANQUILLA	42
	ATLANTICO	SOLEDAD	12
	BOLIVAR	CARTAGENA	35
	CESAR	VALLEDUPAR	12
	CORDOBA	MONTERIA	12
	MAGDALENA	SANTA MARTA	14
SUBTOTAL			127
BOGOTA	BOGOTA	BOGOTA	231
			231
CENTRAL	ANTIOQUIA	MEDELLIN	70
	ANTIOQUIA	BELLO	12
	CALDAS	MANIZALES	14
	HUILA	NEIVA	12
	QUINDIO	ARMENIA	12
	RISARALDA	PEREIRA	16
	TOLIMA	IBAGUE	14
	CUNDINAMARCA	SOACHA	12
SUBTOTAL			162
ORIENTAL	META	VILLAVICENCIO	12
	NORTE DE SANTANDER	CUCUTA	24
	SANTANDER	BUCARAMANGA	18
SUBTOTAL			54
PACIFICA	NARIÑO	PASTO	14
	VALLE	CALI	77
SUBTOTAL			91
TOTAL			665

Distribucion de la Muestra ciudades con menos de 300,000 habitantes

Region	Departamento	Municipios	Muestra
ATLANTICA	BOLIVAR	SAN JUAN NEPOMUCENO	28
	CESAR	SAN MARTIN	28
	CORDOBA	AYAPEL	28
	CESAR	PELAYA	28
	MAGDALENA	CERRO SAN ANTONIO	28
	SUCRE	MAJAGUAL	28
	SUCRE	CHALAN	28
SUBTOTAL			196
CENTRAL	ANTIOQUIA	CALDAS	26
	ANTIOQUIA	ABEJORRAL	26
	ANTIOQUIA	MARINILLA	26
	TOLIMA	ICONONZO	26
	CALDAS	MANZANARES	26
	TOLIMA	COYAIMA	26
	RISARALDA	SANTUARIO	26
	RISARALDA	QUINCHIA	26
SUBTOTAL			208
ORIENTAL	BOYACA	MUZO	26
	BOYACA	TUNJA	26
	CUNDINAMARCA	FACATATIVA	26
	CASANARE	MONTERREY	26
	META	GUAMAL	26
	BOYACA	SUTAMARCHAN	26
	SANTANDER	CONFINES	26
	SANTANDER	RIONEGRO	26
SUBTOTAL			208
PACIFICA	CAUCA	POPAYAN	24
	CAUCA	PADILLA	24
	NARIÑO	CUMBITARA	26
	NARIÑO	SANDONA	24
	VALLE	BUGA	26
	VALLE	FLORIDA	24
	VALLE	RIOFRIO	24
SUBTOTAL			172
TERRITORIOS NACIONALES	CAQUETA	FLORENCIA	18
	PUTUMAYO	PUERTO ASIS	18
	VAUPES	VILLA FATIMA*	18
SUBTOTAL			54
TOTAL			838
GRAN TOTAL			1.503

MUESTRA

ATLANTICA

CENTRAL

ORIENTAL

BOGOTA

PACIFICA

TERRITORIOS NACIONALES
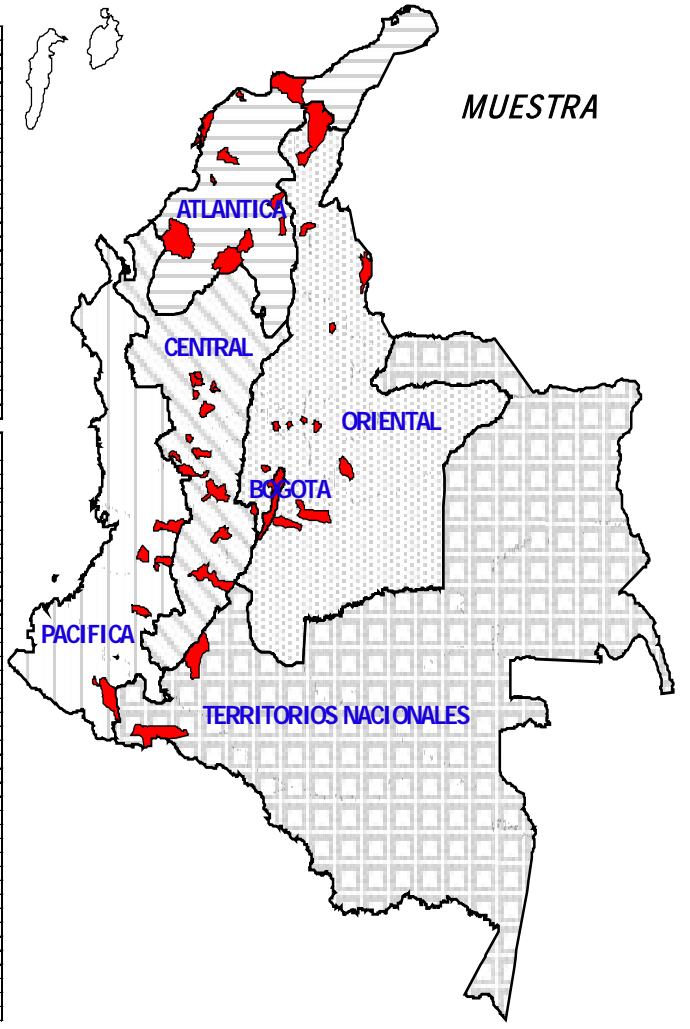
Final Comments on Survey Fieldwork

About the questionnaire: the questionnaire was long, 50 minutes on average, but in general the respondents were willing to answer the questions and we had very few uncompleted interviews.

Due to Colombia's internal conflict, some questions were especially delicate. Some interviewers mentioned that the series on page 8 of the questionnaire generated certain discomfort among some respondents. But to our surprise, we encountered very few refusals to answer those questions.

About the fieldwork: For security reasons, it is was not possible to do 12 surveys in Chalán (Sucre) and 4 in Muzo (Boyacá).

In the rest of the country, some of our interviewers were stopped and interrogated by the illegal armed groups, who finally permitted them to continue their work.

Despite the fact that some respondents were located in areas with a strong presence of illegal armed groups, there was no reports of any kind of pressure to induce answers from respondents. On the contrary, the interviewers emphasized the free will of those who agreed to be part of the study.

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