

DESIGN OF THE SAMPLE

LUIS CARLOS GOMEZ

The universe of the National Health Survey was the civilian non-institutional population of the departments—an intermediate political administrative division—whose surface covers 52.7 per cent of the country, but which contained 98.7 per cent of the total population. Because of their low population density (1.3 per cent of the population) and their rudimentary communication, extensive geographical areas of the country (47.3 per cent of the country's surface) did not become part of the sampled universe. The inclusion of this 1.3 per cent of the population would have involved enormous procedural difficulties and a disproportionate increase in the costs.

GENERAL PROCEDURE

The design of the sample was carried out in two stages.

In the first stage 716 primary sample units were defined and grouped into 40 strata. One unit was selected from each stratum.

In the second stage, the selected primary sample units were subdivided into segments of approximately ten dwellings per unit from which 24 were selected. The 960 segments composed the sample for the household interviews. For the clinical examination, a subsample of approximately one of every ten interviewees was taken.

In Table 1, the size of the samples of the National Study of Colombia and the National Health Poll of the United States are compared.

The Primary Sample Units

The primary sample units were composed of municipalities with a

TABLE I. COMPARISON OF SAMPLE SIZE, COLOMBIA AND UNITED STATES

	<i>Colombia</i>		<i>United States</i>	
	<i>Per Cent of Total</i>		<i>Per Cent of Total</i>	
	<i>N</i>	<i>Population</i>	<i>N</i>	<i>Population</i>
Total population	18,000,000		18,000,000	
Persons interviewed	52,000	0.3	180,000	0.1
Persons examined:				
Adults	2,600	0.014	7,200	0.006
Children	2,600	0.014	21,600	0.036
Surface covered by the selected prime units.				
Interviews	350/1900	20.0	40/700	6.0
Examinations	40/1900	2.0	40/700	6.0

In Colombia a person who was interviewed represented 340 of the universe under study, and a person who was examined represented 3,400.

population of more than 5,000 and with an available hospital, health center or other health service. The municipality was selected as the primary sample unit because it is the smallest political-administrative division about which information is available and which offers operative advantages. The limit of 5,000 population was established to guarantee a minimum probability of selection of 0.01 per unit. The existence of a health service was to guarantee facilities for conducting the study.

Cities with fewer than 5,000 inhabitants or without some health service were annexed to the nearest municipality that fulfilled both of these requirements. Since the population of a stratum should be approximately 450,000, the cities that reached or surpassed that limit were considered to have a number of primary sample units proportionate to their population.

Regional and Strata Division

The departments of the country were grouped into four regions to establish the prime units for the presentation of the results in the study, and the calculation of the reliability of the estimates. Each region with a minimum population of three million contains departments with similar ecological characteristics so that regional estimates were sufficiently reliable for the majority of the data.

Bogotá, the capital of the country, because of its special characteristics, was separated as a region to become an individual area of tabulation since this circumstance was considered to compensate for its

smaller population. Five tabulation areas were formed for the four regions of the country.

Within each region, the primary sample units were arranged in 40 strata so that they would have the maximum homogeneity among the grouped units and the greatest heterogeneity between one stratum and another from the point of view of the size of population, percentage of urban population and average altitude.

Ten of the strata classified as "Defined" coincide with the prime unit, and therefore had an assured selection. These strata correspond to the prime units in the most populous municipalities of the country. The other 30 "Undefined" strata were formed from the remaining prime units.

Sample of the Prime Unit of the Undefined Strata

A prime unit was selected as the representative of each of the remaining 30 strata. The probability of being selected was proportional to its population within the strata.

The "controlled selection" technique was used in the process. The use of controls increased the probability of selecting preferred combinations of prime units beyond those obtained in the stratified sample without hampering the real probability of each individual unit.

The following control criteria were used:

Groups of departments formed within each tabulation area considering the homogeneity of its socioeconomic and geographic characteristics and a minimum population of 1.3 million.

Index of medical attention services, simultaneously considering the number of beds available per 1,000 inhabitants and the size of the hospitals.

Subdivision of the Sample into Representative Semisamples

The entire sample was subdivided into two semisamples of the 20 most similar sampling units, so that, independently, results that could be generalized to the country might be obtained from them. This necessitated a security mechanism in case the operative contingencies might hamper the development of the study in all of the selected units. On the other hand, the subdivision facilitated the calculation of the validity of the results.

Selection of the Local Sample

The second stage of the design dealt with the procedures that fol-

lowed the selection of the prime units and that were verified for each one of them.

The organization and treatment of the basic information about dwellings and cartography was somewhat unusual:

1. Dwellings: Information about the number of completed census tickets in the sectors of the 1964 census was obtained for each of the 40 prime units. Since the information refers to buildings and not to dwellings, it was necessary to make the conversion to the latter measurement by applying adjustment factors established by a rapid revision of the corresponding census tickets to each municipal district.

2. Cartography: Cartographic information about Colombia is still general and in some cases deficient and outdated. Only municipal districts fulfilled the requirements necessary for a rapid sample selection. Therefore, all of the selected rural areas had to be visited beforehand to complete the necessary subsamples and complementary cartographic framework that would permit their location at the time of the study. This work, called "segmentation," decisively influenced the general procedure of the study because of its magnitude and importance.

Once the information necessary for the selection was available in an adequate form, it was completed in various steps both at the central level and at the local level.

Procedure at the central level. The number of theoretical segments of an "average size" of ten dwellings was established for each minimum census sector. (The minimum census sectors were established by blocks in municipal districts and in rural zones, by areas of approximately 100 dwellings.) A systematic selection of 24 "size samples" (units of ten dwellings) was made, and the census sectors in which they were located were identified.

In the second step, the selected sectors were outlined on a map of the unit. Those corresponding to municipal districts offered no difficulty and were later ready for a rapid verification in the "segmentation." Those in rural areas were finally subdivided to obtain the smallest geographical unit that could later be easily managed.

Procedure at the local level. The blocks in the municipal districts, and the rural areas selected at the central level were submitted for verification and a complementary cartographic framework immediately prior to the field work. At that time, rural areas of more than 15 dwellings were subdivided to facilitate the later interview. The work was

completed at the departmental and local levels by health workers necessarily trained in the territory, with precise written instructions and with adequate supervision by central level officials.

When the segmentation of each unit was completed, all the documents were sent to the central level where they were reproduced or included in the folders of the interviewers.

When the selected municipal sectors had more than 15 dwellings, they were subdivided by the interviewers immediately before beginning the interviews, and under the direction of the field supervisor.

The subsample of persons for the clinical examination was taken while conducting the interviews using a list of persons interviewed to control distribution by age.

Table 2 refers to the selected and covered sample in the household interview and the clinical examination.

REAPPLICATION AND ADJUSTMENT TO THE UNIVERSE

To obtain correct estimates from the sample, it was necessary to inflate or weigh the studied units with the reciprocal of its probability of being selected to recreate the universe, and to adjust the reconstituted universe to the predictions of the 1964 census.

Restitution. Basic weight of inflation. Each family and individual selected for the home interview had a basic weight of about 340, indicating the number of people in the universe represented by one person in the sample, and equal to the reciprocal of his selection probability.

The basic weight of the people selected for the clinical examination

TABLE 2. SELECTED AND COVERED SAMPLE

	<i>Home Interviews</i>			<i>Clinical Evaluation</i>	
	<i>N</i>	<i>Per Cent</i>		<i>N</i>	<i>Per Cent</i>
Families selected*	8,920	100.0	Persons selected	5,258	100.0
Families interviewed	8,669	97.2	Persons examined	5,027	95.6
Families not interviewed**	251	2.8	Persons not examined	231	4.4

* 9,798 dwellings were selected of which 878 corresponded to families with permanent residence in another site, vacant houses and other causes; therefore the number of families actually selected was 8,920.

** The 251 families not interviewed are those who refused or who were momentarily or temporarily absent.

was equal to their weight in the interview times the reciprocal of the fraction of the subsample. This number fluctuated around 3,400.

Special weight of the subsample. When the segment was subdivided, the dwellings and persons in the sample received another weight or "special weight," which was the reciprocal of the fraction of the subsample. When the number of dwellings was less than anticipated, the weight was calculated from the relation: dwellings found/dwellings anticipated.

Adjustment of coverage. Families selected but not interviewed, and individuals chosen but not examined received a special treatment that consisted of attributing to them the replies of similar families and individuals.

Justification of estimates. The population reconstructed from the sample for the home interview and the subsample for the clinical examination were adjusted to the urban and rural totals for the 1964 census. This adjustment, even if it might slightly slant the results, reduces the variability of the estimates, corrects the lack of coverage for families not located, assures consistent population bases and overcomes the deficiencies of estimated population used in the selection of prime units.

MEASURE OF VARIABILITY

By dealing with a random sample, it is possible to obtain correct estimates of the variability of error in the sample. The variability of the sample, the interviewers or examiners, can be measured as a total, or separately for each one of the components. Nevertheless, the estimate of total variability was used to measure the trustworthiness of the results.

TABLE 3. COMPARISON OF OBSERVED AND EXPECTED RATES AND STANDARD ERRORS

	<i>Standard Error</i>	<i>Estimated Rate</i>	<i>Confidence Limits at 95 per cent Level</i>	
			<i>Lower</i>	<i>Higher</i>
Expected rate	.009	10.00	8.2	11.8
Obtained rate	.005	10.9	9.8	12.0
<i>Tabulation Area 4</i>				
Anticipated rate	0.018	10.0	6.4	13.6
Obtained rate	0.012	12.1	9.7	14.6

TABLE 4. DISTRIBUTION BY AGE AND SEX IN THE 1964 CENSUS, HOUSEHOLD INTERVIEW AND CLINICAL EXAMINATION

Age (in Years)	Men			Women		
	Census %	Household Interview %	Clinical Examination %	Census %	Household Interview %	Clinical Examination %
0-14	47.9	49.7	50.6	45.4	45.0	45.5
15-44	39.0	36.4	35.2	41.3	39.9	40.2
45 and above	13.1	13.9	14.2	13.3	14.2	14.3

TABLE 5. DISTRIBUTION BY SEX IN THE 1964 CENSUS, HOUSEHOLD INTERVIEWS AND CLINICAL EXAMINATION

	Census %	Household Interview %	Clinical Examination %
Men	49.3	48.3	47.3
Women	50.7	51.7	52.7
Total	100.0	100.0	100.0

The obtained variability is sufficient for the purposes of the study, and somewhat lower than anticipated in the design of the sample.

The difference between the standard errors is important. The national rate obtained for the number of persons incapacitated for two weeks is almost one-half the expected error for any characteristic of similar frequency, as shown in Table 3.

For Tabulation Area 4, whose population is five million, the difference between the standard errors anticipated and those obtained is still considerable.

REPRESENTATIVENESS OF THE SAMPLE

The design of a sample is efficient insofar as it represents the characteristics of the universe in the combination of unlisted units. Nevertheless, this analysis is practical only for those characteristics about which information already exists in other sources such as the population census or various systematic reports.

Table 4 shows the percentage distribution by age for each sex in the 1964 census, the home interview and the clinical examination.

The largest difference, 3.8 per cent between the census and the clinical examination, occurs in the group of men between ages 15 and 44. This is explained by the exclusion of the institutional population from the study and by the size of the interval in the age group. The rest of the differences vary around 0.1 and 2.7 per cent.

The minimal differences between the interviewed and examined population demonstrates the effectiveness of the system adopted for taking the subsample of persons for the clinical examination.

The differences in the percentage distribution according to sex are of even less importance, as seen in Table 5.

APPENDIX

Selection Process for the Sample for the National Health Survey

I. Division of the country into 716 prime sample units.

Political administrative division and number of units per department.

Universe: 716 prime sample units in 18 departments; 98.7 per cent of the total population of the country; 52.7 per cent of the total surface.

Criteria:

1. political administrative division.
2. population.
3. minimal health resources.
4. accessibility.

Note: Bogotá, Medellín, Cali, Barranquilla and Bucaramanga were treated independently.

II. Grouping of the prime units into 40 strata.

Regional division and number of strata per division.

Criteria:

1. population of approximately 450,000.
2. homogeneous characteristics within each stratum: total population, per cent of urban population and average altitude.
3. combination of prime units from the same region.

III. Selection of 40 prime units, one prime unit per stratum.

Criteria:

1. selection controlled by two criteria: groups of departments and index of health services.
2. probability of selection of each unit, proportional to the volume of population within the stratum.

IV. Selection of 960 segments for the home interview. One segment per ten dwellings; 24 segments per unit; 9,600 dwellings = 52,500 persons.

Criteria:

1. proportionality of the number of urban and rural segments established by the size of their populations.
2. systematic selection.

V. Selection of subsample of 5,250 people for the clinical examination. 130 persons per unit, approximately five persons per segment.

Criteria:

1. systematic selection among people who were interviewed.
2. control of the distribution by age.