

SAMPLE DESIGN – MEXICO METAGORA PILOT SURVEY

1. Coverage

Target Population

Population targeted in this survey is all inhabitants of the Federal District of Mexico, aged 15 years or older that is, Mexico City without the rest of the Metropolitan Area.

Survey Population

For the purpose of this survey, we included only inhabitants who are 15 years old or older living permanently in households within Mexico City. This population, according to the 2000 census, is 6.4 million inhabitants. During the survey, we decided to exclude the population of the borough of Tlahuac.¹

Reference Period

The questionnaire asks respondents for all the contacts they had with the authority within one year. To reduce a potential recall error, the survey places the respondent on the last year previous to the data collection period using month-to-month reference (i.e. from November 2003 to October 2004).

Key Variables to be measured

The key variables to measure are contact with the authorities and situations of ill-treatment occurring during these contacts. There are no existing measures for these variables, neither on incidence nor for its variance.

2. Pilot Test

A pilot test was conducted from September 28th to October 19th. The purposes of the test were

- proving the questionnaire,
- testing the fieldwork procedures,
- estimating the non-response rate,
- estimating the DEFF,
- obtaining estimates of both value and variance for the key variables: contact with the authority and ill-treatment .

The sample for the pilot test was selected using simple random selection from the list of AGEBs. By design, a selection of 2 AGEBs for each socioeconomic level was made. The AGEBs were previously grouped by socio-economic level and then a simple random selection was made from the list of AGEBs. Within each AGEB, a random selection of 4 blocks was made in the office. Then, a random selection of 3 dwellings by block was carried out.

The pilot test included 18 AGEBs and 216 dwellings selected. Most important findings were:

Incidence of contact with the authorities for population 15+	39%
Expected error (at 95% confidence level)	5.7%
Non-response (including households not contacted or out of reach)	69%
Design Effect (DEFF)	9.424

¹ During the fieldwork a tragic event occurred in Tlahuac where two police officers were lynched by local population. Given the impact of this event, we decided to exclude the area from the sample.

3. Sample Design.

To estimate the sample size, we used both a budget restriction and the results obtained from the pilot test, i.e., non-response rate, incidence of contact with the authority, variance of the variable contact and design effect.²

The sample selection included 8,688 dwellings. To achieve this size, we used the following formula:

$$n = \frac{z^2 * P * Q * DEFF}{r^2 (1 - TNR)}$$

where:

- n = Sample size (dwellings).
- z = Value in a normal distribution that assures a established confidence level
- P = Estimated value for the incidence of contacts with the authority.
- Q = (1-P).
- DEFF = Design Effect.
- TNR = Maximum expected non-response rate (including household not contacted or out of reach)
- r = Maximum expected error

These 8,688 dwellings were selected in **362** AGEB's.

Sampling Frame

We used SCINCE from INEGI (Sistema de Consulta para la Información Censal). This software is a list of the basic geographical maps called Áreas Geoestadísticas Básicas (AGEB). An AGEB is a basic unit area map created by the National Statistics Institute (INEGI) that include around 2,500 inhabitants with a size that ranges from 25 to 50 blocks for most of the maps of the Federal District.

The SCINSE includes the maps with a group of information obtained from the Population Census: population by AGEB, number of dwellings, education levels, income levels, etc. The maps for every AGEB allow the geographical reference of the blocks, their random selection and printing in the office.

Sample Selection

A multi-stage sampling was carried. First, we created strata from the official geographical maps (AGEBs). Then, we selected AGEBs from all stratum using probability proportional to size of the AGEB based on the number of inhabitants who are 15 years old or older (PPT). Then, a selection of 6 blocks was carried randomly in the office. Next, a field census of each of the selected blocks was carried out and, from that Census; we randomly selected 4 households in each block. Inside of the household, the respondent was selected using the last birthday method.

Stratification

The primary sampling units, AGEBs, are very heterogeneous in terms of the conditions of the population. The structure of Mexico City is such that within one AGEB there may be extremes types of households in terms of income, education and other variables. This heterogeneity was highly evident during the pilot test and we created stratum to reduce the variance of certain variables. We ended up with 20 stratum using the following variables reported by AGEB:

- Socio-economic level classification of the AGEB according to AMAI rules (Mexican Association of Research Agencies)
- Share of the Employed Population who earn less than one minimum wage

² There is no existing data that may be used for this variance

- Share of the Employed Population who earn more than five minimum wages
- Share of the population 18+ without college education

Sample Allocation

The sample was allocated among strata proportional to the number of AGEBS within each stratum. Within stratum, the AGEBS were selected proportional to the size of the population 15+ living in the AGEBS. The following table shows the resulting distribution by AGEBS.

STRATUM	AGEBS in stratum	%	AGEBS in sample	STRATUM	AGEBS in stratum	%	AGEBS in sample
1	164	7	26	11	326	14	52
2	139	6	22	12	193	8	31
3	63	3	10	13	62	3	10
4	9	0	1	14	485	21	77
5	36	2	6	15	12	1	2
ABC+	311	18	65	D+	1078	47	172
6	106	5	17	16	105	5	17
7	128	6	20	17	61	3	10
8	39	2	6	18	103	5	16
9	82	4	13	19	4	0	1
10	108	5	17	20	53	2	8
C	463	22	73	D/E	326	15	52
TOTAL AGEBS DF		2278	TOTAL AGEBS in sample				362

Selection Procedure

Selection of Primary Sampling Units (PSU)

Inside each stratum primary units were selected with probability proportional to the size of the population 15+ in the unit (AGEBS)

Selection of Secondary Sampling Units (Blocks)

The AGEBS maps have a draft of a list of all blocks inside of the area. The blocks were numerated and a random selection of six blocks was made with equal probability through a systematic procedure (random number tables)³

Selection of Dwellings

Inside the selected blocks, we carried out a census of each block counting the total number of dwellings in the block.⁴ Then, the random selection of 4 dwellings was done using a random number table.

Selection of Households

In cases where more than one household was found inside of a dwelling, the interviewers were instructed to carry-out a pre-established random selection of the household⁵. The criterion more commonly used was applied to distinguish households, i.e., the possession of separate cooking facilities.

Selection of Respondent

For every household, a list of the members who were 15 years old or more was elaborated. The list included several data: gender, age and date of birth. The member who had the most recent birthday was chosen as informant.

³ Cf. Appendix A3: *Random Number Table for the selection of Blocks and Dwellings*

⁴ Cf. Appendix A7 *Format to Register the Dwelling Census*

⁵ Cf. Appendix A4, *Random Number Table for the selection of Households*

4. Expansion Factors

The expansion factor for the respondent w in the household l in the dwelling k corresponding to the block i , inside of the PSU j in the strata h is:

$$F_{wlkijh} = \frac{1}{n_h} * \frac{P_h}{P_{jh}} * \frac{M_{jh}}{m_{jh}} * \frac{V_{ijh}}{v_{ijh}} * HO_{kijh} * IN_{lkijh}$$

where:

- P_{jh} = Population 15+ years old in the PSU j within stratum h
- P_h = Population 15+ years old in the stratum h
- n_h = Number of PSU selected within the stratum h
- M_{jh} = Number of blocks in the PSU j in the stratum h .
- m_{jh} = Number of blocks selected in the PSU j in the stratum h .
- v_{ijh} = Number of dwellings selected in the block i in the PSU j in the stratum h .
- V_{ijh} = Number of dwellings in the block i in the PSU j in the stratum h .
- HO_{kijh} = Number of households in the dwelling k in the block i in the PSU j in the stratum h .
- IN_{lkijh} = Number of inhabitants in the household l in the dwelling k , in the block i , in the PSU j in the stratum h .
- F_{wlkijh} = Expansion factor for the respondent w in the household l , in the dwelling k , in the block i , in the PSU j in the stratum h .

Weighting

Weights were used for making adjustments for non-response. Weights were created to adjust the surveyed population to have the same distribution as the target population on gender, age and socio-economic level:

$$WA_{wlkijh} = F_{wlkijh} * FA$$

