The Census in Mexico; its history, recent developments, achievements and challenges

Carole Schmitz *
Instituto Nacional de Estadística y Geografía (INEGI), Aguascalientes, Mexico –
carole.schmitz@inegi.org.mx

Brief historical overview of the census organisation and methodological developments in Mexico. Criteria and procedures to enhance the quality of the process and data for the 2010 census: planning, data collection, data processing, post enumeration survey and analysis of main variables. Main results of inter-census survey 2015, including recent experience with the use of technologies to improve the quality of the data and reduce the time for presentation and dissemination of the results. Methodological perspective and challenges for 2020 census.

Keywords: population; dwellings; coverage; information quality.

1. Historical Overview

The first population count in Mexico dates back to the Pre-Columbian era, according to codices and monuments, in the year 1116, 3.2 millions of inhabitants were counted. Afterwards, during the Aztec empire, population counts were conducted for controlling the payment of tributes. During the XVI century, two population censuses and five demographic counts took place; from the colonial era, the so called "Revillagigedo" Census (1790-1791) stands out. In 1895, the first General Census of the Mexican Republic was carried out as a national statistical exercise. Since then, the population censuses have been carried out every ten years, in the years finishing by zero, with an exception in 1920 due to the Mexican Revolution. The most recent census took place in 2010, and additionally to the censuses, in 1995 and 2005, population counts were carried out, for which a full enumeration of the population and dwellings was conducted. For 2015, this methodology changed and instead of a full population count, an Intercensal survey was carried out, with a sample size of nearly 6 million dwellings and from which the population of the country was estimated in around 119 million persons.

Since 1930, Mexican population and housing census has been taken on a de jure basis, through a face to face interview using printed questionnaires. For the last two population and housing censuses, two kinds of questionnaires have been used, a short form for regular enumeration, and a long form to be collected in a probabilistic sample of dwellings that allows to obtain more detailed information of the socioeconomic characteristics. Although the data collection period has being increasing over time, one day in 1980, one week in 1990, two weeks in 2000 and four weeks in 2010; advances in technology have made possible to reduce the time for results release, going for example from 24 months in 1990 to 9 months in 2010. Even though modern technology has not been used directly for data collection, it has allowed making the processes more efficient and reducing the time for results publication, particularly by improving the monitoring and supervision systems that contribute the coverage and data quality.

2. Quality in population and housing census 2010

As well as methodologies and census processes, quality assurance processes have evolved. The National Institute of Statistics and Geography (INEGI), created in 1983 and consolidated as an independent public body in 2008, has sought from its beginnings to provide quality information to the society and the state. Accordingly, in 2014 the Governing Board approved the Quality Assurance Standard, and in 2015 the Institutional Quality Policy, which establishes the principles of quality in statistical and geographical products: Pertinence, Accessibility, Opportunity and Punctuality, Consistency and Comparability, Standardized Metadata and Truthfulness (Accuracy and Reliability). Although the quality of products and statistical processes has several dimensions, this document will focus on the principle of truthfulness

and, in particular, when dealing with population and housing censuses, in terms of coverage and data quality, taking as reference the 2010 Census and some experiences of the Intercensal Survey 2015.

The following subsections briefly explain the strategies used in each of the stages of the 2010 census project and that seek to contribute in the fulfillment of the institutional quality policies previously mentioned.

2.1 Planning

Prior to the collection of information, several tests are carried out on census questionnaires, operational processes and data processing; also, training strategies are designed with special emphasis on aspects of coverage insurance and topics difficult to collect. The control and data gathering are carried out based on the national geostatistical framework, which INEGI continuously updates, in particular before the census is taken, an update is made that ensures that a geostatistical framework and corresponding cartography is available in order to guarantee a complete coverage of the national territory.

2.2 Data collection

The operational structure is established considering control sections that allow an adequate distribution of workloads, and considering procedures for a continuous monitoring and supervision. Recruitment is carried out through a web system, in which all candidates for each operational position are subjected to various filters, including a post-training evaluation, with the aim of recruiting the best elements in each position of the designed operating structure. During the census taking, monitoring is done mainly through the use of computer systems, which provide information in real time at all levels of the structure. Among the systems that are used for the field operation, stands out the "operational risk map", which presents the geographical areas with some kind of risk (insecurity, terrain conditions, weather conditions, etc.), and allows decision making during planning and even during the operation, with the purpose of reducing, as much as possible, under coverage.

Once the questionnaires are collected in the field, and prior to being sent to the capture area, a manual validation process is carried out, in which an operative agent reviewed the appropriate filling for some basic sections of the questionnaire (geographical identification, list of persons, sex, age, fertility, economic characteristics, international migration), in case of inconsistencies or lack of information, the questionnaires are returned to field for correction.

For the censuses prior to 2010, the task of revision of census result for the dwellings was assumed only by staff belonging to the supervisory structure of the enumeration itself. For the 2010 census, for the first time, an independent of enumeration and specialized structure was assigned to carry out the verification of a large sample of uninhabited, temporary use or dwellings reported with a no contact result. Its objective was to verify and, if necessary, correct the uninhabited condition to improve the coverage of the final census figures and ensuring a higher quality of the information corresponding to the uninhabited or temporary use dwellings, also it enabled to create a perception of work surveillance for the enumeration structure responsible for identifying the inhabited condition of the dwellings.

The verification operation was carried out for 20 days and began two weeks after the beginning of the census operation. About 2.3 million of buildings were verified at national level, representing about 40% of uninhabited and temporary dwellings; through this operation, 139 695 interviews were recovered, that is 6.2% of the buildings verified. It's important to note that, to facilitate both monitoring and verification tasks, as well as avoiding over coverage and invasion of areas, enumerators placed stickers outside the dwellings during the data collection in accordance with the outcome of the visit.

The enumerated dwellings by the verification operation accounted for 0.49% of the total number of private dwellings in the country, thus, for the 2010 census, considering the verification operation and the efforts to guarantee coverage and reduce the non-response, the percentage inhabited dwellings that could not be enumerated was 1.6% of total planned dwellings, one percent less than recorded for the 2005 Population and Housing Count.

2.3. Data processing

After two weeks of field operations, data processing begins, which includes the processes of: capture, coding, validation, release of figures and dissemination. The first step is the reception of materials in the capture area, which includes processes to control the matching between materials received and those reported by the operating area, and whose monitoring is maintained throughout all the data processing stage. During capture and coding, acceptance/rejection sampling methods are used to improve quality and ensure that errors are within acceptable parameters. For coding processes, it's sought to maximize automatically coded descriptions, and for those coded in an assisted form, mechanisms are established for quality monitoring both by sampling process and according to conceptual criteria. Once the information has been coded, the automatic validation process is initiated, which ensures the integrity and congruence of the data based on criteria of a conceptual and logic-based nature; the results of this processing are analyzed with disaggregation to the municipality level, observing that changes and non-response rates are below pre-established tolerances. For the majority of the variables the changes represent less than 1% of the cases, in particular, for the variables of age and sex, the modifications to the information originally collected in the field due to omissions or inconsistencies were applied in 0.05% and 0.08% of cases, respectively.
Before the processing of results for publication is launched, the historical behavior of the main indicators of each municipality is analyzed, the atypical cases are documented and justified. This is done through a system called "Figures Release System", which can be consulted at all levels of the structure to facilitate decision making. In particular, for the data processing, the use of technologies has allowed the significant reduction of processing times, the automation of processes and the publication of results through the Internet and through geostatistical systems.

2.4. Post-enumeration survey

Traditionally coverage was estimated in Mexico though the ratio between the enumerated population and the corrected population, obtained from an analysis that takes into account each demographic component. This estimation of undercoverage is conducted by the National Population Council (Consejo Nacional de Población, CONAPO). For the first time in Mexico, for the 2010 Census, INEGI decided to conduct a post-enumeration survey to obtain a measure of population coverage at the national level, for urban and rural areas, for five-year age groups and by sex, and to identify conceptual and procedural improvements for future censuses. The staff that conducted the survey was independent of the enumeration structure and was highly trained both for the data gathering and for the comparison of the census information with the survey information, carried out on a case-by-case basis. The post-enumeration survey was carried out from July 26 to August 6, that is, approximately one month after the end the field enumeration. The content of the post-enumeration questionnaire included the resident population in the dwelling and their residence status during the census taking (confirmation, entry or exit), sex, age and relationship to the head of the dwelling. The comparison between the census questionnaires and the post-enumeration questionnaires was done manually, it was decided not to use an automated system and returns to field to corroborate the cases with significant differences were implemented, mainly due to erroneous location.
The survey had a planned sample size of about 1.2 million private dwellings nationwide, the sampling scheme was stratified and by clusters in a single stage, the interviews were also conducted face-to-face using paper questionnaires. It should be noted that some of the areas originally selected for post-

enumeration were not covered or were excluded from the comparison due to location errors or insufficient information to compare the results, mainly due to lack of streets names and numbering. In the areas finally considered for the survey, 1.3 million dwellings were counted, for which an interview was obtained in almost 1 million dwellings, about 248 thousand dwellings were found uninhabited or of temporary use and no interview was conducted in 58 577 dwellings .

2.5. Data quality

From the post-enumeration survey, the coverage rate estimates were obtained for the different domains of study using a combined ratio estimator, thus, at the national level, the coverage rate was of 98.7%, for urban areas of 98.59% and for rural areas of 99.07%. The coverage rate obtained was similar between men and women (98.64% and 98.78% respectively), and for the age groups, the best coverage rates were observed for the five years age groups 5-9 (99.1%) and 10-14 (99.08%); on the other hand, the lowest coverage rates were observed in the groups aged 20-39 years (98.4% on average). Another result from this survey was the percentage of migrant population, population residing in the dwelling in May 2010 and that during the post-enumeration survey no longer resided in that dwelling, estimated at 0.82% of which 0.77% were migrants and 0.05% were deceased.

In reference to quality of the data by age and sex, the Whipple index for the 2010 Census was 111.46, which indicates an approximate high quality of the data (110-125). On the other hand, the summarized index of Myers presented a value of 6.67, indicating an intermediate digit attraction level (5.1-15) while the United Nations index was 16.40, meaning that the quality of the information was considered "Satisfactory" (less than 20). Although the quality of data doesn't reach the ideal values, as shown in the following table, efforts have been made to improve the quality of information both at the level of coverage and concerning structure by age and sex:

Indicator		**1980**	**1990**	**2000**	**2010**
Under coverage	Consejo Nacional de Población (National Population Council) (CONAPO)	2.7%	6.1%	3.3%	1.4%
	Post-enumeration Survey (INEGI)	N/A	N/A	N/A	1.3%
Whipple index[1]		126.27	120.37	114.15	111.46
Myers index[1]		13.44	10.79	7.99	6.67
United Nations index[1]		19.69	16.43	18.13	16.40

N/A: Not available

Table 1: Indicators of under coverage and quality in Population and Housing Censuses. Mexico 1980-2010

3. Intercensal Survey 2015

As mentioned before, in 2015 a complete population count was not performed unlike the two previous events between censuses, instead, a probabilistic sample survey was carried out that allows estimates to be obtained up to municipality level. Although the methodology followed is equivalent to that of the population and housing censuses, in which face-to-face interviews are conducted using printed

[1] Internal estimates, Instituto Nacional de Estadística y Geografía (2016).

questionnaires, the control and monitoring computer systems were strengthened and improved, including the use of geostatistical systems for the planning and data gathering. One innovation that was implemented for this event is the use of unique QR codes (Quick Response Codes) in each printed form, which allowed to control the questionnaires from its distribution to the operational structure, the control and monitoring during the data collection and throughout the data processing. An important use for the control of the instruments was the association through the QR codes of each questionnaire to the list of buildings, which strengthened the control of the geographic identification associated to each questionnaire, and the control of the non-response by enumeration area.

4. 2020 round perspective

For the 2020 Census, it is estimated that in Mexico there will be a population close to 130 million persons; although the global tendency is towards the use of mobile devices, currently in Mexico the options that could allow this change in data collection are analyzed, considering the inherent conditions of the country (population size, communication routes, connectivity, insecurity, etc.), seeking to maximize the benefit that would be obtained, which is the publication of the results in a shorter time, without affecting or even reducing costs and maintaining the data quality, and considering the information needs that will influence the size of the census questionnaire.

Assuming a duration for the data collection of one month, with a short and a long questionnaire of approximate length of those used for the 2010 census, about 180 000 mobile devices would be required to conduct the census through a Computer Assisted Personal Interviewing (CAPI) scheme. Although one of the main challenges of this scheme is the subsequent use of the equipment and its associated high cost, it's considered that some alternatives of hybrid methods can be explored, such as the data collection with printed forms for problematic areas, or collecting only the long questionnaire with mobile devices, or even extend the census taking period to two months which will reduce to half the number of devices required, but arouses concerns about over and under coverage.

INEGI already has the experience in surveying, Economic Censuses and the Census of Schools with mobile devices and the institutional capacity for the management of information collected by CAPI, however, for the Population and Housing Census, the magnitude of the structures is a factor to consider, as well as the implications in terms of information security, coverage, and data quality.

Regardless of the collection method, the country's conditions make it increasingly hard to obtain answers by the respondents, and therefore maintain or reduce levels of under coverage, represents a significant challenge for the census operation, and for which specific strategies will be designed, both for the operational level and for communication campaigns, for example, test will be conducted for a self-enumeration through the Internet in case of denial of response.

5. Conclusions

The evolution of censuses has been closely related to technological advances, at the same time, the experience acquired in the Institute has allowed to strengthen statistical information generation processes, which now even have quality standards and policies.

The Population and Housing Census is a fundamental statistical project for the country, since it’s used for the planning and evaluation of programs and public policies, the focusing of vulnerable populations among many other aspects, in this sense, INEGI seeks to satisfy the information needs of the users, contributing thus to the national development, which implies to look for the continuous improvement of the processes and the assurance of data quality.

In this context, the change to data collection with mobile devices is contemplated for the 2020 round, but even more so, it will seek the strengthening of actions for the use of administrative records, which, although in the medium term are not seen as a replacement for the population censuses in Mexico, until now have been a source of data thinly used for this statistical project.

References

- INEGI (2009). 125 años de la Dirección General de Estadística 1882-2007, INEGI.
  (125 years of the General Directorate of Statistics 1882-2007, INEGI.)

- INEGI (2010). Síntesis Metodológica y Conceptual del Censo de Población y Vivienda 2010, INEGI.
  (Methodological and Conceptual Synthesis of the 2010 Population and Housing Census, INEGI.)

- INEGI (2012). Resultados de la encuesta de posenumeración del Censo de Población y Vivienda 2010, INEGI.
  (Results of the post-enumeration survey of the 2010 Population and Housing Census, INEGI.)

- INEGI (2012). Resultados de la encuesta de verificación de viviendas deshabitadas y de uso temporal del Censo de Población y Vivienda 2010, INEGI.
  (Results of the survey of verification for uninhabited dwellings and of temporary use of the Population and Housing Census, INEGI.)

- INEGI (2015). Norma Técnica para el aseguramiento de la calidad de la información estadística y geográfica del Instituto Nacional de Estadística y Geografía, INEGI.
  (Technical Standard for the quality assurance of the statistical and geographical information of the National Institute of Statistics and Geography, INEGI.)

- INEGI (2015). Política de calidad institucional, INEGI.
  (Institutional quality policy, INEGI.)

- Cabrera Gustavo (1997). Cambios en el tamaño y crecimiento de la población total, Revista Demos Núm. 10, México.
  (Changes on size and total population growth, Demos Magazine N°10, Mexico.)

- Consejo Nacional de Población (2016). Estimaciones demográficas 1990-2009. CONAPO.
  (Demographic estimations 1990-2009, National Population Council CONAPO.)
  Available at: http://www.conapo.gob.mx/es/CONAPO/Proyecciones_Datos

- Consejo Nacional de Población (2016). Proyecciones de Población 2010-2030. CONAPO.
  (Population projections 2010-2030, National Population Council CONAPO.)
  Available at: http://www.conapo.gob.mx/es/CONAPO/Proyecciones_Datos