

## **Has census coverage in Latin America improved with time?**

*Gabriel Mendes Borges*

Researcher at the Brazilian Institute of Geography and Statistics (IBGE) and Ph.D candidate in Demography at University of California, Berkeley. Co-founder, Latin American Observatory of Population Censuses (OLAC).

*Nicolás Sacco*

PhD Candidate in Social Demography at the University of Buenos Aires. MA in Applied Statistics (University of Tres de Febrero) and MA in Social Demography (University of Luján). Co-founder, Latin American Observatory of Population Censuses (OLAC).

Date of last revision: April/2016

**DO NOT CITE WITHOUT AUTHORS' PERMISSION**

## **ABSTRACT**

Taking into account the 2020 census round, we evaluate historical and recent trends in censuses coverage. We describe how censuses coverage has changed in Latin America since the mid-twentieth century to the 2010 round. Carrying out our analysis in a comparative and historical perspective, using figures based on direct and indirect demographic estimations, we also explore the outcomes on census coverage related with its cost. We find out that census coverage in Latin America showed a downward trend in omission levels. However, the last census round (2010) made clear growing difficulties faced by national statistical offices to mount the census operation successfully and a great heterogeneity among countries.

Key words: censuses coverage; Latin America; 2020 round; census cost.

## INTRODUCTION

Population and Housing Censuses are the most important source of demographic information in a country, playing a particularly important role in Latin America, where administrative data are often incomplete, of poor quality and/or inadequately detailed. Furthermore, the reduced number of household surveys comparing with developed countries posits censuses as the only reliable source of information for many matters. Household surveys also have the disadvantage of not releasing information at the geographic desegregation level available in the censuses.

Censuses are used in a variety of ways for federal and local governments, private companies and researchers. A usual Constitutional target of a census is providing states and local areas information to support apportionment and congressional redistricting. This specific use makes the precision of census figures particularly necessary. Furthermore, due to socioeconomic disparities in Latin America, distinct net undercoverage is likely to impact geographic areas and specific groups differentially. Visibly, the general goal of measuring the quality of a census is to assess the scope of census coverage error by domain and by socioeconomic and demographic group.

The existence of coverage and quality problems is intrinsic to activities of the magnitude of a census. Thus, evaluating and measuring such problems are of fundamental importance. Coverage assessment is a set of techniques that measure the disparities among census figures and the corresponding true counts<sup>1</sup> for groups or areas. Census coverage estimation is one aspect of the overall quality assessment of a census. Coverage measurement is the quantitative aspect of coverage evaluation, which additionally includes qualitative techniques. The dissimilarity between the census counts and the related true counts at the individual or household levels is known as census coverage error.

Some previous efforts have been made to evaluate Latin American censuses in a comparative perspective (Arretx and Chackiel, 1985; CELADE, 1968; Mortara et al., 1964; Tacla

Chamy, 2006). However, the last census round (2010) opens up a new series of questions, not addressed by previous researches.

Recognizing the importance of content errors and the differential coverage by distinct groups, this work concentrates on the analyses of the overall coverage of census. The main objective of this paper is thus to describe how censuses coverage has changed in Latin America since the mid-twentieth century, with special focus on the 2010 round.

We evaluate the hypothesis that censuses' coverage has improved over time, carrying out our analysis in a comparative and historical perspective. A series of questions drives the guidelines and hypothesis of this article: what was the overall coverage of the Latin American censuses in the last round and how is it different from the historical coverage levels? Trying to answer this question, we also aim to search for possible explanations of the observed trends: what are the main reasons that have been leading to census undercount in the region?

The paper is divided into four parts: we first describe previous research on census coverage estimates in the region, especially those papers that also address census rounds in an historical and comparative perspective. In the Data Sources and Methods section we describe the data and methods used for further analysis. In the Findings section we provide empirical evidence of changes in census coverage over the period under analysis, with especial focus in the last round. As part of the results section we provide a case study for four countries (Argentina, Brazil, Chile and Paraguay), in addition to discuss the relationship between costs and the technology adopted with undercount estimates, in order to make a first approach to possible explanations of the observed differences in coverage by country. Finally, the Discussion and Conclusions section places the results in a broader theoretical perspective and more general framework of analysis.

## **LITERATURE REVIEW**

Mortara (1964), presenting an overall picture of the deficiencies in the Latin American censuses around 1960, took into account largely conjectural estimates and observed

that the faulty enumerations of 1940 and 1950 rounds (where available) were subject to a wide variations between countries. He argued that “in most Latin American countries, the inefficiency of public administrations, added to the lack of civic-mindedness, the widespread ignorance and poverty of the citizens, make reliable censuses difficult”. As a result, “some corrections would have been needed to bring census results nearer to reality” (Mortara et al., 1964, p. 60). However, he recognizes that great progress had been made in this field, comparing to censuses undertaken in the beginning of the century, and acknowledges the efforts made by the national statistical services, stimulated and directed by the international institutions.

Analyzing the modern Latin American censuses at that time, CELADE<sup>2</sup> (1968) estimated census coverage in the region in the 1960s given the figures provided by the Post-Enumeration Surveys (PES) and Demographic Analysis (DA). Only six countries performed the PES in this census round: Costa Rica, Chile, Guatemala, Honduras, Panamá, Uruguay. The census omission then ranged from 5.4% in Chile to 2.1% in Uruguay, Panamá and Costa Rica. By calculating estimates with DA, the authors found important omission differentials by age and sex, and a relevant misreporting of ages involving young and old age groups.

Arretx and Chackiel (1985) evaluate the 1980 census round in Latin America and conclude that despite previous beliefs that censuses in the region would necessarily improve their quality and coverage over time, following the technologic, economic and social improvements in the region, some countries’ censuses deteriorated in that decade. They assign responsibility for the worsening of census quality in some of the countries they analyze to the economic crisis and political conditions in the region.

More recently, Tacla Chamy (2006) mentions that, with a few exceptions, the censuses taken in Latin America and other regions during the eighties showed a substantial deterioration both in terms of coverage and quality of the data, when compared with censuses undertaken in the previous decade. She considers that, in this round, the methodology applied for collection, processing, evaluation and dissemination of results and the topics included in the censuses had some overlapping aspects, but

were not uniform within countries. Tacla Chamy (2006) also addresses the fact that the preparation of the 2000 round had been affected by the economic situation of the countries of the region and had suffered a strong constraint on fiscal spending, which had caused important repercussions in census operations, as well as what had happened in the 1980s and 1990s.

Tacla Chamy (2006) presents figures of census omission for the census rounds from 1950 to 2000, for which assessments were made. The estimates used were implicit in the population projections and estimates prepared by CELADE, sometimes in collaboration with national statistical offices. She finds out that, in the 30 years between the first censuses carried out in the 1950s and the last during the 1980s, coverage had oscillated significantly. Eight out of the seventeen countries managed to lower the omission rates in 1980, comparing to the 1970 census round.

As noted in a document produced by CELADE (1968), Tacla Chamy (2006) mentions that there were several reasons to explain the deterioration of the 1980s census, including: inappropriate design of the questionnaires, for instance, the inadequate questions and format, the bulky number of questions, the order of the answers; not always correct use of sampling in the survey; problems regarding to coding or pre-coding of responses; automatic assignment of inconsistent or nonexistent values; use of computer equipment in discordance with individual characteristics of each country etc. Other problems were found, such as stem from unrealistic organization, not adapted to the conditions of the countries of the region; lack of appropriate statistical mapping; insufficient and inadequate instruction method to a heterogeneous number of people training.

For the 1990s round, Tacla Chamy (2006) claims that only five out of the 16 countries that conducted census in this decade managed to maintain or lower the census omission<sup>3</sup>. The preparation of the new 2000 round had also been affected by the economic situation of the Latin American countries, as a result of difficulties resulting from a strong restriction on fiscal spending. This caused significant impact on census,

as high-cost operations require timely funding from governments for meet each of the stages and ensure the quality of the results.

This same budget constraint led to three countries to adopt extreme measures. Uruguay, for example, chose to implement a census in 2004 investigating only the population by sex and age. Colombia applied sampling in the census, i.e., used a basic questionnaire to the entire population and another extended questionnaire to a sample. At the same time, Peru made a basic counting first and then applied a mega survey, both conducted in 2005. Estimates for the 2000 round of census indicate that the census omission in the countries available in Tacla Chamy's (2006) paper fluctuates between 2.5% and 5.8%, except for Venezuela, with an undercount of 7.5%.

What has happened with census coverage in the 2010 round? It is well known that the 2000s was a period of high economic development in almost all Latin American countries, with increasing levels of social mobility and well-being for different segments of the population. These changes in the regional socio-economical context opens up a new series of questions to assess whether the trends in census coverage observed in previous years remained. Did the coverage improve in parallel with the general economic context of the region? Are these two factors related in a regional level or do we have to look at the internal heterogeneities and the context of each country?

Thinking of the next 2020 round, in addition to these working-questions, the evaluation of historical and recent trends in censuses quality can also shed light on the future perspectives for censuses in a region, since many countries are now facing important economic and social crisis.

#### **DATA SOURCES AND METHODS**

Errors in the census results are classified into two general categories – coverage errors and content errors. Coverage errors are, in general terms, related to failure in counting persons or housing units, leading to missing or duplicated cases. Content errors refer to mistakes in the given information of persons or housing units effectively

enumerated. All the information collected in the censuses are subject to error. Two common examples of information that affect age of people are, for instance, the preference for age with terminal digits and the age exaggeration among elderly population. Content errors have been improving over time, as a result of better questionnaire designed, improvements in educational level of population and the better quality of administrative records. An additional element significantly affecting the sex-age structure of the population is the differential of coverage within these two variables. Some coverage pattern seems to be recurrent in different contexts, such as the omission of children and the differential undercount by sex. Notwithstanding the fact that census errors can happen in multiple ways, this paper concentrates solely on the analyses of the overall census coverage, understanding that this measure gives a comprehensive picture of census quality.

There are many census evaluation methods, but two have stood out as prevailing. The first is the Post-Enumeration Surveys (PES), in which a sample of households is revisited after the census, some data collected again and then compared to that collected by the census in the same areas. The second method of evaluation censuses is performed through Demographic Analysis (DA), which consists of evaluating data using internal consistency within the same census and/or the application of demographic techniques using administrative records for developing population estimates, used later to compare to the census results. These two methods represent the main examples of the so-called direct and indirect techniques for measuring census coverage.

In principle, these two methods are not competitive, and can even be considered complementary, since the PES could provide additional information to incorporate to the populational dynamic analysis. Demographic analysis and the need for consistency between the populational estimates and the dynamics of its components could ratify PES results or demonstrate the need for adjusting its figures, or review the reconciliation exercise attempt to explain the differences. Moreover, the Post-Enumeration Surveys can provide geographically disaggregated information. The latter

is more complex to obtain by indirect evaluation, in particular because of the distortions which may occur with open populations regarding mobility, or the occurrence of local events without an almost unimportant national expression (Chackiel, 2009).

In this paper, we use both indirect and direct techniques to measure census coverage, in order to offer the widest possible range of estimates. Using multiple sources and methods is also useful to give an overall perspective on the quality of the estimates.

The performance and use of the PES have not been uniform in Latin America, raising many criticisms about the effectiveness of this method in evaluating censuses in the region. Chackiel (2009) argues that, despite the fact that many countries have included the PES in their census planning, not all cases have used the same methodology and technical rigor. Another issue is related to the lack transparency and technical information regarding the publication of the PES reports. Notwithstanding these issues, we report the results of direct evaluations available to our knowledge. We also report indirect evaluation performed by the National Statistical Offices (NSO), normally using the “Demographic Reconciliation” method, often with the support of CELADE.

Furthermore, in order to maintain methodological consistency over time and between countries, in addition to obtain estimated statistics for the region as a whole, we calculate our own indirect estimates, by comparing censuses figures to population projections published by the United Nations (2015). Our measure of census undercount is given by the following formula:

$$U_{ij} = [1 - (C_{ij}/P_{ij})] \times 100$$

where  $U_{ij}$  is the overall estimated undercount,  $C_{ij}$  is the census count and  $P_{ij}$  is the population projection for country  $i$  and year  $j$ .

In addition to estimate census coverage for each country, we also calculate three measures of centrality (weighted average, simple average, median) and one measure of dispersion (interquartile range) in order to provide a broader characterization of the overall census quality in the region and the diversity between countries.

## **FINDINGS: CENSUS COVERAGE THROUGHOUT HISTORY**

Table 1 shows undercount estimates published by the NSO, using both direct and indirect techniques. We observe that census assessments both through direct and indirect methods were not made or are not publicly available for many years and different countries.

Furthermore, in some cases, indirect and direct evaluation for the same census are not consistent. In Brazil, for example, direct estimations have given consistently higher undercount rates than demographic analysis and even the general time trend is not consistent. In the 2000 Argentinian Census, the PES estimated 7.0% of census omission, whereas indirect evaluation resulted in 2.8%. The main problem with these inconsistencies is that statistical offices have been unable to justify the differences. Thus, the general procedure, with just a few exceptions, has been evaluate and adjust base population for projections using demographic analysis, undermining the use of the Post Enumeration Surveys.

Despite these limitations, Table 1 depicts some interesting results. For the 2010 round, figures derived from indirect methods indicate that census omission fluctuates between 0.9% (Cuba) and 9.3% (Chile), not considering the outlier case of Paraguay. Ecuador, Venezuela and Argentina seem to have improved coverage in 2010, compared to the 2000 census, down from almost 6 to 4.3% in the first case, from almost 8 to 6.5% in Venezuela and from about 2.8% to 2.0% in Argentina. Costa Rica and Uruguay showed an important decline in coverage between the last two rounds of census, with undercount increasing from 2.9% to 6.2% and from 1.9% to 3.1%, respectively. In the case of Chile, the National Statistical Office reported an increase in census omission from 3.8% to 9.3% between the 2002 and the 2012 censuses. Finally, Paraguay presents a very special case, with omission raising from 6.9% to 25.6%, a substantial deterioration in the last decade.

Table 1. Census undercount (%) by country and census round, according to indirect (Demographic Analysis) and direct (Post-Enumeration Surveys) evaluation released by the National Statistical Offices - 1950-2010

Country	Census round												
	Indirect evaluation (DA)							Direct evaluation (PES)					
	1950	1960	1970	1980	1990	2000	2010	1960	1970	1980	1990	2000	2010
Arg						2.8	2.0						7.0
Bol													2.9
Bra				1.8	3.6	2.3	2.6		7.3	4.3	4.7		5.8
Chi						3.8	9.3	5.4					4.1
Col			9.6	7.3	12.0	3.7							
Cos						2.9	6.2	2.1					
Cub							0.9						0.3
Dom	9.9	7.8	9.8	7.7	5.3	4.6							
Ecu	7.9	7.2	3.3	4.3	5.7	5.9	4.3				6.7	5.5	
Els													
Gua								3.7					
Hai													
Hon						7.0		5.3		5.3	18.0		
Mex							1.6						1.3
Nic													
Pan						6.3	6.7	2.1					
Par						6.9	25.6		7.4	10.0	7.4	6.9	
Per		4.9	4.1	4.3	2.6	3.9							
Uru					2.1	1.9	3.1	2.1		2.1			4.9
Ven						7.7	6.5					6.8	

Source: National Statistical Offices

Table 2 shows the estimated undercount of the censuses undertaken in Latin America since the 1950 round using a comparable methodology, described in the previous section. It also presents summary statistics for all countries combined.

Table 2. Census undercount (%) by country and census round, according to indirect evaluation comparing censuses figures and UN population projections - 1950-2010

Country	Census round						
	1950	1960	1970	1980	1990	2000	2010
Arg		4.1	3.6	1.8	2.2	4.2	3.5
Bol	13.4		11.3		10.7	3.8	3.1
Bra	5.1	5.2	4.5	4.1	5.0	4.2	4.5
Chi	8.2	6.1	7.6	2.5	2.1	3.1	9.8
Col	12.4	7.4	5.2	11.7	9.8	5.2	
Cos	17.3	10.7	7.0	10.2		3.8	6.9

<b>Cub</b>	7.6		2.8	1.9		0.2	1.7
<b>Dom</b>	11.3	9.3	11.0	8.6	5.4	4.2	5.7
<b>Ecu</b>	9.8	9.3	5.2	6.1	7.5	6.6	4.4
<b>Els</b>	16.2	12.5	6.7		6.0	4.1	
<b>Gua</b>	12.0	7.5	13.1	17.6	18.1	10.4	
<b>Hai</b>	4.8		10.7	16.4		6.7	
<b>Hon</b>	9.0	9.7	12.3	9.3		5.7	
<b>Mex</b>	9.0	9.7	7.2	4.5	5.4	5.3	5.9
<b>Nic</b>	19.3	21.7	24.7		6.1	4.9	
<b>Pan</b>	14.0	7.6	6.9	10.0	6.5	6.9	6.5
<b>Par</b>	11.7	11.1	10.3	11.3	7.8	7.3	26.1
<b>Per</b>		5.6	4.8	5.6	5.8	4.0	
<b>Uru</b>		2.4	1.6	2.4	2.8	2.5	3.2
<b>Ven</b>	11.8	11.6	12.3	9.9	10.5	8.9	8.5
<b>Weighted average</b>	<b>8.3</b>	<b>7.0</b>	<b>6.2</b>	<b>5.7</b>	<b>6.0</b>	<b>4.9</b>	<b>5.2</b>
<b>Sinlge average</b>	<b>11.4</b>	<b>8.9</b>	<b>8.4</b>	<b>7.9</b>	<b>7.0</b>	<b>5.1</b>	<b>6.9</b>
<b>Median</b>	<b>11.7</b>	<b>9.3</b>	<b>7.1</b>	<b>8.6</b>	<b>6.0</b>	<b>4.6</b>	<b>5.7</b>
<b>Interquartile range</b>	<b>5.1</b>	<b>5.1</b>	<b>6.3</b>	<b>7.5</b>	<b>4.2</b>	<b>2.9</b>	<b>4.3</b>

Source: (United Nations, 2015) and National Population and Housing Censuses.

Overall, there have been improvements since the beginning of our series. In 1950, 8.3% of the population living in countries that undertook a census was not counted. This proportion fell to 4.9% and 5.2% for the 2000 and 2010 rounds, respectively. Countries such as Argentina, Brazil, Cuba, Peru and Uruguay have presented historical high quality censuses in terms of coverage. Venezuela, Paraguay, Guatemala, Panama and Haiti are countries that tend to have censuses high omission rates. Bolivia is an important example of a country that has improved the quality of its censuses in terms of coverage since the 2000 round.

The 1980 round of census was particularly diverse in terms of coverage, with a few countries presenting relatively good censuses, with omissions of less than 5% (Argentina, Brazil, Chile, Cuba, Mexico and Uruguay), whereas many countries presented really poor censuses, worsening the coverage comparing to the previous census round (Colombia, Costa Rica, Guatemala, Haiti, Panama, Paraguay).

The censuses around 2000 were, in general, of high quality, with an average omission around 5%. Even in countries with high undercount rates (Guatemala, Paraguay,

Venezuela), this census round was one of the best of their series. Despite been affected by the regional economic situation and suffered a strong constraint on fiscal spending, this was, in general, a good census round for Latin America.

The censuses undertaken around 2010 had, in many countries, lower coverage than the previous round. This was the case for Brazil, Costa Rica, Cuba, Dominican Republic, Mexico and Uruguay. Chile and Paraguay are extreme examples, in which censuses have worsened significantly in the last census round, from 3.1% to 9.8% and from 7.3% to 26.1%, respectively. The 2010 Census in Paraguay is the worst of our series in terms of coverage. These two cases will deserve special attention in the next sections.

Some of the inconveniences reported in the 2010 census round in Latin America had to do with the financial situation in the region, exacerbated by the international crisis. The insufficient human and financial resources, in addition to the lack of autonomy of the Institutes have also been reported as factor contributing negatively to this census round (DANE, 2014).

These result show that institutional, economic and/or political crisis, recurring events in Latin America, appear as important factors that affect the quality of censuses in the region. Two examples are the 1991 Census of Brazil and the Argentinean 2001 Census, which presented high omission levels compared to the historical series under evaluation.

What are the reasons to the observed census coverage fluctuations over time? We must observe each particular country to respond these questions. For this purpose, in this paper we take the examples of the overall trends in Argentina, Brazil and also evaluate the most recent censuses in Chile and Paraguay.

### **Argentina**

In Argentina, the analysis of census results has traditionally been done at the provincial level. INDEC (2005) made an exercise to estimate the 2001 census coverage, using demographic methods. The indirect methods result showed an omission of 2.75%, while the post-census survey gave figures of about 10% of omission in the City of

Buenos Aires, Buenos Aires and Greater Córdoba, 7% in Great Rosario and for the total, but this information was not published (INDEC, 2010, p. 4).

Complementary to that analysis, which took into account the preliminary results of the 2001 Census, Massa et al. (2005) compared the pre-census dwellings counting and then the dwellings census enumeration to estimated coverage. This procedure had the possibility to make estimates based on pre-census information, i.e., the count and the previous list of dwellings on which it could be estimated, in turn, the total number of interviews realized and unrealized. On this basis, they were able to trace some of the problems arising from the census operation and the differences between dwellings actually surveyed and those expected based on the count. In addition to this analysis using the information from dwellings with absent inhabitants, they managed to estimate the undercount dwellings census and its possible influence on the associated population undercount. Thus, the estimate of total population by province was made under the strong assumption of closed population. With this exercise it was possible to rebuild some of the problems arising from the organization and implementation of the 2001 round.

However, in 2010 this preliminary list of dwellings has not been published to date. Furthermore, the reasons for discarding a post-census survey as an element for measuring coverage in 2010 are unknown. Probably it was due to the financial cost of the method with consideration of different experiences in Latin America in which little use of their results (CELADE, 1984; Chackiel, 2009, pp. 57–61) was observed, an issue that strengthens the uncertainty the cost-effectiveness of the operation (Whitford and Banda, 2002). For 2010 this complementary direct analysis with pre-census information is not possible.

Sacco (2015), using indirect estimates, calculated a total omission of 2.1% for 2010. For the male population the omission was 2.9% and for women 1.3%. Gonzalez et al. (2013) presented a study of the level of coverage of the 2010 census based on population projections. Since the proposed method derives from demographic components, it is a method similar to the technique using indirect estimation. They

discovered that the total level of omission was 2.4%, higher for men (3.3%) than for women (1.6%). Since other methods were not available, these two studies were the only alternative proposals, for the moment, to the INDEC official estimates at the level of the whole country. In that document, using the indirect method of demographic components INDEC reported that the level of omission was 1.9% for both sexes, and is also bigger in men (2.4%) than woman (1.5%).

This results, although higher than in previous censuses of 1980 (1.0%) and 1991 (0.9%), for both sexes, respectively, are in line with historical levels of coverage in Argentina and can be considered as satisfactory taking into account the overall level of omission of Latin America for 2000 and 2010 rounds, as showed before. However, this results display the importance that reached census estimated coverage for Argentina in 2010: the outcome of this exercise is close to a significant magnitude of 900,000 people of the population actually enumerated. Although shows that there was a low omission with respect to the regional context (Chackiel, 2009; Ruiz, 2013), however continues to be very significant both in relative and absolute terms. INDEC (2010, p. 3) found high levels of omission, 2001, therefore, while a lower level seen over the previous census is still maintained with minor substantive improvements in 2010 and still above it those observed in the 1980 and 1991 census.

### **Brazil**

The first census in Brazil was in 1872. The 1880, 1910 and 1930 censuses were not accomplished due to political issues. The 1890, 1900 and 1920 were very problematic censuses in terms of quality and coverage. The 1940 Census is a landmark in the population statistics of Brazil, the first under the responsibility of IBGE (Brazilian Institute of Geography and Statistics). This was a particularly important census, thanks to the effort of the Italian demographer Giorgio Mortara, who participated in all stages of this census. Mortara had identified limitations of previous censuses, both in terms of data quality and content of the questionnaire, creating new conditions to the demographic studies in the country, which had also influenced the field in the Latin America as whole (Oliveira and Simões, 2005).

In fact, the 1940 and 1950 censuses in Brazil seem to have quality. Table 2 shows that undercount in the 1950 Census was only 5.1%, far below from its counterparts. The 1960 Census had a similar degree of omission (5.2%), but due to political instability and internal crisis in the beginning of the 1960s, their results were not published until the late 1970s. The 1970 Census has been also considered a landmark in the history of the Brazilian population censuses, both regarding to its organization and the reliability of its results. In fact, our evaluation shows that the undercount rate dropped to 4.5%.

In a context in which many countries saw worsening in census quality, the 1980 Census was, in a certain way, successful, showing omission of only 4.1%. An increase in the omission rate appears in the 1991 Census (5.0%) and reflects the context in which it was undertaken. Fiscal and political crisis postponed the census, initially planned for 1990, but it was still run under unstable circumstances.

Persistent economic difficulties have pushed IBGE towards the use of computing and technology in collecting and disseminating information. The 2000 Census was the first one to use scanner in data processing. Following the general regional trend of improvements in census quality, the 2000 Census was, along with the 1980 Census, the best one in terms of coverage.

The 2010 Census was very innovative in terms of technological advances, being the first one to use PDA/Handheld-based in data collection. The census coverage was similar to the 2000 Census, with a slightly increase (4.5%).

Overall, the Brazilian censuses have been presenting a regular and satisfactory undercoverage error, ranging from 4.1% to 5.2%. Even though some contextual factors have been associated with census coverage, small differences might be also attributed to limitations of our estimation method.

It is worth noting that fiscal and economic crisis remain highly influential in the decision regarding census taking in Brazil. The 2015 Intercensal Count, which had been planned to have the format of a census, was first postponed to 2016 and then definitely cancelled due to budget restraint.

### **Chile and Paraguay: the recent unsuccessful experience**

As discussed before, the 2010 round of census in Chile and Paraguay was the worst in these countries. The causes of the problems are distinct in each country and the uses of the final census figures were also different.

The 2012 Census in Paraguay had two main operational innovations: i) the use of PDA to collect and transmit the data; ii) change from a "de facto" to "de jure" format. An effort was made to limit the per capita cost of the census to around US\$2.70, below the regional average of US\$3.08. The Inter-American Development Bank, which provided financial support to the 2012 Census in Paraguay recognized that despite improvements in terms of technology, organization, management, and the availability of human resources, previous censuses in Paraguay had suffered from problems of quality, coverage, and timeliness of both the delivery and use of the information generated (IDB, 2011). This analysis is supported by the results showed on previous sections showing that Paraguay use to have higher census undercount than the regional average. The (IDB, 2011) loan proposal document also claims that the problems relating to statistics in Paraguay are structural and the production of statistics does not operate as a national system. Seeking for factor contributing to this problems, the document lists some issues related to the weaknesses of the NSO in Paraguay (DGEEC) in areas directly related to its work, which is lack of planning, organization and coordination of census activities. Two specific issues are lack of human and financial resources, and it has been showed that DGEEC has a much poorer structure in terms of staff and budgeted than other Latin American countries. The NSO of Bolivia (INE), country with relatively similar socioeconomic conditions to Paraguay, has 7 times more employees and almost three times of the statistics bureaus' per capita budget than DGEEC.

There have been reported many specific problems regarding to the 2012 Census in Paraguay, such as conflicts regarding enumerators' payment and PDAs being stolen. Nevertheless, we think that the above-mentioned structural problems regarding the

national statistical system in the country are essentially the main reason to the failures in this census.

In regard to the use of the results, even with more than  $\frac{1}{4}$  of the population not being counted, the DGEEC decided to use the 2012 census to estimate fertility, mortality, migration and the proportion of population by sex in the districts (DGEEC, 2015). This can be very problematic, considering that this omission is likely to be biased according to different demographic and socioeconomic groups. The base-population used as starting point of the projection is the adjusted 2002 Census. Even in this better census, the adjustment was differential by the geographic strata. The omission in Asunción was 3.3% and in other big cities 9.7%. The use of the 2012 Census for population projections purposes is now assuming that undercoverage was roughly constant over districts in 2012, which is an unrealistic hypothesis. However, in a country with an incomplete vital statistical system and lack of household surveys, the census is the only source to estimate demographic parameters, such as fertility, mortality and migration, and the geographic distribution of the population. In this kind of situation, the census has to be used regardless its quality.

In the case of 2012 Census in Chile, an external committee was called to revise the census process, evaluate the results and propose recommendations (Bravo et al., 2013). The authors of the report consider many possible reasons for the problems in the census, which can be summarized as lack of planning, organization and management of the census activity. More specifically, the decision to change the format of the census from “de facto” to “de jure” less than one year apart to the census date and all the consequences arising from this decision seem to have been decisive to the failure of the census (Bravo et al., 2013). This decision does not seem to have been a pure technical and statistical decision, but appear to have also contained some specific political strategies (Yuri and Carlos, 2014).

The committee recommended not to use the 2012 Census data for official estimates and public policies purposes (Bravo et al., 2013) and the final decision made by the INE was to discard the census. It is interesting to compare the use of the data in the case of

Chile with the final decision to use the 2012 Census in Paraguay. In the case of Chile, the census was discarded with a much lower omission rate than in Paraguay (around 9%). However, Chile is a country with better vital registration and could manage to produce population projections and estimates without the most recent census.

Another suggestion made by the external committee was the realization of a short census in 2015. In fact, a new census has been organized to happen in April 19, 2017, returning to a “de facto” model.

### **CENSUS COST AND TECHNOLOGICAL ADVANCES**

In this section we explore the relationship between cost and use of technology with census coverage, assessing the hypothesis that these factors would be able to explain the observed differences in coverage described before. Technological advances have always been one of the main hopes to increase timeliness, provide a better use of the data and improve the overall quality of censuses. While the benefits of technological advances in census taking might have been true in many ways, they do not seem to affect necessarily the quality of census, at least in terms of coverage.

Perfit et al. (2012) create a typology to divide countries according with the technology used in each census: first, second and third and fourth generations. Five countries used the third generation collection method in their last census round, which means the use of mobile devices to collect data: Brazil, Colombia, Paraguay, Uruguay and Venezuela.

As we have seen before, Brazil and Uruguay, countries with traditionally good censuses, presented a slightly increase in census undercount after the implementation of the PDA. In the case of the 2011 Census in Paraguay, there was also a shift from “de facto” to “de jure” model, such as in Chile and Paraguay. Colombia presented a significantly improvement in census coverage in the last round of census (2005) and Venezuela had a slightly improvement comparing the 2000 and 2010 census round. Paraguay, as discussed earlier, had an unsuccessful experience in the first census which high technology methods were used.

Perfit et al. (2012) claim that technological changes adoption requires profound changes in the census methodology. A paradigmatic example is the necessity of undertaking a “de jure” census once the PDA is used, given to the infeasible large quantity of devices needed to run a “de facto” census. This is an odd example of important technical and conceptual decisions been driven by the desirability to implement technological advances, when we would expect this relationship working in the opposite way.

An issue closely related to the implementation of technology is the cost of a census - the use of technology is supposed to, among other benefits, reduce cost (Perfit et al., 2012). We have seen that many censuses have been affected by financial and economic crisis and the lack of resources. Thus, we evaluate the relationship between the cost per capita (in US\$) and the overall census coverage in the 2000 and 2010 census rounds, being aware of the limitations of comparing costs from different countries.

Figure 1 shows no clear relationship between these two variables, meaning that: i) it is possible to undertake good censuses with low costs; ii) expensive census operations do not result necessarily in high quality census.

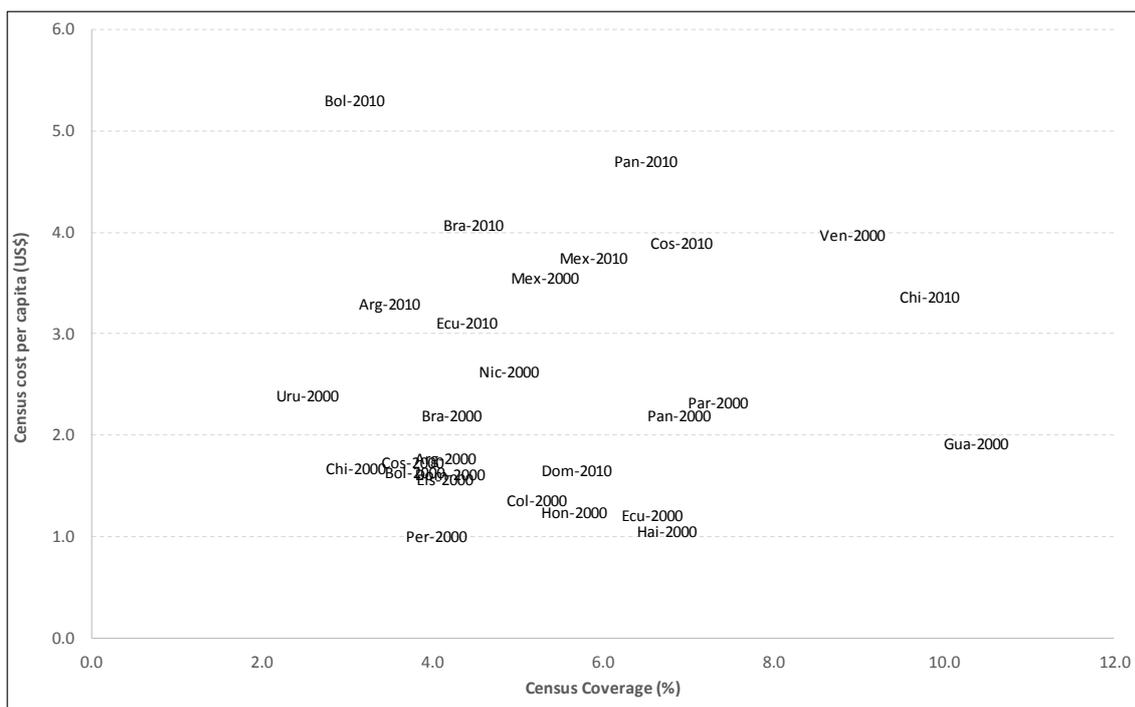


Figure 1. Census cost per capita (US\$) and Census coverage. Latin American countries, 2000 and 2010 round.

## DISCUSSION

Almost all the 2010 Population and Housing Census round planned to take place in Latin America and the Caribbean were undertaken. This latest round was characterized by the introduction of novel questions on census forms, such as those which identify and characterize international migrants, and conceptual changes in existing questions as to measure the prevalence of people with disabilities and variations on data collection concerning indigenous and black populations. Technological innovations, use of electronic devices capture by enumerators, incorporating new tools for cartographic updating and use of social networks for public broadcasting operating were also examples of changes from previous censuses. Furthermore, the number of countries that joined "de facto" censuses has increased, which was confirmed as the predominant type in the region.

However, the 2010 census round also made evident the growing difficulties faced by national statistical offices to organize the census operation successfully. Furthermore, an increasing heterogeneity among countries on infrastructure and capacity, as well as the conceptual and operational definitions used still persists. Other themes are in

constant debate, such as the duration of the census, the use of a single questionnaire compared to the option of short and extended questionnaires, allocation of persons in dwellings occupied by absent residents, the release of census microdata for free use, dissemination mechanisms result in general, among others. Census evaluation and the use of census data after evaluation have been also vastly discussed topics.

The results deriving from census evaluation can be used in several ways, for instance: i) guiding improvements in future censuses and surveys; ii) assisting users in their interpretation of the results; iii) adjusting the census results (Baldrige et al., 1985). The first two uses are extremely important, although it is not clear how the countries have been dealing with them. The third one is a controversial topic and, to our knowledge non Latin American country have ever adjusted the final census results based on the evaluation. In fact, although extensively discussed, adjusting census figures is not a common practice even in other regions. The most symptomatic case of census adjustment is, perhaps, the UK, where the last two censuses (2001 and 2011) fully integrated coverage measurement processes, resulting in the development of the One Number Census (ONC) methodology (Abbott, 2009). A more common practice in Latin America is to use evaluation to adjust censuses in order to derive the base population for projections. This procedure has been used, for instance, in Argentina, Brazil, Colombia, Costa Rica, Chile, Ecuador, Mexico, Paraguay, Peru.

Censuses remain the most important source of statistical information in the region. However, the quality of the information provided has not always been as high as desired. In many cases, conjunctural factors such as political, economic and institutional crisis have affected censuses directly. Furthermore, structural aspects regarding the organization of the National Statistical Offices have also played an important role in census quality. This is not an issue that can be solved in the short term and in the “night before” the census date. In this sense, we have shown that the cost of a census is not necessarily guarantee of success. We argue that assessing these issues require long-term and consistent investments to develop solid national statistical systems, in which statistical offices play an important role.

The results deriving from the of Post-Enumeration Surveys and Demographic Analysis should be, in principle, complementary. However, we find out that, in many cases, coverage estimations from the two different data sources are inconsistent. In order to achieve the main purposes of the evaluation, the National Statistical Offices and academic researchers should be able to understand and explain these differences, but there has not been satisfactory answer so far.

Despite the use of the PES in many countries, large part of the census evaluation in Latin America has been done using the DA method. This method has been used by the NSOs, CELADE (Chackiel, 2009) and the United Nations (Gerland, 2014). This was also the preferred approach used in this paper. However, DA methods have several limitations, the most important being the possible inaccuracy of demographic parameters estimations, resulting from the lack of good vital statistics in the region. New alternatives have been proposed to deal with the limitations of the methods currently available. Wheldon *et al.* (2013), for instance, propose a method that reconstructs historical demographic parameters using a Bayesian hierarchical approach, estimating age-specific population count. In addition to simultaneously estimate demographic parameters, considering the uncertainty associated with historical demographic data, it also incorporates measurement errors, offering census coverage estimates as one of the outcomes. This is a promising method, which can be implemented in Latin American countries, which tend to have defective demographic data and censuses with distinct coverage levels.

There has not been yet any census of the region able to achieve and have a response from all the inhabitants in their territory at the time of the survey. It is likely that the evasion to answer, rejection and false responses will continue, meaning that there will still be under-represented and incomplete counts of people within households. There will also be non-detectable errors in the census operations. In short, there will be under coverage. Furthermore, in some population groups, the least socially integrated, recent migrants, isolated people, who do not have legal status, which are constantly moving, the poor and the less educated, those indifferent to public charges, are the

usually the census register does not count. In short, the omission will be differential. Developing methods able to deal with this issues is still an unfilled demand.

## **CONCLUSION**

In general, census coverage in Latin America showed throughout the analysis of various censuses conducted in the region since the 1950s, presented a downward trend in omission levels. Factors such facility of accessing remote areas, due to improvements in infrastructure, transport, technology, partly explain these improvements. Increasing educational levels of the population also tends to improve the quality of information, though it is not clear how this could affect census coverage. Despite the general trend of data quality improvement, there have been many fluctuations over time, which depend on a series of structural and conjunctural factors driven the production of statistics in Latin American countries.

In addition to analyzing the structural and cyclical factors affecting the quality of the census, future research should look at which subpopulations remain omitted in censuses in Latin America. While studies addressing this issue indicate strong differences in terms of sex and geographic region, one would find the spreads are maintained over time, not only for these two variables, but other features.

Although drawing an association amid census coverage and defective census process was a challenging task, we think that a substantial improvement can be made in this path. While it is essential to measure census coverage as we did, moreover, it would be particularly useful to use that assessment to develop the quality of subsequent censuses. Therefore, a central procedure of coverage measures is to support and recognize principal sources of coverage errors and feasible to recommend complementary process to moderate the incidence of those errors in the upcoming round. With the historical information available, the tradeoff concerning the outcome on accuracy and on census process cost might nowadays be better comprehended, so this evidence can subsequently be used to assign resources for developing alternative census designs and processes that might offer count with higher quality in 2020. It is

also possible to consider the idea that the use of such a feedback might likewise offer savings in census cost, in addition to quality.

## **BIBLIOGRAPHY**

- Arretx, C., Chackiel, J., 1985. Algunos problemas relativos a la recolección de datos demográficos en los censos de población de América Latina en la década de 1980. Notas de población.
- Baldrige, M., Brown, C., Jones, S., Keane, J., 1985. Evaluating Censuses of Population and Housing. Department of Commerce, United States of America/US Bureau of the Census.
- Bravo, D., Larrañaga, O., Millán, I., Ruiz, M., Zamorano, F., 2013. Informe final - Comisión externa revisora del censo 2012.
- CELADE, 1984. Métodos para proyecciones demográficas (No. Vol. E). Centro Latinoamericano de Demografía, San José, Costa Rica.
- CELADE, 1968. Métodos de evaluación en los censos de población: Algunas aplicaciones hechas por CELADE. A/83, Santiago.
- Chackiel, J., 2009. Evaluación y estimación de la cobertura en los censos de población: la experiencia latinoamericana.
- DANE, U., 2014. Taller Experiencias de los Censos de Población y Vivienda de la Ronda de 2010 - Documento Memorias. Bogotá, Colombia.
- DGEEC, 2015. Paraguay Proyección de la Población Nacional, Áreas Urbana y Rural por Sexo y Edad, 2000-2025 - Revisión 2015. Dirección General de Estadística, Encuestas y Censos, Fernando de la Mora.
- Gerland, P., 2014. UN Population Division's Methodology in Preparing Base Population for Projections: Case study for India. Asian Population Studies 10, 274–303. doi:10.1080/17441730.2014.947059
- González, L., Ribotta, B., Torres, E., 2013. Estimación indirecta de la omisión del Censo 2010 a través de proyecciones de población. Argentina y provincias seleccionadas. Presented at the XII Jornadas Argentinas de Estudios de Población, Bahía Blanca.
- IDB, 2011. Program to Develop the National Statistical System Through the 2012 Population and Housing Census.
- INDEC, 2010. Censo Nacional de Población, Hogares y Viviendas 2010 (Documento metodológico preliminar). INDEC, Buenos Aires.
- INDEC, 2005. Metodologías aplicadas para estimar la cobertura de población en el Censo Nacional de Población, Hogares y Viviendas de 2001. INDEC, Buenos Aires.
- Massa, C., Bankirer, M., Raimondi, M., 2005. Un aporte metodológico para el estudio de la cobertura del censo 2001 a nivel nacional. Presented at the VII Jornadas Argentinas de Estudios de Población.
- Mortara, G., Whelpton, P.K., Grauman, J.V., Behm, H., Montenegro, T.H., Miró, C.A., Kirk, D., Davis, K., 1964. Appraisal of Census Data for Latin America. The Milbank Memorial Fund Quarterly 42, 57–85. doi:10.2307/3348658

- Oliveira, L.A.P. de, Simões, C.C. da S., 2005. O IBGE e as pesquisas populacionais. *Rev. bras. estud. popul* 22, 291–302.
- Perfit, J.T., Russell, M., Moncada, G., Muñoz Ayala, J.E., 2012. Estimación de algunos efectos y determinantes del uso de distinta tecnología en los censos de población de América Latina y el Caribe. *Notas de Población*.
- Ruiz, M., 2013. Análisis de cobertura. Presented at the III Reunión regional sobre evaluación y estimaciones demográficas con base en información censal, Santiago de Chile.
- Sacco, N., 2015. ¿Cierran los números con la gente adentro? Principales resultados de una evaluación general del censo argentino de 2010. Presented at the XIII Congreso de la Asociación Argentina de Estudios de Población (AEPA), Ciudad de Salta, Provincia de Salta, Argentina.
- Tacla Chamy, O., 2006. *Omission Censal en America Latina 1950-2000*. United Nations Publications.
- United Nations, 2015. *World Population Prospects: The 2015 Revision*. Department of Economic and Social Affairs, Population Division.
- Wheldon, M.C., Raftery, A.E., Clark, S.J., Gerland, P., 2013. Reconstructing past populations with uncertainty from fragmentary data. *Journal of the American Statistical Association* 108, 96–110.
- Whitford, D.C., Banda, J.P., 2002. ¿Vale la pena hacer encuestas de post-empadronamiento censal? *Notas de población* 75.
- Yuri, C.B., Carlos, H.F., 2014. Censo de población chileno 2012: de la frialdad cuantitativa a las pasiones estadísticas. *Estatística e Sociedade*.

---

<sup>1</sup> The true value is actually an unknown parameter, but it can be estimated through different methods, which will be discussed further.

<sup>2</sup> Spanish acronym for the Economic Commission for Latin America and the Caribbean

<sup>3</sup> According to her revisions, Argentina maintained a default equivalent to just over 1% in 1980 and 1990; in Chile was very similar, rose 1.6 to 2.0%, Guatemala fell from 15.6 to 14.5%, Panamá from 6.1 to 3.1% in 1990 and Paraguay from 8.4 to 7.1%. Draws attention to the cases of Colombia and Guatemala, whose omissions exceeded the 10% of the registered population. The first with 11.3% in circumstances in the census that was detected above 5.8%, i.e. it deteriorated significantly. Meanwhile, in Guatemala's new review show a decline in the decade 1980-1990 despite continued with an omission almost 15%. Compared with the rest, censuses of both countries can be considered the poorest in terms of coverage of the 1990s.