Evaluation of Brazil's population census coverage and quality through demographic analysis

Gabriel Mendes Borges

1. Introduction

Population and Housing Censuses are the main activity undertaken by an official institute of statistics, and are also the most important source of demographic information in a country, principally for small geographic areas. The existence of coverage and quality problems is inherent in activities of this magnitude. Thus, it is essential to evaluate and measure such problems, in addition to a continuous search for improvements of the collected information.

Nevertheless, census figures that are subject to error are still valuable if data limitations are understood by users and if the errors do not adversely affect the major data uses. In that sense, evaluation studies which examine the results, procedures and operations used in undertaking a census are necessary for providing both the producers and users of the data with information needed to assess census quality (U.S. Bureau of Census, 1985).

At the time of census result publication, the Institute of Official Statistics has a commitment with transparency regarding the quality of the statistics produced, which comprises providing users with information necessary to evaluate the quality and accuracy of questionnaire answers.

> *"The purpose of census evaluation is to provide users with a level of confidence when utilizing the data, and to explain errors in the census result. It is therefore important to choose an appropriate way of sending out these messages to the right group of people."* (United Nations, 2008)

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.

The uses of the evaluation census process results can be made in several ways: guiding improvements in future censuses and surveys; assisting users in their interpretation of the results; adjusting the census results (U.S. Bureau of Census, 1985).

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, 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 (CEPAL, 2009).

2. Objectives

This paper aims to present an indirect evaluation of Brazil's 2000 and 2010 Population Censuses in order to achieve the main purposes of an evaluation, especially adjusting census results with the purpose of utilizing these adjusted populational figures as the base population used for a starting point of projections and estimates.

While recognizing the importance of the PES and the combined use of direct and indirect techniques in the assessment censuses programs, this paper uses only the indirect technique for measuring the coverage of the Census through demographic analyzes[1].

For the main purpose of this work, graph analyzes and comparisons between the population structure observed with the "expected" pattern using some indicators like sex ratio and census survival ratios (CSR) are used. Based on these supports, it is used the "Demographic Reconciliation" method in order to obtain a more reliable population number which is finally compared to populations observed in censuses. Besides measuring the coverage error, especially in certain ages in which omission is greater, like children, these procedures also enable the estimation of some kinds of content errors, like age exaggeration in elderly populations.

[1] In the Brazilian population projections, the adjustment factor obtained for the 2000 Brazilian Census by conducting demographic analysis was used as a parameter for adjusting the populations from each state. As it is known that omissions in the censuses are differential by geographic regions, the factor estimated for the country as a whole was weighted using coverage rates from the 2000 Brazilian Census Post Enumeration Survey (PES) (IBGE, 2013).

It is important to clarify that no technique has the ability to correct the census data with absolute precision. What has always been sought is eliminating irregularities, evaluating and adjusting not-expected behaviors from the knowledge of the characteristics of the studied population, looking for a more believable populational structure. The objective is using the censuses as the main source of demographic information, seeking coherence between them and the demographic components. Care is also necessary in order to not overcorrect the census data.

3. Census Population by age and sex

The structure of the population by age and sex provides subsidies for the calculation of several indicators, allowing change assessment of the sociodemographic profile of the population over time. This information also constitutes a powerful tool which fulfills the purpose of supporting public policy planning, and provides parameters to be considered in evaluating processes of several social and economic programs.

There are several problems that normally affect information about the age statement of people in a census or survey. Firstly, it is important to say that it is very common for only one person interviewed to provide the information for the entire household unit members, which can result in not knowing date of birth and/or declared age of other household members.

A significant problem that affects the people's age information is the preference for age with terminal digits, in that some people tend to round their ages, accumulating statements in numbers ending in 0 and 5. In the Brazilian censuses, an important improvement in in this issue has been observed in the last 50 years, as a result of advances in information capture, in addition to the decreasing difficulty of the population to provide their age precisely, which reflects the improvements in information quality indicators. Despite these improvements, this trouble still remains, but can be minimized by grouping ages by five year groups.

An additional and recurrent error observed in many populations is that the elderly tend to declare themselves older than they really are, an increasing trend with growing age (COALE; CASELLI, 1990; PRESTON; ELO; STEWART, 1999; DEL POLOLO, 2000; ROMERO; FREITEZ, 2004).

Besides these content errors mentioned before, an additional element significantly affecting the sex-age structure of the population is the differential of coverage within these two variables. This may vary from region to region, depending on particular socioeconomic and demographic characteristics. However, some facts seem to be recurrent in different places, like the omission of children and the differential undercount by sex.

Regarding children undercount, there are several factors that may be related to this differential omission, one being cultural issues, where a child could not be viewed as a person, nor their existence being considered (CHACKIEL, 2009). Furthermore, the propensity of the occurrence of this problem can be enhanced if the characteristics of households are connected to a larger number of children,

especially due to differential fertility in accordance with these characteristics. It is known that fertility is higher in regions where enumeration is more difficult (WEST, ROBINSON 1999; O'HARE, 2009).

Another characteristic that affects population distribution by sex and age is the differential undercount by sex, especially among young adults, in which the male population tends to be more undercounted that the female. This differential coverage may be related to family and work relationships, in which adult men have greater activity rate in the labor market than women, besides greater representation in single-person households.

4. Evaluation of Brazilian censuses through demographic analysis

Two of the most basic demographic indicators – Census Survival Ratios (CSR) and Sex Ratios (SR) – provide important contributions to understanding the main problems in the population by sex and age information discussed in the previous sections.

CSR refers to the ratio of the observed population in the census in a given age group and this same population ten years younger in the previous census. This indicator expresses the evolution of the same cohort over 10 years and can be formally written as:

$$CSR_{s,(x+10,x+15)}^{t,t+10} = \frac{P_{s,(x+10,x+15)}^{t+10}}{P_{s,(x,x+5)}^{t}}$$

where $P_{s,(x,x+5)}^{t}$ is the population at year *t*, sex *s* and age group between *x* e *x+5*.

The Sex Ratio shows the number of men per 100 women in a certain year and age group and could be expressed by the following formula:

$$SR_{(x,x+5)}^{t} = \frac{P_{h,(x,x+5)}^{t}}{P_{m(x,x+5)}^{t}}$$

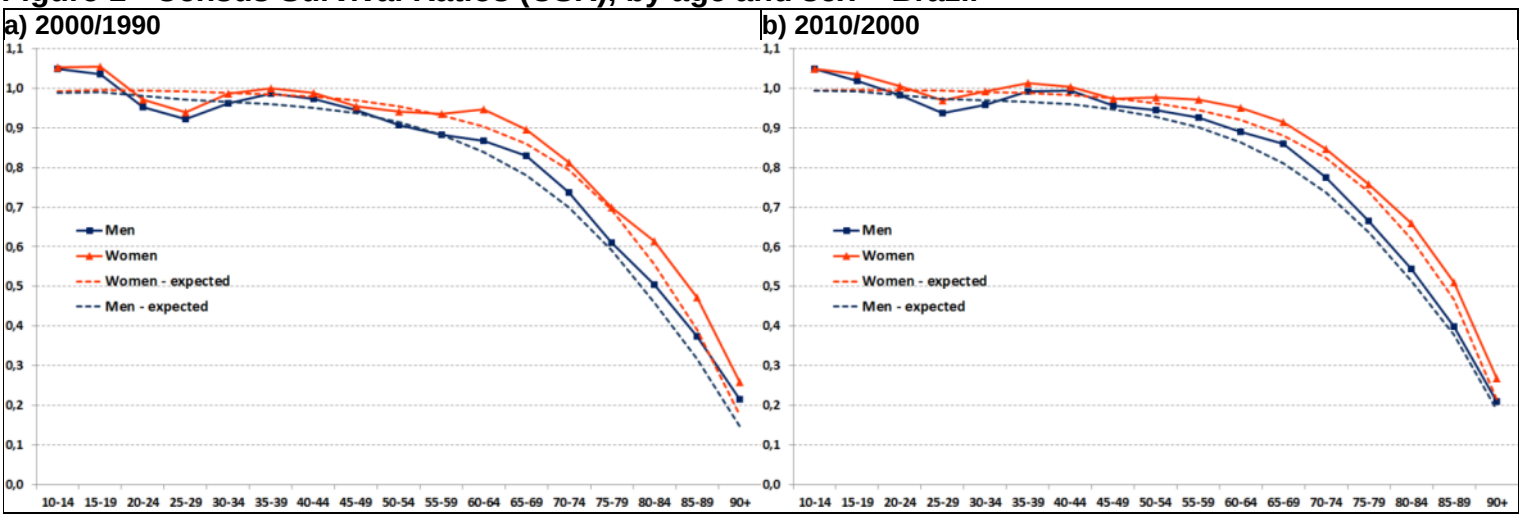
Figure 1a shows the census survival ratios (CSR), by age in 2000, between the 1990[2] and 2000 censuses and Figure 1b shows the same information for the 2000 and 2010 censuses, respectively. They show similarity between both CSR of 2000/1990 and 2010/2000, presenting a similar coverage and content error pattern. It is noted that CSR curves both from men and women do not follow an expected pattern in a closed population, without differential coverage between two censuses and without problems in age report. The theoretical pattern[3] always shows values greater than 1.0, with a decreasing tendency with the increasing of age – with the exception of the first groups – as a result of the mortality rate increase. It would be

[2] The population of 1991 Census was shifted to 1990.

[3] The theoretical pattern (expected CSR) was calculated using life tables from 1990, 2000, 2010 and their interpolations.

expected also the curve for women being always superior to that for men, reflecting the mortality differentials by sex, favorable to women.

Figure 1 - Census Survival Ratios (CSR), by age and sex – Brazil



Source: IBGE – Brazilian Demographic Census of 1991, 2000 and 2010

For the first two age groups the CSR is greater than 1.0, indicating that there would be fewer children from 0 to 4 years old and from 5 to 9 years old in 1990 than 10-14 and 15-19 ten years later, in 2000. Because an immigration of children at these ages is unlikely, without the correspondence of this phenomenon in young adults, this indicates clearly an undercount of children in the first census, which is higher among 0-4 age group. The same pattern is observed comparing 2000 and 2010 censuses.

The analysis of the CSR for the 25-29 age group also shows an incompatibility on data from the same cohort in the period under examination (1990/2000). However, differently from the younger age groups, the observed CSR are lower than the expected values. The literature has shown that the 10-14 and 15-19 age groups are traditionally well-enumerated (CHACKIEL, 2009). So, an undercount of the 25-29 population is likely, which is consistent with the known young adult's coverage patterns.

Next age groups show closeness between observed and expected women CSR curves from 30-34/20-24 to 55-59/45-49, demonstrating consistency between censuses at these ages.

The gap between observed and expected CSR curves both from men and women for elderly people can be explained by age exaggeration of this group.

These differences could also be justified by international migration, errors in estimating mortality, which is used to calculate theoretical CSR and differentials between two census enumeration. Regarding the first explanation it is unlikely to occur, according to the knowledge of migration patterns in Brazil, a significant international migration, especially among the elderly (CAMPOS, BORGES and

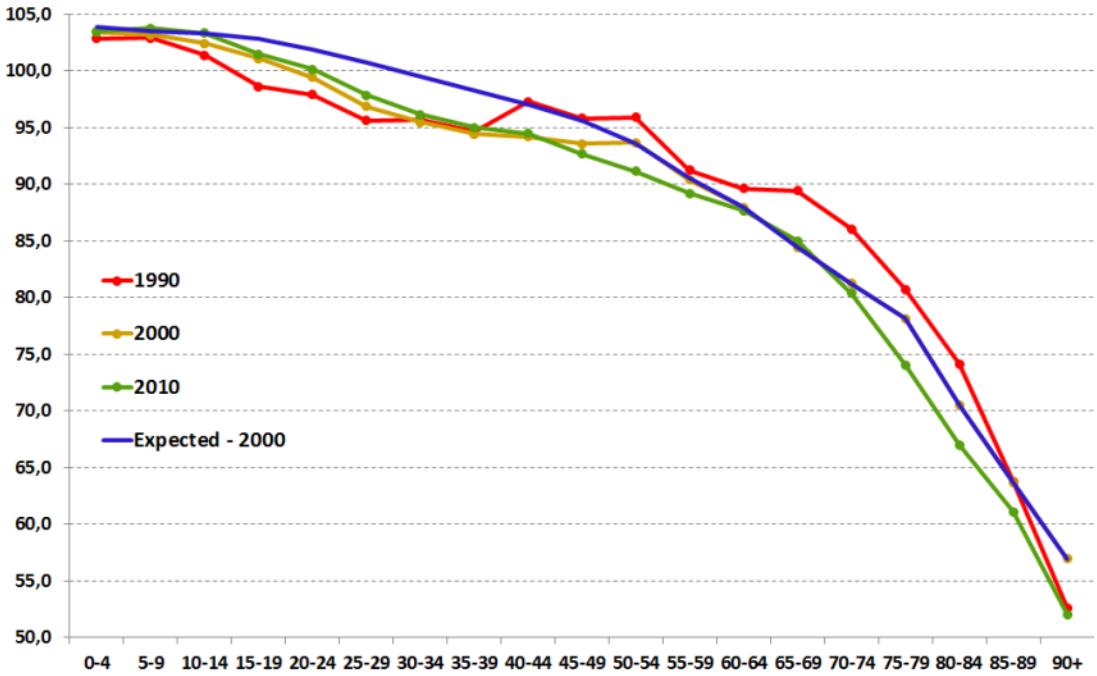
SILVA). Concerning enumeration differentials, it is difficult to assume that older people are better enumerated than these same cohort ten years earlier. With concern to the explanation by errors in mortality estimates, simulations considering changes in the correction of deaths factors through indirect techniques, also showed little impact on the differences founded between CSR observed and estimated. Thus, the most plausible explanation for this phenomenon is related to reporting errors in age between the elderly populations.

Errors in this age group tend to be toward overkill in declared ages, which causes over-enumeration of older ages. The effect of the cohort size is added to this effect, in which size decreasing is very rapidly at these ages (CONDRAN, HIMES, PRESTON, 1991; DEL POLOLO, 2000). Overkilling age by 5% of the male population from 80 to 84 years old for the next group in 2010, for example, would increase by 13% the 85-89 population. Besides the trend to exaggerate their own age, it is also important to note that the percentage of people who provide their own information tends to decrease rapidly for elderly, reaching only 20% for those who over 90 year old.

Figure 2 presents observed and expected[4] Sex Ratios from 1990, 2000 and 2010, showing differences between the expected and observed sex ratios from 10 to 49 years, evidencing that a differential undercounted by sex remains, with greater omission for men at these ages. Despite predictable fall in sex ratio with the increasing of age, the figures observed in demographic censuses do not match the expected trend for this indicator, even considering the high overmortality existing in Brazil, especially among youth.

[4] Expected Sex Ratio was calculated using a population projection with base population in 1940, in which population was subjected to mortality rates calculated from 1940 to 2010. The resulting sex ratio does not depend on fertility, but only the sex ratio at birth and mortality differentials by sex in that period.

Figure 2 – Observed and Expected Sex Ratios – 1990, 2000 and 2010 – Brazil



Source: IBGE – Brazilian Demographic Census of 1991, 2000 and 2010

5. Demographic Reconciliation

Population censuses assessment was taken from a demographic reconciliation. This process, combined with the cohort-component method, is a tool that aims to obtain adjusted structures by age and sex from the enumerated population in the census using the knowledge about the parameters that represent the country's demographic dynamics.

Considering the principle that there are no prescriptions that could be automatically applied in each case, more than a statistical / mathematical procedure, it is an analysis of the data with demographic criteria. This technique is based on own census data besides the knowledge and experience accumulated over the error patterns that affect this information. From there, the most plausible demographic dynamics for the country are represented and, as a by-product, the functions of census error can be obtained by sex and age. This dynamics includes, besides the estimation of population by age and sex, the components of growth that explain its evolution: fertility, mortality and international migration (CHACKIEL, 2009).

In the exercise made for this paper, 1990 and 2010 populations were, respectively, projected and back-projected to 2000 – using fertility and mortality estimations – in order to compare with the observed population in the 2000 census.

Based on the previous analysis of the sex ratios and the census survival ratios, resulting in coverage and quality evaluation, age groups of each census were chosen.

In general, the choice of the information to be used for each group was as follows:

• not using the information observed in Census 2000 for groups 0-4 and 5-9. Population for 0-4 group was estimated from an average between the group generated from the estimates of births between 1995 and 2000 and the back-projected 10-14 population. For 5-9 group the average was made again, but only using the female population, since the sex ratio in 2010 for 15-19 age group have presented values lower than expected, indicating a higher undercount between men comparing with women;

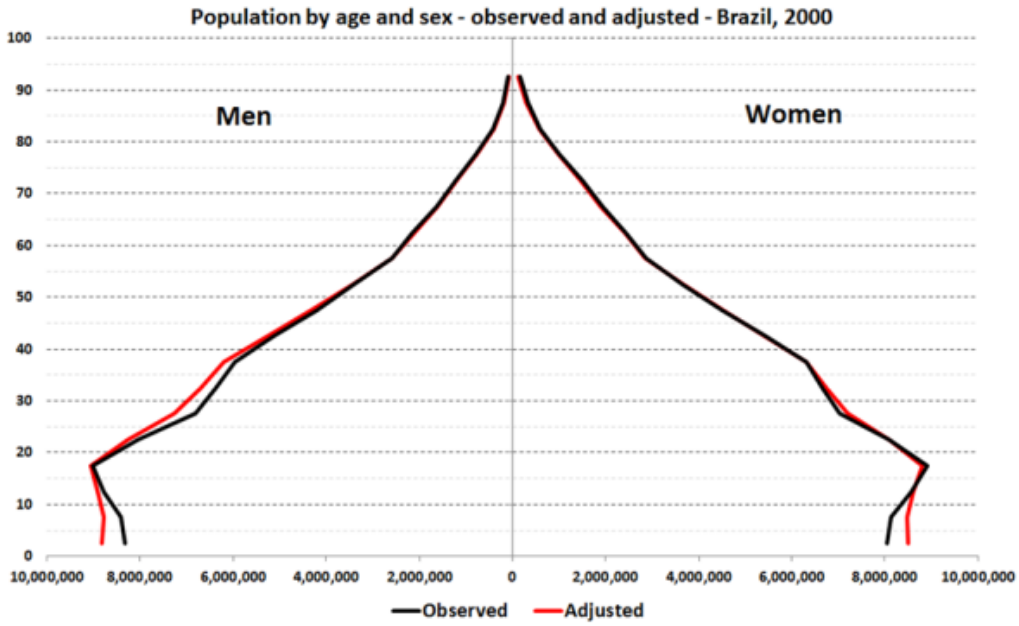
• from 10 to 49 years old, due to higher male population underreporting, only information coming from women was used, preferably those well-enumerated groups discussed above, and;

• for age groups older than 50 years old, information was used both from men and women, since, for these ages, a certain consistency was shown in the observed sex ratios in demographic censuses. Regarding the choice of the groups according to the censuses used, it was seen that the biggest problem affecting the enumeration at these ages comprises errors in age declaration, in which there is a general tendency to exaggeration, an increasing trend with age. Thus, the information coming from the 1990 Census was prioritized, as it refers to groups of younger ages. Descriptive analyzes and literature considered show that people who were between 60 to 64 years in 1990, for example, would offer more reliable information than the same cohort 20 years later, from age 80 to 84 in 2010.

Figure 5 shows the observed and adjusted structure by age and sex in 2000. It shows, for both sexes, an omission of children from 0 to 9 years old. Furthermore, omission of men at working age was also observed, being higher in the 20-24 age group, a range that also presents a minor omission among women. For the elderly, over-enumeration was observed, which is, as discussed above, associated to problems related to age statement among this group.

Demographic reconciliation resulted in a total population adjustment around 2.1% for the 2000 Census.

Figure 5 - Population by Age and Sex - Observed and Adjusted - Brazil, 2000



Source: IBGE – Brazilian Demographic Census of 1991, 2000 and 2010

6. Conclusions and Challenges

There are several types of errors in the census information, which can differently affect the population by age and sex. The exercises of evaluating and adjusting population lead to a more reliable estimate, which can be used, among other ends, as the base population for projection starting point.

An important challenge is the coverage estimation for small areas, like states and municipalities, areas in which this demographic technique cannot be performed. Another issue is related to looking for more accurate fertility, mortality and migration estimations, in order to improve the quality of the basic information used in the method.

7. References

CAMPOS, M. B. de; BORGES, G. M.; SILVA, L. G. de C. e. Estimativas de migração internacional no Brasil: reversão do saldo migratório internacional negativo. In: ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS, 18., 2012, Águas de Lindóia. Anais.- ABEP, 2012.

CEPAL, *Evaluación y Estimación de la Cobertura en los Censos de Población: la Experiencia Latinoamericana*, Seguimiento a los Avances de la Preparación de la Ronda de Censos 2010 en America Latina: Taller del Grupo de Trabajo de la CEA/CEPAL, Santiago de Chile, 2009

CHACKIEL, J. Evaluación y estimación de la cobertura en los censos de población: la experiencia latinoamericana. “Seguimiento a los avances de la preparación de la ronda de censos 2010 en América Latina: taller del Grupo de Trabajo de la CEA/CEPAL”, Santiago de Chile, 2009.

COALE, A. J.; CASELLI, G. Estimation of the number of persons at advanced ages from the number of deaths at each age in the given year and adjacent years. Genus, Roma: Università degli Studi di Roma “La Sapienza”, v. 46, n. 1-2, p. 1-23, Jan./Jun. 1990.

CONDRAN, G. A.; HIMES, C.; PRESTON, S. H. Old age mortality patterns in low-mortality countries: an evaluation of population and death data at advanced ages, 1950 to the present. Population Bulletin of the United Nations, New York: United Nations, Department of Economic and Social Affairs, n. 30, p. 23-60, 1991.

DEL POPOLO, F. Los problemas en la declaración de la edad de la población adulta mayor en los censos. Santiago de Chile: Comisión Económica para América Latina y el Caribe - Cepal, Centro Latinoamericano y Caribeño de Demografía - Celade, 2000. 53 p. (Serie población y desarrollo, n. 8).

IBGE, Projeções da População – Brasil e Unidades da Federação. Série Relatórios Metodológicos. Volume 40. Rio de Janeiro, 2013

O’HARE, W. P. Why are young children missed so often in the census? Kids count working paper, Baltimore: The Annie E. Casey Foundation, Dec. 2009. Disponível em:http://www.aecf.org/~/media/Pubs/Other/W/WhoAreYoungChildrenMissedSoOftenintheCensus/final%20census%20undercount%20paper.pdf>. Acesso em: out. 2013.

PRESTON, S. H; ELO, I. T.; STEWART, Q. Effects of age misreporting on mortality estimates at older ages. Population Studies, London: University of London, London Scholl of Economics and Political Science, v. 53, n. 2, p. 165-177, July 1999.

ROMERO, D.; FREITEZ A. Problemas de calidad de la declaración de la edad de la población adulta mayor en los censos de América Latina de la ronda del 2000. 2004. 32 p

U.S. Bureau of Census, *Evaluating Censuses of Population and Housing*, Statistical Training Document, ISP-TR-5, Washington D.C, 1985.

United Nations. *Principles and Recommendations for Population and Housing Censuses*, *Revision 2*. Statistical Papers, Series M nº 67/Rev.2. Sales nº. E.07.xvii.8., 2008

WEST, K. K.; ROBINSON, J. G. What do we know about the undercount of children? Washington, DC: U.S. Census Bureau, 1999. 13 p. (Population Division working paper, n. 39).