# Assessment of drawing impacts of the questionnaire in Brazilian Census of 2010

Luiz Felipe Walter Barros[1,2], Gilson Gonçalves de Matos[1], Marden Barbosa de Campos[1], Gabriel Mendes Borges[1]

[1]Brazilian Institute of Geography and Statistics, Rio de Janeiro, Brazil
[2]Corresponding author: Luiz Felipe Walter Barros, e-mail: luiz.w.barros@ibge.gov.br

## Abstracts

The Brazilian censuses questionnaires of 1980, 1991 and 2000 had one sheet per person, forcing the reference to each person specifically in each question. For the 2010 census, with the change of the questionnaire on paper by the electronic, filters and jumps were automated and the questions were asked by block, in other words, all household members responded the same question every time. This format has brought significant operational advantage by facilitating fieldwork in the data collection. On the other hand, may have brought a worsening quality of information, especially with regard to a tendency of generalize the answers for all household members. In this sense, the objective of this study is compare the results of the 1980, 1991, 2000 and 2010 censuses, and of two other Brazilian surveys, evaluating the cases with occurrence of the same answer among all household members. Preliminary analyzes have showed that for certain variables, especially for households with more residents, the census of 2010 tends to present a higher occurrence of cases with the same answer for all dwellers, which could not be explained solely by socioeconomic and behavioral changes in the last 10 or 20 years. This may indicate a tendency to generalize the responses as a consequence of changes in the form of collection the information.

Key Words: Demographic censuses, information quality, questionnaire design

## 1. Introduction

According Chackiel (2009), efforts to conduct a census of quality should be a central concern of statistical institutes and must involve, besides the development of processes to quantify the errors of coverage and quality, the adoption of new procedures for their evaluation and correction. The implementation of evaluation projects provides security in the use of data, which may explain any problems/errors in census results (UN, 2008). Nevertheless, these evaluations have often been neglected in comparison with other stages of the project.

With the experience of conducting population censuses in recent decades by statistical institutes around the world, a greater democratization of information and improvements in coverage and fieldwork is expected, facilitated by greater access to the territory and the introduction of new technologies. On the other hand, the coverage and quality of censuses depend on other factors related to the complexity of modern life and the organization of society, some of which tend to hinder this task. Moreover, in recent decades, there has been an overload on the census questionnaires (with the inclusion of new questions and new issues) (Chackiel, 2009).

With the introduction of the electronic questionnaire, new possibilities in data collection emerged, bringing new challenges to statistical institutes. In Brazil, for example, the census questionnaires in 1980, 1991 and 2000 were designed so that each person responded to all the questions individually. For the 2010 census, with the change of the questionnaire on paper by the electronic, filters and jumps were automated and the questions were asked by block of residents, in other words, all members of the household responded to the same question at the same time.

There are studies that show the effects on the quality of data related to changes in the questionnaire design. Hess *et al*. (2001) present results of an Experimental Research of the U.S. Census Bureau, 1999, which included a test comparing questions on person level with questions on household level. In the first approach, the questions were made individually to each person, while the questions on the household level sought to identify if someone in the household has the characteristic of interest, to then identify the individuals. The authors found some evidences that the use of the questionnaire on household level increases the risk of undercount compared with a drawing of questions at the person in some specific topics. The nonresponse is usually similar for both treatments. Overall it was not proven superiority for either design. Finally, they found an increase in efficiency of the interview with the drawing at the household level, and evidence that interviewers prefer this style. Although these studies were not conclusive in favor of one or another drawing, there is a clear indication of the need to perform tests before opting for one or another design, weighing up advantages and disadvantages. It is noteworthy that after the studies, the questionnaire at the person level was adopted in the American census of 2000.

In Brazilian Census of 2010 the new format of the questionnaire brought significant operational advantages, facilitating fieldwork in data collection and making the process less tiring for respondents and interviewers. On the other hand, may have had a negative impact on the quality of some information, particularly by a tendency to generalize the responses for all household members. In this sense, the purpose of this article is to analyze the trends presented in the censuses of 1980, 1991 and 2000, in the National Household Sample Survey (PNAD) of 2001 and 2009 and in the Household Budget Survey (POF) of 2002-2003 and 2008-2009 with the results of the census of 2010, assessing cases of occurrence of the same answer for all household members for selected variables.

## 2. Methods

It is believed that the change made in the drawing of the questionnaire can cause at least two effects:

*i* - Generalization of answers for independent questions[3] ;

*ii* - Confusion in filling dependent questions[3] ;

The aim of this paper was to verify the existence of the effect *i*, in other words, if the response of one question at the same time by all the dwellers may have brought some bias, especially as regards to a greater tendency to generalize the answers. For this, we used two strategies: 1) to compare the trend in the last decade with those presented in previous decades, using data from the last Brazilian censuses - 1980, 1991, 2000, and 2) to compare the trends of the last decade found in the census data with data from PNAD (2001 and 2009) and POF (2002-2003 and 2007-2008).

We selected four variables common to these censuses and other surveys: color or race, was born in this city, religion, and can read and write. The results were controlled by the number of household members.

## 3. Results

The Figure 1 (a, b, c, d, e, f) shows the percentage evolution of households with the same response for all members in the selected variables for the three surveys used in this study (Census, PNAD and POF). Observing figure 1 (a, b, c), which refers to the variable "color or race", as it increases the number of residents in the household there is a tendency of decrease the percentage of those with the same response for all

---

[3] It is understood as independent those questions whose answers do not depend on previous questions (color and race, religion, etc.), and how dependent those whose answers that depend of other questions (fertility, education, labor, etc.).

members, in all the censuses. Regarding the evolution of the indicator, it is observed that for all households with at least two residents there is a significant decrease in these percentages between 1980 and 2000, with a slight increase between 2000 and 2010 according to census data. It is also noticed that the trend reversal in the series is more abrupt in households with more residents.

Ribeiro and Silva (2009) show that there is evidence of increasing interracial marriages in Brazil and therefore a gradual rise of miscegenation between 1960 and 2000, indicating that Brazilian society seems to become significantly more mixed, with marriages between people of different educational levels and races. The authors also report a growing trend of acceptance of different racial groups and a strong tendency of reduction the barriers between them. The information of Figure 1 (a, b, c) goes against this trend and most seems to indicate a difference due to the method of capturing the information that a change in the miscegenation patterns of families and households, confirming the influence of the questionnaire design in capture the phenomenon.

In order to check if indeed there was a trend reversal, we investigated the trend in the last decade observed in the PNADs and POFs, which are IBGE surveys with the same information captured of a very similar form, but individually. Note that the choice of PNAD as a comparison is interesting, because in addition to not perform the research questions collectively as the census, the search began using electronic questionnaire from 2007. Thus, we are "isolating" the effects of "electronic questionnaire" in the results, although of an approximate and indirect form (but the only way available). We should mention that our intention is to compare the trends of the curves, not their levels. Because of the specificities of each survey, we expected different results between Census, PNAD and POF, but pointing to the same direction.

From Figure 1 (a, b) it is clear that the tendency was reversed only in the census for the last decade. Both PNAD and POF indicate an increase in diversity within households, trends very close to those presented by censuses in previous decades.

We also inquired whether in households with more than one informant this effect would be less, namely, if there would be less tendency to repeat the same answer for all residents, since each person supplied their own information. For this, we performed a filter by selecting only those households in which all adult residents (18 years) supplied their own information (Figure 1c). Surprisingly, these households concentration of same answer for all residents was higher than average, which goes directly against the initial expectations. The only explanation to justify such result is that the question "Assine who provided the information" was also performed in block, and may have suffered a similar effect of the variable "color or race".

The second variable analyzed was the question "was born in this city". This variable belongs to the migration topic. Figure 1d shows the evolution of homogeneity in the responses of households through the last four censuses and its comparison to the last decade with the data from PNAD. According to this chart, we can see a slight decrease in the percentage of households with the same answer for all residents between 1980 and 2000. This trend was reversed in the 2010 census, when there was a significant growth in this percentage. As this variable indicates a stock of migrants and non-migrants, from the perspective of naturalness (being cumulative result of population movements of the past), great changes in this indicator were not expected between 2000 and 2010, period of slight reduction in Brazilian migration. Just as in the variable on "color or race", the trends presented in Figure 1d by PNAD are very similar to those of censuses for previous decades, namely, a slight downward trend or small increases (for households with more than 9 dwellers), which keeps the average falling as expected. Again, the effects found in census data seem to be more related to the change in the form of data collection than with changes in demographic trends.

**Figure 1 - Percentage of households whose response was the same for all residentes by total residents for selected questions. Brazil, 1980 - 2010.**
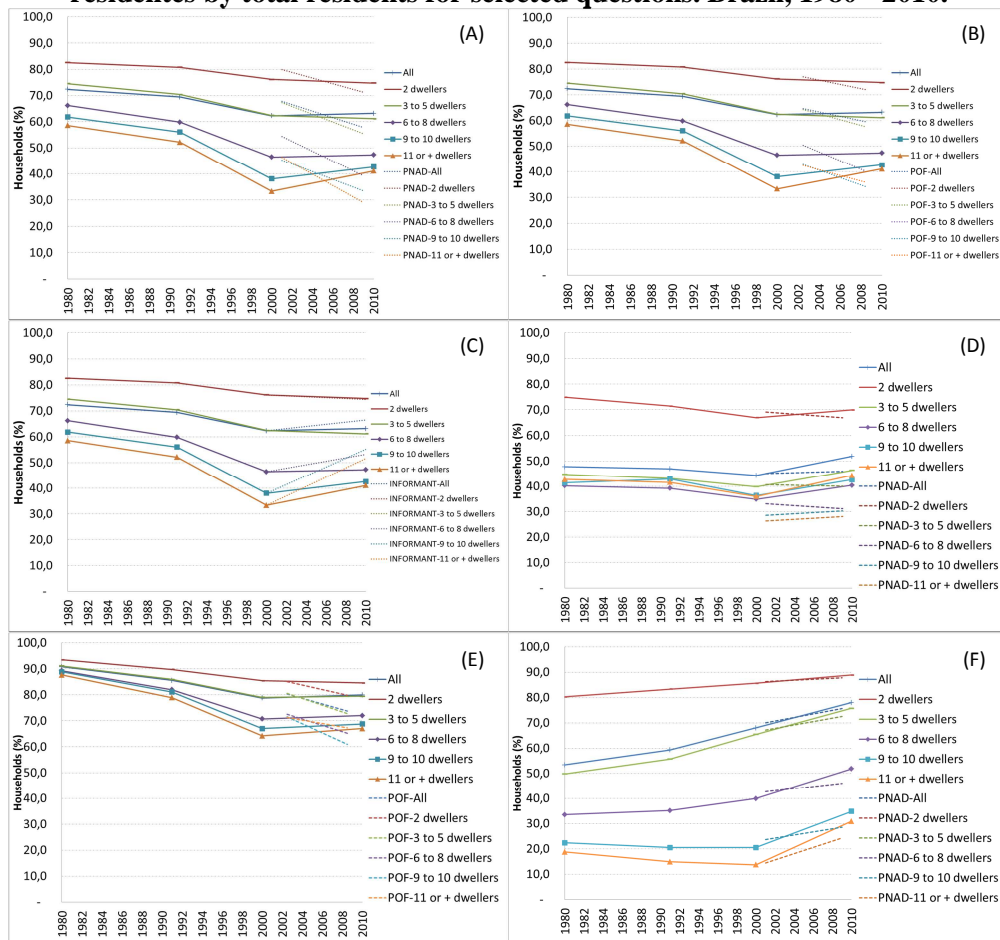


Note: A,B,C: question "race/color", comparing censuses trend with PNAD (A), with POF (B), and when all adult residents (18 years) supplied their own information (C). D: question "was born in this city", comparing censuses trend with PNAD. E: question "religion", comparing censuses trend with POF. F: question "do you know reading and writing", comparing censuses trend with PNAD. Households with 2 or more residents.
Source: Brazilian Censuses of 1980, 1991, 2000 and 2010, PNAD of 2001 and 2009, POF 2002-2003 and 2008-2009.

The homogeneity of responses in the Religion variable was analyzed too. This homogeneity presented a significant drop between the years 1980 and 2000. In the last decade, however, the census data show an increase in the percentage of households in which all residents have declared the same religion. This indicator also showed differences according to the number of household members: those with two residents continued their downward trend, although milder than in the previous decade, while the remaining households showed a reversal of the trend, increasing with the number of residents. The trends until 2000 are directly related to the loss of Catholic hegemony (already established in the literature), giving way to evangelicals and those without religion, among others (Alves *et al.*, 2013). The decline of religious diversity within the households also goes against the trends observed in the actual 2010 census, of growing diversity of religious groups in Brazil.

Also according to Figure 1e, the differences in trend verified between census and POF are clear for the last decade. It is important to remark than the trends presented by POF are closer of the expected.

The analphabetism was the last phenomenon analyzed, through the question "can you read and write?". Figure 1f shows a substantial increase in the percentage of households whose answers were the same for all residents over the past censuses. This

trend is related to the increase literacy in recent years in Brazil, and consequently a significant reduction in the illiteracy rate among people of 15 years or older, of 25.5% in 1980 to 9.6% in 2010 (IBGE, 2011). This means that there are an increasing percentage of households in which all members have declared to read and write. From Figure 1f is not possible to observe a clear change in the trend, as noted in previous analyzes. However, the growth in the proportion of households with the same answer is quite steep in the last decade for larger households, which may indicate some influence of the new design of the questionnaire. This trend is even clearer when comparing the results of census and PNAD. For all sizes of households listed, the trends observed in the PNAD for the last decade have inclinations less than those of Census, especially for larger households.

## 4. Conclusions

With the change of the questionnaire on paper by the electronic, filters and jumps were automated and all members of the household answered each question at the same time (the questions were asked by blocks). This format has brought significant operational advantages in terms of practicality and flexibility on the interviews. However, the presented results indicate that this new design of the questionnaire may have caused an impact on data quality.

The objective of this work was to verify the effect of generalization in answers of independent questions, in other words, if the design of the questionnaire in block brought some bias, especially with regard to a tendency to generalize the answers for all dwellers of a household in some specific variables. From the results, the hypothesis of "generalization" seems to be correct, indicating a tendency to generalize the responses, which increase with the number of household members. This fact has significant magnitude in the case of rare characteristics that occur in isolation within households. There are variables with higher correlation between household members, such as education and religion. However, other variables may be presented separately within the household, occurring in only one of its members, as some types of disabilities. If there is generalization of responses, these features would not be detected. In this sense, the impact of the questionnaire made in block can also be differentiated by type of question.

Another difficulty imposed by this "new" design of the questionnaire is to conduct individual interviews with each resident, because the questions are answered for all residents at the same time. This can result in increased response given by a third party, namely, the interviewed providing information relating to other residents without the desired precision.

It should be noted that questionnaires with different designs have advantages and disadvantages. These aspects need to be evaluated for each type of research. Those that require greater agility, as censuses, aspects that emphasize agility may be considered more important than for a detailed survey. The ultimate goal would be to achieve a balance between speed and accuracy of data collection, in pursuit of maximum efficiency of the interview, knowing and accepting the specificity of each drawing. The question to be answered is: What would be the ideal model of questionnaire to ensure the maximum quality with minimal effort in terms of duration of the interview?

It is hoped, through these studies, raise discussions about the design of the questionnaires and the implementation structure of innovations and changes in household surveys of IBGE and other statistical institutes.

**References**

Alves, J. E. D.; Barros, L. F. W.; Cavenaghi, S. (2012). **A dinâmica das filiações religiosas no Brasil entre 2000 e 2010: diversificação e processo de mudança de hegemonia**. Revista de Estudos da Religião (REVER), V. 12  nº2, pp. 145-174. http://revistas.pucsp.br/index.php/rever/article/view/14570/10595

CELADE/CEPAL (2011), **"Guía para asegurar a calidad censal"** - Serie Manuales N° 74. Santiago de Chile, Comisión Económica para América Latina y el Caribe, División de Población. Publicación de las Naciones Unidas.

Chackiel, J. (2009), **"Evaluación y estimación de la cobertura en los censos de población: la experiencia latinoamericana"**, documento presentado al seminario "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", del 3 al 5 de junio. Santiago de Chile. http://www.eclac.cl/celade/noticias/paginas/0/37790/JChackiel_doc.pdf

Hess. J., Moore, J., Pascale, J., Rothgeb. J., Keeley A, & U.S. Census Bureau (2001). **The effects of personal-level versus house-level questionnaire design on survey estimates and data quality**. Public Opinion Quarterly, 65, 574-84.

IBGE (2012). **"Censo Demográfico 2010 - Características Gerais da População, religião e pessoas com deficiência"**. Rio de Janeiro, p.1-215. ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Caracteristicas_Gerais_Religia o_Deficiencia/caracteristicas_religiao_deficiencia.pdf

IBGE (2011). **"Censo Demográfico 2010 - Características da População e dos Domicílios - Resultados do Universo"**. Rio de Janeiro, p.1-270. http://www.ibge.gov.br/home/estatistica/populacao/censo2010/caracteristicas_da_pop ulacao/resultados_do_universo.pdf

RIBEIRO, C. A. C.; SILVA, N. do V. **Cor, educação e casamento: tendências da seletividade marital no Brasil, 1960 a 2000**. Dados: Revista de Ciências Sociais, Rio de Janeiro: Instituto Universitário de Pesquisas do Rio de Janeiro - Iuperj, v. 52, n. 1, p. 7-51, 2009. <http://www.scielo.br/pdf/dados/v52n1/v52n1a01.pdf>.

United Nations (2008); **Principles and recommendations for population and housing censuses**. Revisión 2. http://unstats.un.org/unsd/demographic/sources/census/docs/P&R_Rev2.pdf