

## **Scoping study for a Brazilian Rolling Census: options to spread data collection over time**

Alvaro de Moraes Frota  
IBGE, Rio de Janeiro, Brazil – alvaro.frota@ibge.gov.br

Andrea Diniz da Silva  
IBGE, Rio de Janeiro, Brazil – adiniz@ibge.gov.br

### **Abstract**

Over its 80 years of existence IBGE has been improving its methodologies and its work processes. For the population and housing censuses, the institution has to take into consideration the increasing demand for more current information than those produced every ten years, in addition to reduced respondent's cooperation and to the need to reduce costs. Seeking alternatives to help meet both the old and the new challenges, IBGE started a scoping study on new methodologies of census. The scoping study was concluded in 2014 and provided information on the impacts of the application of the rolling census methodology in Brazil. One of the important aspects considered was how to spread the data collection over time. The study took into consideration the aim of obtaining a good balance between quality of the estimates and cost of the operation. Six alternative ways of spreading data collection over time are presented in this paper.

**Keywords:** Census methodology, Brazilian census, rolling census, data collection.

### **1. Introduction**

The mission of the Brazilian Institute of Geography and Statistics – IBGE is "To portray Brazil by providing the information required to the understanding of its reality and the exercise of citizenship". In this context falls the responsibility to conduct the census once every ten years. This is the largest operation undertaken by the institution and provides high quality information about the Brazilian population and households.

Over its 80 years of existence IBGE has been improving its methodologies and its work processes always taking new challenges into consideration. As other statistical offices around the world, IBGE faces the increasing demand for more current information than those produced every ten years, in addition to reduced respondent's cooperation and to the need to reduce costs.

Seeking alternatives to help meet both the old and the new challenges, IBGE started a scoping study on new methodologies of census. The work was initiated in 2004 and aimed to evaluate the technical and operational feasibility of adopting an alternative methodology of census. A preliminary assessment revealed that the rolling census would be an appropriate alternative in the context of the Brazilian statistical system. This initiative was called Estudo de Modalidades Alternativas de Censo Demográfico and later had the name changed to Estudo da Modalidade de Censo Demográfico Contínuo - EMCDC.

Activities conducted under the EMCDC programme include international and national seminars, multi-disciplinary working groups, meetings with IBGE's experts, a six-year pilot test and a simulation of the rolling census methodology and fed a scoping study for the Brazilian rolling census.

The scoping study was concluded in 2014 and provided information to assess the impacts of the application of the methodology of a rolling census in Brazil.

One of the important aspects considered was how to spread the data collection over time. The study took into consideration the aim of obtaining a good balance between quality of the estimates and the cost of the operation. Six alternative ways of spreading data collection over time are presented in this paper.

## **2. Operating model of the Pilot Test**

To conduct the 2010 Census, the 5,565 Brazilian municipalities were represented in a cartography base of 310,114 enumeration areas. This base covered without overlaps or failures nationwide, adapting to the political-administrative divisions of the municipalities. During the operation, as in previous censuses, each enumeration area represented the workload of one enumerator at a single moment in the decade.

For the Rolling Census Pilot Test the same cartography was used, over a five-year accumulation cycle, and a fifth of the enumeration areas in each municipality was associated to each rotation group or year of collection. After five years all municipalities had been completely covered and their data were accumulated and used to make estimates at the municipal level. Such procedure was applied to the four municipalities included in the Pilot Test. The municipalities were visited every year, but only once a year. The same reference date as the previous census (evening of July 31 to August 1) was used and the accumulation was made by joining five sets of information collected along five consecutive years.

The operating model adopted in the Pilot Test is one among many possibilities, so it is importante to study how the data collection will be distributed over time within each accumulation cycle.

## **3. Alternatives for spreading data collection**

Three possibilities regarding the frequency of data collection within a cycle of accumulation of five years were considered: only once in the five-year cycle, five times (annual) and 60 times (monthly). In the first case one fifth of the Brazilian municipalities would be visited once a year<sup>1</sup> and on this occasion all enumeration areas would be covered. In the second case every municipality would be visited once a year, but only one fifth of the enumeration areas would be covered each year. In the third case all municipalities are visited 12 times a year and 1/60 of enumeration areas are covered every month. In all three cases, one fifth of the enumeration areas, ie 62,023, would be covered each year and estimates for the municipal level would be made by the accumulation of data collected over the most recent five years.

To guide the present study, the following simplification were made: 1) the more spread in time is the data collection, the less complex is the operation and the more current is the information, so it is the best quality we can get; 2) the more concentrated in time is the data collection, less returns to the municipality are necessary so the cost of operation is lower. Based on these assumptions, the challenge is to find a balance between cost and quality.

The reasoning is that a model based entirely on the five-year collection does not include returns to municipalities within a cycle of accumulation thus presents a lower operating cost than other models. However, this model does not incorporate as updated data on the population as a model based on monthly collection, and present the same degree of complexity which the decennial operation. On the other hand, the full scattering collection over the 60 months of accumulation cycle in all municipalities reduces the size and complexity of the operation and allows for incorporation more current information, but does not represent a better overall quality assurance. An important consideration of this issue is that the greater the amount of returns to remote and difficult to access areas the greater is the risk of problems and loss of data quality. Furthermore, such a model is of a high operating cost and may not be feasible.

The first exercise was to think that spreading the data collection over time can have less impact on the cost when it is adopted in the municipalities where IBGE's offices are running and count on a well established team and the distances to be covered are shorter. In this case, the first scenario is to spread the data collection in such municipalities out over 60 months and make it only once every five years in the other municipalities.

---

<sup>1</sup> A total of 1,113 municipalities, considering the cartography base used in 2010.

### Scenario 1: Five-year and monthly data collection

In this scenario, about 43% of the population is listed under the five-year model and the remaining under the monthly model. A total of 5,169 enumeration areas are visited every month, 2,442 in municipalities under the five-year model for which all enumeration areas are covered on this occasion and 2,727 in municipalities under the monthly model so that a 1/60 of its enumeration areas are covered per month. Table 1 shows the average number of enumeration areas to be surveyed every month under Scenario 1.

**Table 1 - Average enumeration areas to be surveyed every month under monthly and five-years models, according to the Federation Units - Scenario 1**

Federation Units	Data collection operating model		
	Both	Five-year model	Monthly
<b>Brasil</b>	5169	2442	2727
São Paulo	1102	354	748
Minas Gerais	543	275	268
Rio de Janeiro	463	132	330
Bahia	396	221	176
Rio Grande do Sul	372	213	159
Paraná	291	141	150
Ceará	221	136	85
Pernambuco	206	121	85
Santa Catarina	198	122	76
Goiás	157	84	73
Maranhão	147	98	49
Pará	146	90	56
Espírito Santo	106	43	63
Mato Grosso	99	53	46
Amazonas	94	37	57
Paraíba	92	60	33
Piauí	88	56	32
Distrito Federal	72	0	72
Rio Grande do Norte	71	44	27
Mato Grosso do Sul	70	31	39
Alagoas	62	33	29
Sergipe	55	34	21
Rondônia	39	22	17
Tocantins	35	22	13
Acre	15	6	8
Roraima	14	7	7
Amapá	14	6	8

Source: IBGE, Base Operacional Geográfica (planejamento).

There may not be a unique arrangement for the 27 federation units, and then alternatively five additional scenarios were designed. A combination of the three possibilities regarding the frequency of data collection within a cycle of accumulation of five years: only once in the five-year cycle, five times (annual) and 60 times (monthly) was taken into consideration. In these scenarios the frequency of collection in a municipality is related to the number of enumeration areas in the municipality in order to give an idea of the workload. In this case the annual basic model would only apply to municipalities with at least five enumeration areas and monthly to

municipalities with at least 60. However, whatever the scenario, the number of enumeration areas to be covered per year is 62.028, which correspond to a monthly average of 5,169 enumeration areas. Figure 1 summarizes the scenarios considered.

**Figure 1 - Data Collection Scenarios**

Scenarios	Data collection operating model		
	Once in five years (in all the municipalitie)	Annual (in 1/5 of the enumeration areas)	Mensal (in 1/60 of the enumeration areas)
Scenario 1	Municipalities without IBGE's office	None	Municipalities with IBGE's office
Scenario 2	1 to 4 enumeration areas	5 to 59 enumeration areas	60 enumeration area or more
Scenario 3	1 to 49 enumeration areas	50 to 499 enumeration areas	500 enumeration areas or more
Scenario 4	1 to 4 enumeration areas	5 enumeration areas or more	None
Scenario 5	1 to 49 enumeration areas	50 enumeration areas or more	None
Scenario 6	1 to 99 enumeration areas	100 enumeration areas or more	None

Source: IBGE, Estudo da Modalidade de Censo Demográfico Contínuo.

### **Scenarios 2 and 3: five-years, annual and monthly surveys**

In Scenario 2, the Brazilian municipalities are divided into three partitions. The first one contains the municipalities that have one to four enumeration areas, in which the survey will be conduct every five years and will encompass the entire city at once. The second contains the municipalities with five to 59 enumeration areas, with annual collection of 1/5 of the enumeration areas of the municipality. The third partition, finally, adds the municipalities with 60 or more enumeration areas, where the collection will be monthly, resulting, in the end of a year, 1/5 of its sectors collected

In scenario 3 the municipalities are also divided into three partitions. The first partition with municipalities that have one to 49 enumeration areas; the second aggregating municipalities with 50 to 499 enumeration areas, with annual collection in 1/5 of the sectors; and the third adding municipalities with 500 or more enumeration areas, with monthly collection in 1/60 of such areas.

In these two scenarios, 2 and 3, to distribute the labour force over time, visits to municipalities with quinquenal collection are staggered over 60 months and visits aimed to collect 1/5 of the enumeration areas in municipalities with annual collection are staggered over 12 months. In other municipalities, the collection is, by definition, monthly.

Scenario 2 shows the following for the 5,565 Brazilian municipalities: 903 municipalities with monthly collection, 4,539 municipalities with annual collection, and 123 municipalities where the collection will take place once every five years. Thus, monthly enumerators will visit 5,169 enumeration areas, with 8 in municipalities with five-year collection, 1,648 in municipalities with annual collection and 3513 in municipalities with monthly collection. Taking as reference the results of the 2010 Census, this design implies having about 72% of the population being enumerated in the monthly model, 28% in the annual model and less than 1% of the population under the five-year model.

In Scenario 3, the municipalities are divided as follows: 64 with monthly collection, 1,083 with annual gathering and 4,418 under the five-year collection. Thus, the visited enumerating areas are distributed in 1635 with monthly collection, 2,099 from municipalities with annual

collection and 1,435 from municipalities with five-year collection. In this scenario 35% of the population is enumerated in the monthly model, 41% in the annual model and 24% of the population under the five-year model.

Figure 2 summarizes the scenarios considered.

**Figure 2 - Total of Municipalities, Monthly Average of Surveyed Municipalities, Total of Enumeration Areas and Monthly Average of Surveyed Enumeration Areas, according the Data Collection Operating Models and Scenarios**

Scenarios	Totals / Averages	Data collection operating model			
		All	Monthly	Annual	Once in five years
Scenario 2	Total of Municipalities	5.565	903	4.539	123
	Municipalities per month	93	15	76	2
	Total of Enumeration Area	310.114	210.762	98.897	455
	Enumeration Area per month	5.169	3.513	1.648	8
Scenario 3	Total of Municipalities	5.565	64	1.083	4.418
	Municipalities per month	93	1	18	74
	Total of Enumeration Area	310.114	98.072	125.919	86.123
	Enumeration Area per month	5.169	1.635	2.099	1.435

Source: IBGE, Base Operacional Geográfica (planejamento).

In both scenarios the collection is distributed over the 60 months of the accumulation cycle, independently of the distribution within each municipality, so both the total number of municipalities as enumeration areas to be visited per month is the same in both scenarios.

The adoption of one of the two scenarios, 2 or 3, requires to conduct a field work every month, throughout the year, so that the reference dates need to be properly equated to the right information to the respondent. In this case, the reference date of the survey may be the last day of the previous month, as the reference date of employment and income questions may be the previous month preceding and the last full week of such month. The reference date of migration should be retroactive five years from the reference month of the survey. In this case there would not be a single reference date for the annual presentation of cumulative results, which could be referenced only in the year or June 30 as an attempt to mark the middle of the year, for example.

#### **Scenarios 4, 5 and 6: five-years and annual surveys**

In scenarios 4 to 6 were considered only two frequencies of return to the municipality: annual and five-year. In Scenario 4, the municipalities with up to four enumeration areas are visited once every five years and the other annually. In scenario 5 the five-year collection is held in municipalities with up to 49 enumeration areas. In Scenario 6 the five year collection is held in those with up to 99 sectors.

After the computation, Scenario 4 has the following configuration for the 5,565 Brazilian municipalities: 5,442 of them with annual collection and 123 with five-year collection. Thus, the 62,023 enumeration areas to be visited each year are distributed as follows: 61,932 from municipalities with annual collection and 91 municipalities where collecting is every five years. Scenario 5 provides annual coverage of 44,798 enumeration areas in 229 municipalities with annual collection and 17,225 in 884 municipalities with five-year collection. Scenario 6 provides annual coverage of 35,767 enumeration areas in 96 municipalities with annual collection and 26,255 in 1,017 municipalities with five-year collection.

Figure 3 summarizes these three scenarios considered.

**Figure 3 - Total of Municipalities, Annual Average of Surveyed Municipalities, Total of Enumeration Areas and Annual Average of Surveyed Enumeration Areas, according to the Data Collection Operating Models and Scenarios**

Scenarios	Totals / Averages	Data collection operating model		
		All	Annual	Once in five years
Scenario 4	Total of Municipalities	5.565	5.442	123
	Municipalities per year	1.113	1.088	25
	Total of Enumeration Area	310.114	309.659	455
	Enumeration Area per year	62.023	61.932	91
Scenario 5	Total of Municipalities	5.565	1.147	4.418
	Municipalities per year	1.113	229	884
	Total of Enumeration Area	310.114	223.991	86.123
	Enumeration Area per year	62.023	44.798	17.225
Scenario 6	Total of Municipalities	5.565	480	5.085
	Municipalities per year	1.113	96	1.017
	Total of Enumeration Area	310.114	178.837	131.277
	Enumeration Area per year	62.022	35.767	26.255

Source: IBGE, Base Operacional Geográfica (planejamento).

Referring to the 2010 Census, the design of Scenario 4 involves having more than 99% of the population surveyed in the annual model collection and less than 1% of the population under the five-year model. In scenarios 5 and 6, 76% and 62% of the population, respectively, were under annual collection.

#### **4. Conclusion**

The way data collection is distributed over the accumulation cycle directly affects the cost of the operation and the quality of the estimates, although these are not the only factors of influence. The scoping study offers an overview on possible designs; however the choice of the one depends on further discussion by IBGE.