Development of an Integrated System of Household Surveys: The Brazilian Experience

Quintslr, Marcia Maria Melo
IBGE, Diretoria de Pesquisas – Coordenação de Trabalho e Rendimento
Av. República do Chile 500, 6º andar
Rio de Janeiro (20.031-170), Brazil
E-mail: marciaq@ibge.gov.br

Hypólito, Elizabeth Belo
IBGE, Diretoria de Pesquisas - Coordenação de Trabalho e Rendimento
Av. República do Chile 500, 6º andar
Rio de Janeiro (20.031-170), Brazil
E-mail: elizabeth.hypolito@ibge.gov.br

The increasing demand for social and economic data and the recurring calls for optimal use of public resources have prompted the Brazilian Institute of Geography and Statistics (IBGE) to redesign its set of household surveys. This paper tackles the most important characteristics of this project that was established to integrate the planning, the methodological aspects (including harmonization of concepts and survey questions) and the operational stages of the surveys and also to expand the potential for crossed analysis. The project, named Integrated System of Household Surveys, is being developed focusing on three fundamental targets: the design of Master Sample, the production of short term labor market indicators for all of Brazilian states and the ongoing estimation of household consumption based on a continuous scheme of survey. The Integrated System of Household Surveys aims to meet user's requirements for a greater variety of outputs from different and variable subjects by means of independent surveys or supplementary module to a survey. These supplementary modules will be carried out in conjunction with two core surveys that will be the labor force survey and the expenditure and income survey. In addition, the paper presents some aspects related to the conduction strategies of the project and the next steps in sight.

Introduction

The increasing demand for socioeconomic and demographic data, which mainly come from various sectors of the government and researchers, has been a reality for the official institutes of statistics all over the world, what comprises a huge challenge as the resources become more limited.

The ongoing model of production of household surveys at Brazilian Institute of Geography and Statistics (IBGE) enables the unique and independent planning of each survey, taking into account totally distinct sampling infrastructure. The differences are also present in some concepts adopted for the observation of the same topic. Two surveys whose main topic comprises labor force and income are being carried out nowadays. One of the them takes place on an annual and national basis, whereas the other one is a monthly survey that covers some metropolis. Nearly every five years an expenditure and income survey is carried out and an urban informal sectors survey has an irregular periodicity.

The annual labor force survey, called National Household Sample Survey (PNAD[1]), investigates other regular topics: housing, education, fecundity, migration and some aspects of child labor. Moreover, it is by means of this survey that IBGE responds to the demand of other subjects, such as health, food security, information and communication technology, among others, what sometimes generates very long questionnaires. It is advisable to distribute the load of the interview among the surveys in a more balanced

---

[1] Called, in portuguese, Pesquisa Nacional por Amostra de Domicílios.

way.

IBGE has been reformulating its work process, which aims to implement, in the next decade, the Integrated System of Household Surveys (ISHS). Differing from what happens now, in this new model, the planning, accomplishment, analysis of the results and dissemination of the main IBGE household surveys will be carried out in an associated way, optimizing the required resources and enabling important advances in the production of data about households, families and people.

The ISHS implementation will adopt a shared selection record (Master Frame) and a shared sample design (Master Sample), the harmonization of concepts, variables and classifications in the various investigations comprising the system, the standardization of editing and imputation process, besides the amplification of the use of data collection technologies.

In general, the expansion of the offered range of data will ideally consider the inclusion of new topics and, regarding those already evaluated, it will deepen the investigation, shorten or fix the periodicity of data collection for some topics. The dissemination of the data in more agile deadlines, the production of results in a more detailed geographic division and the inclusion of longitudinal investigations complete the improvement required to support the evaluations of public policies or their designs. This new data basis must also regard comparative studies among Brazilian regions and the comparison among countries.

In the context of ISHS construction, the Continuous National Household Sample Survey (Continuous PNAD) will enable the production of short-term indicators of labor force and income on a national basis and detailed by state. The execution of the family budget surveys will consist of the Expenditure and Income Survey (POF[2]) and the Simplified Expenditure and Income Survey (Simplified POF). This POFs scheme will enable the production of continuous data on consumption, what mainly aims to carry out studies about life conditions, national accounts and inflation, as well as poverty.

Since they serve multiple purposes, the Continuous PNAD and POFs scheme will naturally be important ways to conduct the investigation of supplementary topics. However, some demands may lead to the execution of surveys which are independent from the nuclear ISHS surveys, then flexibility must be taken into account in the construction of the system.

Methodologically speaking, the Master Sample is fundamental for the development of this project as it ensures a unique sample infrastructure for all the surveys in order to allow the systemic organization of the various data sources.

These three basic elements of the Integrated System of Household Surveys, Continuous PNAD, POFs scheme and Master Sample are discussed in this paper. Some aspects related to the conduction strategies of the project and the next steps in sight are also presented.

Master Sample

The Master Sample is defined as a set of enumeration areas, that is, census sectors, selected from the Master Frame with proportional probability to a size measurement. This frame consists of the sectors covering all the national territory and contains data about administrative divisions, population counts and other sociodemographic characteristics obtained from the 2000 Demographic Census[3].

From this Master Sample of enumeration areas, it is possible to build subsamples of sectors and households for the several surveys inserted in the ISHS. For instance, the POFs will use subsamples from approximately 40% of the Master Sample sectors, whereas the Continuous PNAD will use 100% of the sectors of this sample but it will only select some households in each of them. Thus, differently from what happens nowadays, in the ISHS, the sectors will be shared among the surveys.

The construction of this unique sample structure for all the household surveys has many advantages, such as sharing cost of listings and the execution of maps, a better knowledge of the selected areas, the

---

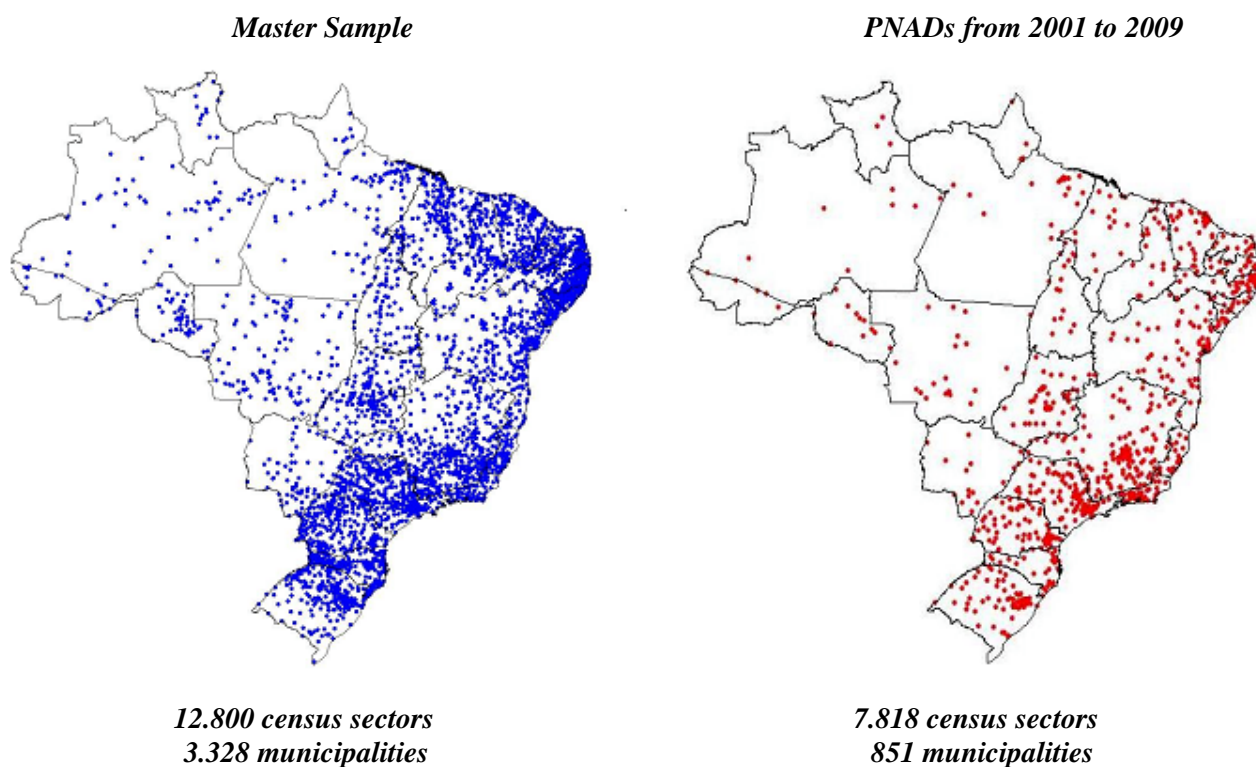[2] Called, in Portuguese, Pesquisa de Orçamentos Familiares.
[3] A new version of Master Sample will be prepared after 2010 Demographic Census.

feasibility to carry out richer analyses of results, besides the fact that new topics may be investigated in an easier way.

The Master Sample will cover all the Brazilian territory. The sample design is stratified in several levels. First of all, the Country is stratified according to its political and administrative divisions. Next, the formed strata are substratified according to the contiguity and homogeneity of the enumeration areas. Then, the census sectors are segregated according to the urban or rural situation. Lastly, income strata are created in order to minimize the variance of the estimator of the total income received by the reference person of the household.

This design allows a better spread of the sample in the national territory if compared to the design adopted nowadays by PNAD, as seen in Figure 1. Among the factors that lead to this difference, we may highlight the stratification of the Master Sample by urban and rural situations and by income, as these stages do not exist in PNAD. Moreover, unlike the Master Sample, in this survey, the sectors are used as the primary sampling unit only in part of the strata.

Figure 1: Municipalities that contain sectors selected for the Master Sample and for PNADs from 2001 - 2009

Master Sample	*PNADs from 2001 to 2009*

12.800 census sectors	*7.818 census sectors*
3.328 municipalities	*851 municipalities*

It is worth pointing out that since 2003, POF has adopted a design which is similar to the Master Sample one. Therefore, there have been no great impacts when selecting a subsample of the Master Sample for this survey.

In order to minimize the end or the total change of sectors from the census years on, a sector rotation scheme will also be used. The idea is to substitute 2.5% of them every quarter. Thus, by the end of 10 years, the sector Master Sample will be all renewed.

ISHS Theme Structure

Regarding the theme structure, ISHS intends to maintain all the current topics that are investigated by IBGE. Consequently, besides the Continuous PNAD and the POFs scheme, the Urban Informal Sectors

Survey will be part of ISHS. The system also aims to include new topics, either as supplementary modules to an existent survey or as independent surveys. Some topics, such as health and victimization, are likely to become independent surveys in the ISHS. Food security, however, will probably comprise a supplementary investigation.

As seen before, due to its annual periodicity, frequency and sample size, nowadays, PNAD is used as a means of investigation of various topics, being them regular or supplementary. According to the amount and urgency of the demands, this survey may be overloaded.

In the new model of survey production, the Continuous PNAD and the POFs scheme will be continuous and serve multiple purposes. Thus, additional topics may be distributed among these surveys what can prevent one of them from being too large.

One of the most important matters to be defined is the thematic content of ISHS in order to attend to the Brazilian needs. Some topics, for example, housing, migration and child labor are investigated by PNAD every year and it is intended to maintain this periodicity for them in the context of the ISHS. Topics such as education of youngsters and adults, food security, information and communication technology, use of time, among others, are strongly demanded to IBGE and must be investigated in the ISHS.

In principle, the link between the additional and the central topics of the survey should be the main criterion to be observed in the survey. The Continuous PNAD will have, as its central topic, labor force and income. Therefore, child labor would be related to this survey, whereas food security would be better placed in the POFs scheme, whose central topic is consumption and income.

Other criteria must also be taken into account when placing topics in these surveys. Among them, we may point out the sample size of each of them, the investigation methodology, the geographic level intended for the results and the periodicity of dissemination.

Figure 2 shows how complex it is to integrate the surveys in only one system. In this context, there is a set of variables such as gender, color or race, age, composition of the household and level of education that comprise the ISHS basic core. These variables will be investigated in all IBGE household surveys in a harmonized way. The composition of the topics and variables in this core are being studied nowadays. Here is presented only one example, in accordance with the definitions of Eurostat, in its Final Report from the Task Force Core Social Variables, referring to the integration and harmonization of the concepts carried out in the several surveys of European countries.

Figure 2: Integrated System of Household Surveys

Survey		Central topic of SIPD	Central topic of the survey	Additional topics
Continuous PNAD		Gender Age	Labour force and Income	?
POFs scheme	POF	Ethinicity (colour, race)	Consumption and Income	?
	Simplified POF			
ECINF		Household composition	Informal sector of economy	?
New survey		Highest level of education	?	?

The Continuous National Household Sample Survey - Continuous PNAD

IBGE, nowadays, carries out two labor force surveys: a national survey, which has been carried out since 1967, providing annual data and the other one, a monthly survey that covers the main metropolises. The current project of revision of the household surveys aims to substitute the two surveys for the quarterly publication of data regarding the economically active population, employment, unemployment, underemployment and income – the Continuos PNAD.

As the current annual survey, the Continuous PNAD, so as to meet the demands, will investigate some topics with a fixed periodicity and other topics will be regarded by means of supplementary modules.

Before going on, it is important to point out the challenge caused by the reformulation of the labor force survey by IBGE, presenting some general characteristics of the current surveys.

The annual survey, called the National Household Sample Survey – PNAD enables data for Brazil, the major regions and states. Every final quarter of the year, 140,000 households are visited and the data are disclosed and published in September of the following year. This survey has been carried out for 40 years. In 2007, the data started to be collected by means of electronic hand held equipment. The most important element to be improved is the availability of the short-term information about labor force on a national basis.

The Monthly Employment Survey collects data in six Metropolitan Regions, where about 30% of the Brazilian urban population reside. It deals with labor force data and employment-related income and, every month, 40,000 households are visited. This survey was fully revised in terms of concept in 2002 and it has used electronic hand held equipment for data collection since then. The main element to be improved is the geographic range, as the survey generates short-term information that does not completely portray the different contexts of the Brazilian labor market.

Box 1: Continuous PNAD – Principles Characteristics

General Objectives of Continuous PNAD
Produce basic information for the study of the socioeconomic development of the country and enable the continuous investigation of the labor and income indicators.
Geographical level of results
Brazil, Major Regions and States, Metropolitan Regions that include the capital cities of the States.
Sample
Brazil will adopt the master sample infrastructure and, in order to carry out the Continuous PNAD, about 170,000 households will be visited every quarter.
Periodicity
Quarterly periodicity for the Labor Force and Income indicators and annual periodicity for some sociodemographic topics, such as child labor and migration. Furthermore, some supplementary modules will be investigated with a variable periodicity, in accordance with the need of results.
Conceptual Framework
The concepts adopted in both current surveys are being revised to make them completely updated regarding the international recommendations. Adaptations may be necessary due to national context.

In the Continuous PNAD, the quarterly estimates referring to labor force and income will be obtained from the collection of data gathered in three consecutive months. Each selected household will be

interviewed five times, during five quarters. Thus, the household will answer the questionnaire for the second time three months after the first interview, for the third time three months after the second interview and so on. In the adopted rotation scheme there will be no monthly household overlap, that is, in two consecutive months, the interviewed households will be completely different. In two consecutive quarters, the household overlap will be 80% and, in two consecutive years, this number will reach 20%.

This rotation pattern was chosen because it has one of the lowest variances in the estimate of the quarterly differences among the patterns previously analyzed by IBGE. The relevance of this fact is associated with the periodicity of dissemination proposed for this survey – every quarter. Besides, users of IBGE data pointed out the importance of an annual household overlap, even if it is small. However, in order to use the overlap in studies, it should be controlled. IBGE has acquired this experience from the monthly labor force survey and should apply it in Continuous PNAD.

The topics labor force and income are the central issue of the survey and will be investigated every quarter in 100% of the quarterly sample. The remaining topics of this survey are not supposed to be investigated within such short periodicity as a quarter and therefore, can be distributed along the five interviews, what prevents the informant from being tired. For example, besides the basic core, households for the first interview will answer questions about a specific topic, households for the second interview will answer another specific topic and so on. These specifics topics will be investigated in 20% of the sample per quarter. In accordance with the topics characteristics or the required precision, they will be collected for a certain period of time (an year, for example) and then disseminated. This is the alternative that shall be adopted concerning child labor and fecundity, for instance. Another possibility is to investigate a certain topic in an established quarter and then repeat this investigation every year or within a desired periodicity. This is the case concerning another regular topic nowadays, internal migration. The structure of distribution of topics is shown in Figure 3, which presents possibilities for the investigation of fixed and supplementary module topics.
The quarterly periodicity proposed to the labor force and income statistics in the Continuous PNAD has been questioned by part of the users, who have highlighted the necessity of monthly information. In order to comply with this demand, IBGE has been looking for alternatives in the international experience, such as the rolling quarterly estimates approach or the monthly estimate for aggregated geographical levels.
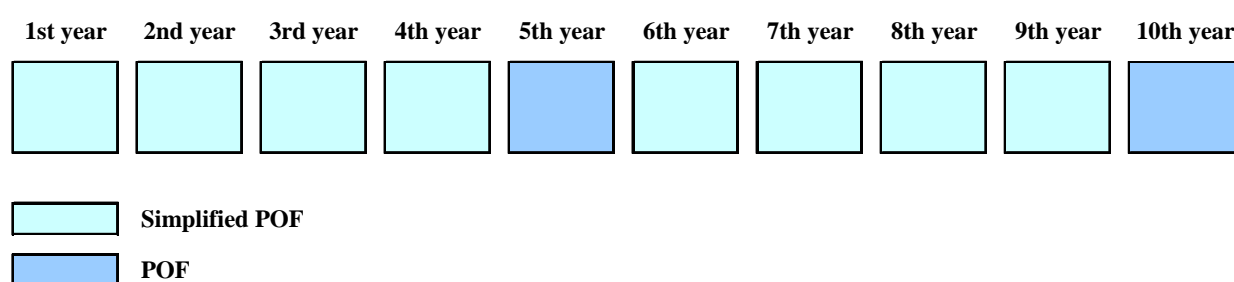
Figure 3: Continuous PNAD - The structure of distribution of topics

Survey	Interview	Central topic		Additional topics				
					by quarter of year			
		SIPD	Survey	by interview	1st quarter	2nd quarter	3rd quarter	2nd quarter
Continuous PNAD	1st interview	Gender	Labour force and Income	Topic 1I		Topic 2Q	Topic 3Q	Topic 4Q
	2nd interview	Age		Topic 2I				
	3rd interview	Ethinicity (colour, race)		Topic 3I				
	4th interview	Household composition		Topic 4I				
	5th interview	Highest level of education		Topic 5I				

The scheme of the continuous POF

The creation of a POFs scheme will offer the evolution of the population welfare taking into account consumption and budget, what will contribute to the establishment of public policies in the social area related to the eradication of poverty, actions concerning nutrition, food orientation, health, housing, among others. It will help follow the behavior of regional differences, urban and rural situations on a regular and systematic basis, in order to optimize the financial and human resources required in the various public. This scheme expects that the complete version of POF will be carried out every five years, according the 2002-2003 POF patterns, which is now being held and that covered all the Brazilian territory, whose methodology prioritizes the collection of both economic and sociodemographic data. These general characteristics will be maintained in the Simplified POF, which will be carried out every year, in order to obtain continuous estimates related to consumption items, thus meeting several applications.

Figure 4: The POFs scheme

The definition of the supplementary modules composing POFs scheme shall take into account the central topic – the domestic budgets, besides the characteristics of the field operations regarding these surveys. The interviewer visits the household more than once in a defined period of time so as to collect the data from each selected household, addressing, in general, to more than one resident.

Regarding the Simplified POF questionnaire, two facts are important: a) it must be a reference to the study of poverty eradication and life conditions and b) it must enable the comparison with the quinquennial POF. The needs of the National Accounts and of the National System of Price Indices will also be taken into consideration.

The definition of the level of details of the investigation of the expenditure items is crucial for the design of the simplified investigation model, which should be compatible with the Classification of International Consumption by Purpose (COICOP), established by the United Nations Organization, as IBGE will adopt COICOP to express all the produced consumption data, i.e. the ones that were published based on expenditure and income surveys, price indices and national accounts.

The definitions about variables concerning income, households and persons will take into account, besides the aspects referring the use of these data, the concept harmonization and the complementarity referring to the investigation of those topics in the quinquennial POF, in the Continuous PNAD, and in all ISHS.

Alternative models of Simplified POF questionnaire were tested, simultaneously with the complete POF in order to compare results, what will be one of the key elements to evaluate the proposed methodologies.

Project strategies and its next steps

The construction of the Integrated System of Household Surveys is to be implemented in 2011, when, using part of the statistical structure installed for the 2010 Census, IBGE intends to start to carry out the

Continuous PNAD. During two years, this survey will happen simultaneously with the two present surveys about labor force.

Moreover, it is important to point out that the surveys now being carried out have been working as a real laboratory for the implementation of the Continuous PNAD. They are going through improvements that will be transferred to the new system, such as the use of computers to collect data in the annual survey, the use of specific softwares for editing and imputation of data and distance training.

The Continuous PNAD questionnaire was tested in November 2008. It has later been updated and is to be tested again in the second semester of 2009, comprising 50,000 households.

The alternative models of the Simplified POF questionnaire were tested in approximately 4300 households, from February to June 2009. Currently, comparative analyses of the results obtained from POF 2008-2009 and from these models are being developed.

The users of IBGE data have been often listened to and informed about the evolution of the project. The technicians have taken part in several events and the IBGE website contains frequently updated information about the advances made. Furthermore, meetings with the users are being held every semester.

Lastly, it is worth mentioning that IBGE leads an interministry committee to discuss the social statistics of the country, which aims to create an integrated vision of the various data bases: Censuses, sample surveys, administrative records. The proposition of topics for the Integrated System of Household Surveys is a highlight in the tasks of this committee.

REFERENCES

[1] EUROSTAT. **Final Report from the Task Force on Core Social Variables.** Meeting of the European Directors of Social Statistics. Luxembourg, 2006.

[2] FREITAS, M. P. S. de, et. al. **Amostra Mestra para o Sistema Integrado de Pesquisas Domiciliares.** Texto para discussão número 23. Diretoria de Pesquisas, IBGE. Rio de Janeiro, 2007.

[3] IBGE. **Pesquisa Mensal de Emprego – PME**, Série Relatórios Metodológicos, vol. 23, Coordenação de Trabalho e Rendimento, Diretoria de Pesquisas. Rio de Janeiro, 2007.

[4] ——. **Pesquisa de Orçamentos Familiares – POF 2002-2003**, **Primeiros Resultados: Brasil e Grandes Regiões**. Coordenação de Trabalho e Rendimento, Diretoria de Pesquisas, IBGE. Rio de Janeiro, 2004.

[5] ——. **Manual de Entrevista da PNAD**. Coordenação de Trabalho e Rendimento, Diretoria de Pesquisas, IBGE. Rio de Janeiro, 2006.

[6] INTERNATIONAL LABOR ORGANIZATION. **Resolution concerning statistics of the economically active population, enployment, unemployment and underemployment**. Thirteenth International Conference of Labor Statisticians. 1982.

[7] QUINTSLR, M. M. M, et. al. **O sistema Integrado de Pesquisas Domiciliares.** Texto para discussão número 24. Diretoria de Pesquisas, IBGE. Rio de Janeiro, 2007.