Census Model Transition: Contributions to its Implementation in Portugal

Carlos A. Dias[1], Anders Wallgren[2], Britt Wallgren[3], and Pedro S. Coelho[4]

Given the high cost and complexity of traditional censuses, some countries have started to change the census process. Following this trend, Portugal is also evaluating a new census model as an alternative to an exhaustive collection of all statistical units. The main motivations for the implementation of this census model transition in Portugal are related to the decrease in statistical burden on citizens, improvements in the frequency of outputs, and the reduction of collection costs associated with census operations. This article seeks to systematise and critically review all alternatives to the traditional census methodologies, presenting their advantages and disadvantages and the countries that use them. As a result of the comparison, we conclude that the methods that best meet these objectives are those that use administrative data, either in whole or in part. We also present and discuss the results of an inventory and evaluation of administrative registers in Portugal with the potential to produce statistical census information.

Key words: Administrative registers; register-based census; traditional census.

1. Introduction

Population and housing censuses are statistical operations performed across the world to collect all data on the statistical units – living quarters (dwellings), households, and persons – within a national universe. Traditionally, census operations are decennial and require significant human, financial, and material resources. In addition to these high costs, a considerable effort is required from citizens, who are "forced" to respond to questions whose answers, in many cases, may already exist in several databases within the Public Administration (Scheuren 1999). Given these constraints, some countries have started to change the census process. This entails collecting data not from the traditional model involving an exhaustive survey of all statistical units, but from administrative sources (Redfern 1986).

This article aims to present a systematic critical review of alternative methodologies to the traditional census, showing their advantages, disadvantages and the countries in which they are used. The methods using administrative data are highlighted and an analysis of

[1] Nova IMS-Information Management School and Statistics Portugal, Av. António José de Almeida, 5, 1000-043 Lisboa, Portugal. Email: carlos.dias@ine.pt
[2,3] BA Statistiksystem AB, Norrhagev.30 SE-71993 Vintrosa, Sweden. Email: ba.statistik@telia.com
[4] Nova IMS-Information Management School, Campus de Campolide, 1070-312 Lisboa, Portugal. Email: psc@novaims.unl.pt

their situation in Portugal is included. The objective is to lay the foundations for the use of administrative data, in whole or in part, in the country's 2021 census round.

This article is organised in seven sections. Section 2 briefly presents the legal framework at national and international level. Section 3 identifies the methods used to obtain census information and explores a critical review of alternative methodologies to the traditional census. Section 4 discusses the census model transition in Portugal. Section 5 analyses the administrative registers in Portugal of potential use for statistical purposes in the census context. Section 6 discusses possible models for the 2021 Census in Portugal. The main conclusions of the article are presented in Section 7.

2. Legal Framework

In order to ensure the harmonisation and comparability of results, the UNECE (United Nations Economic Commission for Europe) recommendations set out the basic rules to be followed in population and housing censuses. They also establish the concepts and definitions associated with the statistical units and variables to be observed (UNECE 2006). Based on these recommendations, for the 2010 census round (covering the period 2005–2014), the EU (European Union) adopted four regulations proposed by Eurostat (Statistical Office of the European Union) after discussion with the representatives of all member states. The regulations introduce a mandatory set of rules on the content to be observed, the geographical breakdown for each variable and the quality indicators that each member state should provide to Eurostat (Eurostat 2011). The existence of this community legislation, as an instrument for regulating the EU censuses, guarantees the availability and harmonisation of census information.

In Portugal, census operations are also supported by specific national legislation. A specific law established the organisational arrangements and executive for the 2011 Census, the last traditional census operation held in Portugal. A feature of this legislation is the explicit reference to the implementation of the census transition process, supported by data from administrative sources. To this end, it includes the possibility for Statistics Portugal, which was responsible for implementing the 2011 Census, to create databases for individualised registers of living quarters/buildings, housing units/dwellings, households, and persons. This legal framework at national level also underlies the Portuguese Statistics Act, which regulates the National Statistical System (SEN) and establishes, for the first time, the principle of the use of administrative data for official statistical purposes.

3. Alternative Methodologies

According to the international recommendations for obtaining census statistics, various approaches can coexist in data collection, covering a wide spectrum from the exhaustive collection of all statistical units (traditional method) to models based solely on administrative information (register-based censuses). The mix of the two methods, supplemented in some cases by sample surveys, allows several combinations.

Numerous studies have presented different method classifications (Valente 2010b; UNECE 2013). In this article, we propose four main groups of methods that build on the classification in Valente (2010b): traditional census, register-based census, combined methods, and rolling census (Figure 1), which are detailed in the following subsections.
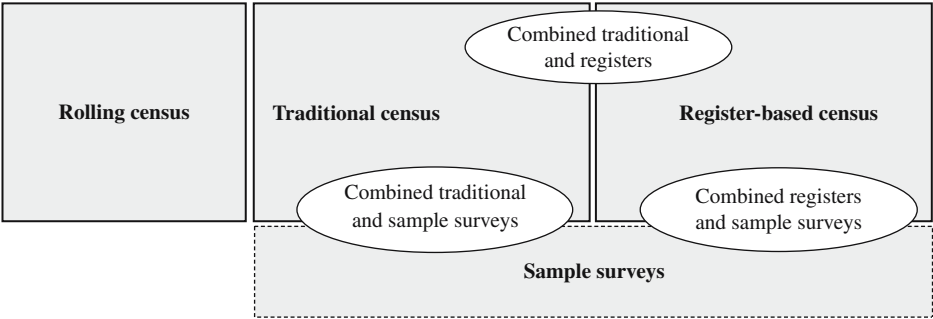
Fig. 1. *Methods of obtaining census information for statistical purposes*

3.1. *Traditional Census*

The traditional census approach collects basic characteristics from all individuals and housing units (full field enumeration) at a specific point in time. In most countries, including Portugal, this is the most common approach to census taking. In the 2011 Census, the census could be taken through a web questionnaire in 13 UNECE countries (Valente 2010b). It is noteworthy that in Portugal approximately 50.5% of the population responded via the Internet.

138 countries responded to the survey conducted worldwide by the UNSD (United Nations Statistics Division) and the UNECE in 2009 concerning the 2010 census round, and of these 83% planned to use the traditional method (UNSD 2010). Of all the UNECE member countries, about 56% of the 50 responses indicated the traditional method as the method to be adopted, while in the 27 EU member states, only 41% indicated this method (Valente 2010a). These results indicate that the most highly industrialised and developed countries are more likely to abandon the traditional method and adopt new alternative census designs.

One of the main disadvantages of the traditional method is the high cost and complexity of census operations regarding the short-term recruitment of a large number of enumerators to carry out the field work. These costs represent about 50–60% of the operation's total budget (Valente 2010b). Another problem associated with this method is presented by the increasing difficulty in conducting population surveys based on field data collection. For security reasons, many citizens, especially the elderly and those living alone, refuse to open the doors of their home to enumerators (Valente 2011). On the other hand, the rate of change in modern societies increases the demand for statistical information and the need for more frequent updates than traditional censuses allow.

3.1.1. Traditional Method with Long and Short Forms

The traditional census approach may also include the use of long and short forms, which can ease the burden on respondents and reduce the cost of census operations. The short form, with wider coverage (majority or all of the population), is intended to collect basic information on the characteristics of housing and/or population (e.g., place of usual residence, sex, age and number of household members). The long form is more detailed

and only answered by a sample of the population, usually between 10 to 20% of the total potential respondents. According to the UNSD, 14 countries, including Brazil, Canada, Mexico, Russia, and the United States (up to the 2000 Census) use or have used a mixture of short and long forms (UNSD 2010).

3.1.2.   Traditional Method with Sampling Annual Updates

This method combines the traditional model, carried out at intervals of five or ten years, with sampling-based annual updates. In the census year, the entire population of the country is enumerated using a short form that collects only basic socioeconomic and demographic characteristics. In the intercensus years, annual sample surveys are conducted with more detailed questionnaires (long questionnaire).

After the 2000 Census, the United States started to use this method, which included the exhaustive census, using a short questionnaire directed at the whole population and repeated every ten years. In the intercensus years, an annual survey is carried out with a larger number of variables – ACS (American Community Survey) – covering approximately two percent of the total population, which in 2010 was about 320 million (Herman 2008; Woods 2009). The ACS was fully phased in by 2005.

When compared with the decennial census, this method has the advantage of being able to provide results with greater frequency and timeliness. In exhaustive census years, it also reduces the complexity and burden involved in the use of the long form. Furthermore, in the intercensus period, it allows for the methods and techniques of sample surveys to be developed and readjusted. However, this method has certain disadvantages: the high financial costs of carrying out the surveys annually and the technical complexity of the associated procedures, especially the construction of the estimators; moreover, the data on detailed characteristics are limited since they only come from a sample survey (the ACS).

3.2.   Register-Based Census

The register-based census method does not use field operations and forms to collect data. The census statistical information is produced solely and exclusively based on administrative data, which are updated regularly according to input information from administrative acts carried out by the population on a daily basis.

The Nordic countries were pioneers of this method. Denmark was the first country in the world to move from the traditional census to a register-based census based entirely on administrative registers. The long-term work and strategy behind the first register-based census in 1981 is described in a book by Statistics Denmark (1995). In Finland, the use of administrative records for statistical purposes began as early as 1970 and has increased since that date (Statistics Finland 2004). Since 1990, the census has been based entirely on information in the registers without using a single form (Myrskylä 1991). In Norway and Sweden, as well as in Austria, exhaustive census operations have also been abandoned and the 2011 Census was fully supported by administrative data (Tönder 2008; Andersen and Utne 2011; Berka et al. 2010).

It is estimated that in Finland, the costs of conducting the census through the register-based census method were less than a tenth of the cost of using a classical method with postal questionnaires but without enumerators (Statistics Finland 2004). However, it

should be noted that the costs required for establishing and maintaining a register-based statistical system can be very significant. Other benefits associated with this method are that the information processing is faster than the traditional method, which also implies greater speed in the delivery of results. The main disadvantages in using this method result from restrictions on access to administrative data and their limitations in terms of content and quality. The variables used are defined by the administrative needs and rules of the organisations that produce them and not from a statistical perspective. Administrative data are often incomplete, inconsistent, outdated, or limited in their coverage. In many areas, some of the mandatory variables, included in international recommendations, may be difficult to obtain or impossible to find. Moreover, administrative data are focused on individuals and generally provide limited information on families, which limits social analysis (Dugmore et al. 2011; Zhang 2011). Finally, the concepts and classifications associated with the variables may not correspond to the statistical concepts that must be observed.

For a system of administrative registers to work effectively, it is necessary to ensure the links between the different records, which is generally possible with a unique identification key. This should correspond to an identifier which is not subject to mutations over time and that unambiguously identifies only one statistical unit. In Denmark a crucial component was the public administration's introduction of a fixed personal identification number for each individual, which replaced the different identifications previously used (Borchsenius 2000). Equally essential was the coding of addresses, which are considered a key link in the whole system. These are assigned a unique number (address code), thus allowing an interconnection between, for example, the Central Population Register and the Buildings/Dwellings Register. In countries without address codes, other methodologies of matching files are under investigation (Maldonado et al. 2010; Winkler 2011; Conti et al. 2012; Zhang 2012).

3.3. Combined Methods

Some countries obtain the census information through a combination of methods, designated as combined or mixed methods.

3.3.1. Traditional Method Using Administrative Registers

Some countries use administrative information to improve the accuracy of enumerations and the quality of data. They use address lists to support field operations and may send the questionnaires to respondents via mail. Part of the questionnaire may already be completed with data obtained from administrative sources (e.g., housing address, occupant names, sex, date of birth, etc.). The respondents or enumerators (through direct interview) only correct or update the information and complete the remaining questions. The average time of interview or completion is substantially reduced, which implies a positive impact on costs and improves the quality of the data. In addition, coverage can be evaluated by comparing the population register and fieldwork results.

When compared to the register-based census model, this method is more expensive, complex in its implementation and increases the burden on respondents (Redfern 1989). Some countries used this model in 2011: the Czech Republic, Estonia, Latvia, Lithuania,

and Italy (Valente 2010b). Spain also considered using it (Ballano 2008), though it finally opted for a combined method of administrative records, exhaustive collection of information on buildings, and sample surveys of dwellings and population.

3.3.2.   Administrative Registers and Sample Surveys

As the administrative registers do not contain all the information required, some countries complement registers with sample surveys. The registers are used to ensure that the entire population is counted and the survey results allow the missing individual characteristics to be obtained. The surveys can be designed specifically for the census (ad hoc surveys) or may already exist.

A mixed method, using administrative registers combined with sample surveys already in use, was implemented in the Netherlands for the first time in the 2001 Census, and was also adopted in 2011. This model, known as the Virtual Census (Nordholt 2005), does not require specific field operations but implies a complex estimation process for the lower levels of breakdown (Houbiers 2004), links between the records (Linder 2004; Nordholt and Linder 2007) and strict quality control (Daas et al. 2009; Nordholt et al. 2011). Slovenia also adopted this model in 2011 (Dolenc 2010). Germany complements this model with additional specific surveys (Eppmann et al. 2006; Szenzenstein 2005). Even using existing surveys, the model can also be combined with ad hoc surveys to evaluate the accuracy and degree of record completion or to add new variables (such as in long forms). Israel adopted this approach in the 2008 Census, thereby improving the accuracy of population registers and adjusting their counts (Valente 2010b).

3.4.   Rolling Census

France is the only country that uses the rolling census method, first proposed by Kish (1986; 1990). Implemented in 2004, it is based on annual surveys, which each year cover about 14% of the total population in parts of the country during a five-year cycle. The nearly 37,000 communities in France are classified into two groups: small and medium-sized communities (with fewer than 10,000 residents) and large communities (with 10,000 or more residents). Small and medium-sized communities, containing about half of the country's total population, are divided into five groups. Every year during the cycle, in rotation, each group is subject to an exhaustive census of all dwellings and people. In large communities, during the cycle, a sample survey, covering about eight percent of the dwellings, is held annually. At the end of five consecutive years the whole population of small and medium-sized communities and approximately 40% of the population of large communities has been surveyed. Overall, about 70% of the French population is covered during the entire lifecycle (Durr and Dumais 2002).

The advantages of this method are the possibility of distributing the efforts and costs over five years and improvements in the frequency of results – annual results in contrast to classical methods. The major disadvantage is the mobility of the census moment, which implies that data are not collected simultaneously for the whole population. Even if the data collected are adjusted to the same period, this poses certain difficulties in comparing areas surveyed at different times. The respondents' mobility over the five years also has implications in the model – it can cause gaps or duplications in the population. It also has

the disadvantage of involving a highly complex methodological approach, especially with the use of sampling techniques and modelling.

4. Census Transition in Portugal

In recent years, Portugal has been promoting policies that will modernise public administration services. In 2006, one of the most visible consequences of this strategy was the creation of SIMPLEX, a national governmental programme for administrative and legislative simplification that aims to improve and facilitate the interaction of citizens and businesses with public administration. One example of such advancements is the IES program – Simplified Business Information. This enables enterprises to reply to the Public Administration only once, through an electronic form, replacing several surveys that contained the same questions and were collected by different entities (Ministry of Finance, Bank of Portugal, Statistics Portugal, etc.). With a view to continuing the modernisation of the data collection process, the Action Programme for the 2011 Census provided the assessment of administrative registers for statistical purposes and a methodological review (INE 2010).

As mentioned above, a fundamental aspect resulting from other countries' experience of the use of administrative data is the existence of specific legislation that allows national statistical institutes access to these data (Wallgren and Wallgren 2007). For the first time in Portugal, the National Statistical Act allows Statistics Portugal to access individual administrative data collected by public sector entities. Following the trend of other UNECE countries, Portugal is thus also able to evaluate a new census model based on administrative data. The need for census model transition is underpinned not only by the high financial resources allocated to traditional census operations, but also the enormous effort required of citizens every ten years. Accordingly, the main motives for the transition are focused on contributions to society: to decrease the burden on citizens, to allow for a greater frequency of census data (annual) and to reduce the high costs associated with census operations. According to the description of census methods presented in Section 3, the methods that best fit these goals are those based on administrative data: the register-based census and combined methods. Their adoption reflects a change of paradigm in census operations in Portugal because it involves (re)thinking the approach and methodological design associated with the production of statistics based on administrative sources. Furthermore, it requires coordination between the different entities that produce and manage the administrative data (Statistics Portugal 2010). Thus the problem is to define a new methodology, based wholly or partly on administrative data, so as to replace the traditional population and housing censuses in Portugal. However, in order to achieve this objective, we must first find answers to the following questions:

- What administrative registers are available and what information do they contain?
- Does the information that exists in administrative data meet the requirements of census users, international recommendations and EU regulations?
- Do the available variables correspond to the fundamental questions of the census in terms of coverage, content and quality?
- What gaps exist in terms of census variables and what methodologies should be implemented to obtain the desired information?

In order to answer these questions, it will be necessary to evaluate the existing administrative registers in Portugal and their potential to produce census statistics.

5.   Administrative Registers in Portugal

One of the main dimensions to consider when evaluating a transition model is whether there are administrative files and registers with individual data (microdata) and unique identification that will be of interest to the census. In countries that have implemented systems of administrative records for statistical purposes, the combination of different sources was also the key factor in the process. To achieve this combination, the quality of the various sources was assessed by comparing and validating the information (UNECE 2007; Wallgren and Wallgren 2007). Following these principles, the research sought to identify administrative sources covering the statistical units used in the census: housing (buildings and dwellings) and population (households and persons). The first step therefore consisted of identifying and assessing sources with statistical units of housing. The second phase of activities focused on population data. Table 1 shows the main administrative registers identified as having potential for the mandatory census variables required by international recommendations, as well as the entities that manage them.

The records in the potential files of interest must be evaluated in terms of coverage, content, quality and identifiers. This task involves the analysis of:

- Individual records, the level of harmonisation, standardisation and consistency of information collected by different entities,
- Updating and management systems,
- Metadata associated with the data.

It is important to note that the information collected by administrative entities does not necessarily correspond to the statistical concepts to be observed according to international recommendations. In order to be used as statistical data, administrative registers undergo several transformations: coding of variables and creation of derived variables and validations, among others (Wallgren and Wallgren 2007).

Under the current legislation, Statistics Portugal has access to some administrative registers with potential for characterising census variables. As far as housing units are concerned, Statistics Portugal has access to the real estate register (municipal property tax), income register (personal income tax) and energy register.

The access to administrative registers of individuals has posed some difficulties. Based on different interpretations of the current Portuguese statistical act, register managers have been reluctant to share their registers. They only allow access when the Portuguese Data Protection Authority gives its consent. Statistics Portugal has assessed, among others, the civil register, the social security register, the employment register and the foreigner register.

The following sections present the preliminary results of the analysis of available administrative records for housing and population statistical units.

5.1.   Housing

With regard to housing information, the real estate register is the core register among the identified relevant administrative registers. It is the most extensive register available in

Table 1. Relevant administrative registers, register holders and content

Administrative registers	Register holder	Content
Real estate register -Municipal property tax	Tax and Customs Authority of the Ministry of Finance	Buildings and dwellings characteristics
Energy register	Energy register – Portugal Energy	Household addresses of electric energy consumption
Civil register	Registers and Notaries Institute of the Ministry of Justice	Some census variables related to population
Social security register	Information Institute of the Ministry of Solidarity and Social Security	Potential interest of some socioeconomic variables
Employment register	Strategy and Planning Office of the Ministry of Solidarity and Social Security	Information about workers under individual contract of employment
Income register - Personal income tax	Tax and Customs Authority of the Ministry of Finance	Personal tax addresses; household structure
Foreigner register	Foreigner register – Foreigners and Borders Office, Ministry of Internal Administration	Characterisation of the foreign population
Unemployment register	Unemployment register – Employment and Professional Training Institute, Ministry of Economy and Employment	Data on registered unemployed
Public administration register	Public administration register – Directorate General of Public Administration of the Ministry of Finance	Data on public administration staff
Pensioner register	Pensioner register – Pensioners Office of the Ministry of Finance	Data on pensioners
University register	Directorate General of Statistics and Science	Higher education administrative data
School register	Directorate General of Statistics and Science	Education administrative data (not higher education)

Table 2.   Available EU mandatory census variables related to housing

Variables available in the real estate register	Variables not available in the real estate register
Type of living quarters	Housing arrangements
Location of living quarters	Occupancy status
Useful floor space and/or number of rooms	Type of ownership
Water supply system	Number of occupants
Toilet facilities	Bathing facilities
Type of heating	Density standard
Type of building/number of dwellings	
Period of construction	

regard to coverage, content and identifiers. Real estate register data are produced by administrative acts covering municipal property tax. All properties (buildings or parts) located in Portugal that pay these taxes are included. As shown in Table 2, the real estate register covers around 57% (eight out of 14) of the mandatory variables of the EU regulations for the 2010 Census Round (Eurostat 2011; UNECE 2006) concerning building and dwelling characteristics. It will not be possible to ascertain the housing arrangements, occupancy status, type of ownership, number of occupants, bathing facilities and density standard.

Moreover, there are significant differences in terms of concepts and categories associated with the real estate register variables and the census variables required. The concepts of *housing property* and *fraction*, used in tax administration, are different from the census statistical units *building* and *dwelling*. In addition to these problems, we identified limitations in terms of harmonisation of fields and low rates of completion of some variables. To complete the missing information, other potentially useful files must be found; alternatively, the possibility of including new fields or making changes to existing forms associated with the real estate register should be evaluated in collaboration with the register managers.

The analysed data file from 2010 (referenced to December 31) only covers administrative acts carried out between 2003 and 2010. The housing properties of this flow of data only represent around 36% of total dwellings obtained in the provisional results of the 2011 Census. Statistics Portugal also needs to request the real estate register keepers'/managers' permission to access the global stock of housing properties. In terms of housing topics, the energy register (2008) and the income register (2008) were also analysed. These registers do not provide census information, thus their potential is limited to the use of additional information to update the housing addresses.

5.2.   Population

As for population variables, about 92% (22 out of 24) of the mandatory census variables of EU regulations (Eurostat 2011; UNECE 2006) are represented in some of the existing registers (Table 3).

Some of these variables can only be obtained through the combination of two or more registers. The data present important limitations in terms of content (suitable concepts) and coverage. For example, with data from the income register and the civil register it is

Table 3. Available EU mandatory census variables related to population

Variables	Civil register	Social security register	Employment register	Income register	Foreigner register
Place of usual residence	x	x		x	x
Location of place of work or school			x		
Sex	x	x	x	x	x
Birth date (age)	x	x		x	x
Legal marital status	x	x		x	
Current activity status		x			
Occupation			x		x
Industry, branch of economic activity		x		x	
Status in employment		x	x		x
Educational attainment			x		
Country/place of birth	x	x			x
Country of citizenship	x	x	x		x
Ever resided abroad, year of arrival					
Place of residence before census year	x				
Relation between household members				x	
Tenure status of households					
Total population	x				x
Locality	x				
Household status				x	
Family status				x	
Type of family nucleus				x	
Size of family nucleus				x	
Type of private household				x	
Size of private household				x	

possible to construct some private households and family nuclei, but these registers do not cover the people exempt from income tax, such as persons with incomes lower than a defined value. For tax purposes, a young adult over 18 years working and living at his or her parents' house (sharing food and possibly other essentials for living) is not included on the same tax form (defining an income register household) as his parents. Instead, he or she is represented with a separate tax form. In census concepts, a young adult and his parents constitute a single household. At the moment, Statistics Portugal is unable to access the complete file of the income register. Some attempts have been made to gain access to these administrative data but so far without success.

The civil register file should be the core register for the population statistical units. However, some population groups are not represented in the civil register file: the foreign resident population (which has no Portuguese civil identification) with the exception of Brazilian citizens with equal status (resulting from the Treaty of Porto Seguro agreed between Portugal and Brazil), and children who were born before 2007 and do not have the

Citizen Card (CC). The "Born Citizen" project, introduced in 2007, means that when children are born and registered they receive a civil identification number even if this is not requested (the CC is only compulsory from six years of age). Those numbers will stay the same throughout their entire life. The only exhaustive coverage of all people legally residing in Portugal is the combination of the civil register and foreigner register files. However, none of these administrative registers covers the illegal population. In contrast, the traditional census gives us a global "picture" and all individuals, regardless of their legal status, should be enumerated. On the other hand, many people who are listed in the civil register with a Portuguese residence effectively reside in another country and are not enumerated in a traditional census. In the 2011 Census, an individual is considered resident when he "lives in his usual place of residence in the twelve months preceding the time of the census with the intention to stay for a minimum period of one year" (UNECE 2006).

The 2009 and 2010 civil register (referring to December 31) data files were analysed. A set of strengths and points that require better assessment for their potential use were identified for the 2010 data:

- The file presents a low percentage of null records (missing data) and a primary numeric key for all records without duplication;
- The address fields of the individuals who already have the CC are standardised and are of good quality, which may allow them to be matched with other sources;
- The administrative division classification maintains the existing code on the date of occurrence. Thus the codes for geographic places of birth and residence are not adjusted to the administrative division changes that have occurred over the years, implying, for example, the existence of the same codes for different administrative areas;
- It is clear that bilateral cooperation between Statistics Portugal and keepers/managers is important in order to improve the quality of civil register data through standardisation and joint verification of discrepancies.

The foreign population residing in Portugal was enumerated in Census 2011 but does not appear in the civil register files. The total resident population in Portugal obtained in the provisional results – the reference date was the census moment (21 March 2011) – was 10,561,614. The civil register (on 31 December 2010) presents 7.1% more resident persons than the 2011 Census. The two sets of data were not collected on the same date. For a correct comparison it would be necessary to make adjustments for the same reference date. However, since the gap is less than three months and the Census 2011 results are provisional, we do not make the adjustments here.

The protocol signed between Statistics Portugal and Social Security permits annual access to the social security register data. The stock of 2008 to 2010 (referring to December 31) social security register data files was analysed. The data present quality for use: they contain an exhaustive primary key (Social Security Identification Number) for all registers; the variables are standardised and present a low percentage of missing data; the percentage of inconsistent data is small; and the classifications are those used by Statistics Portugal. According to the keeper/manager of these administrative registers, the gaps and incoherencies (e.g., geographic codes) in some variables are the result of a process which occurred in the late 1990s. The Ministry of Solidarity and Social Security

gathered all regional registers into one centralised database without a verification or confirmation process. Currently, any update or correction of data concerning beneficiaries is automatically registered in the database and subject to an intense validation process, contributing to increased quality of data. The numeric field NIF (corresponding to the Tax Identification Number) presents a fill rate of around 96%. Thus this field could be used as a key liaison with other administrative files.

The employment register is obtained from the administrative data submitted annually by all employers with workers employed under individual contract of employment. The online application used to capture information establishes a set of validations that guarantee overall quality and consistency of this file. The stock of 2009 (referring to 31 October) employment register data files was analysed. The data present quality for use: they contain an exhaustive primary key (NISS – Social Security Identification Number) for all registers; the variables are standardised and present a low percentage of missing data; and the classifications are those used by Statistics Portugal. Later, it was found that the employment register keeper/manager has other available mandatory variables that have not yet been provided: date of birth and branch of economic activity.

Finally, the foreigner register can ensure the coverage and content of a very specific portion of the population (foreigners legally residing in Portugal), completing the information in the civil register files. The stock of 2009 (referring to 31 December) of the foreigner register was analysed. The data present some problems for use for statistical purposes: the identification key corresponds to a code number assigned by the foreigner register, which prevents matching with other files; the geographical breakdown only exists at municipal level; and some fields are not validated, notably in the address fields. Later, it was found that the foreigner register keeper/manager had not provided other available mandatory variables: country of last residence, Social Security Identification Number and Tax Identification Number, which could be the key liaison with other administrative files such as the social security register and employment register.

5.3. Integration

The development of an information system to store the individual registers of housing and population must be part of the transition census process. This can be based on existing models in other countries (UNECE 2007; Wallgren and Wallgren 2007), adapted to the Portuguese reality. For statistical purposes, this system shall include the national files of dwellings and persons, which may be based on the 2011 Census microdata and subsequent updates with housing and population administrative data.

As already mentioned above, an important part of administrative information is dispersed across multiple files managed autonomously by different entities. In addition to storage issues, the various registers within the system must be linked if the administrative registers are to be transformed into statistical records.

The geographical key location of buildings and dwelling, to be defined, could be address type (full address information and postcode), coordinate ID type (X, Y) or a combination of both. This key is fundamental in linking housing registers to population registers. The real estate register uses geographical coordinates and address.

The structure of the National Population Register for statistical purposes based on administrative data requires a unique key. This key, which we shall call the administrative key, is of great importance since it will ensure that all the administrative information on the population can be successfully integrated. The key would ideally be composed of only a single field, but due to the complexity of linking files (already experienced in the analysis), we cannot rule out the need to develop a composite key. There are files that, due to the absence of common fields, can only be related through a key from a third file. Table 4 presents the possible fields that might allow the links between files. For example, the numeric field NIF (Tax Identification Number) could be used as a key liaison between the income register, the social security register and the real estate register. In the same way, the numeric field NISS (Social Security Identification Number) could be used as a key liaison between the civil register and the social security register. In addition to the possible connections through the numeric codes NIC (Civil Identification Number), NIF and NISS, other possibilities could be tested, since the variables are standardised. For example, "Address" connects statistical units of housing to units in the population associated with aggregated "Name and Date of Birth" or "Address".

Table 4.   *Identities in registers that can be used as matching keys*

Administrative registers	Name	Address	Geo code	NIC	NISS	NIF
Variables related to housing						
Real estate register		Address	Geo code			NIF
Energy register		Address	Geo code			
Variables related to population						
Civil register	Name	Address		NIC		
Social security register		Address		NIC	NISS	NIF
Employment register					NISS	
Income register		Address				NIF
Foreigner register	Name	Address				

NIC Civil Identification Number.
NISS Social Security Identification Number.
NIF Tax Identification Number.

6.   Possible Models for the 2021 Census in Portugal

To prepare for the 2021 Census, a model for use could be established by comparing the results of the 2011 Census with existing administrative data. To avoid an exhaustive comparison of all administrative data and 2011 Census data, it should be possible to implement a test structure: choosing a number of areas in the country (e.g., a sample of municipalities) to compare data collected from the 2011 Census with the corresponding administrative data at the micro level of each statistical unit. The comparison with existing administrative records could provide answers to the following question: if the 2011 Census had not been carried out using the traditional model, could we have obtained consistent and relevant data on population and housing through administrative sources?

As identified in Section 4, the methods that best fit the objectives defined for the Portuguese census transition are those based on administrative data: *register-based*

census, or *administrative registers and sample surveys*, or *traditional method using administrative registers*. These different approaches to census taking involve trade-offs between: overall quality of census information; the cost, complexity, and frequency of census data; and burden on citizens required for change. The register-based census is the first priority model to develop for the post-2011 Census in Portugal because of the low cost, the high frequency of data, and the fact that it is no burden on individuals. However, there may be difficulties in moving directly from a traditional to a complete register-based census as a result of the problems identified in the preliminary analysis of existing registers (performed in Section 5) and the lack of experience in the use of administrative registers for statistical census purposes. Only Austria has passed directly from a traditional census in one round (2000) to a register-based census in the next round (2010) (Valente 2011; Ralphs and Tutton 2011). This transition has generally taken several decades, according to the experience of the Nordic countries.

The second priority could be the combined method of administrative registers and sample surveys because of the reductions in cost and burden on individuals, when compared to the traditional approach. In order to implement it and complement the missing information in administrative registers, it will be necessary to evaluate all existing surveys in the Portuguese statistical system. The integration of administrative data and sample survey data requires a complex process of estimation and calibration, especially for areas with lower levels of disaggregation. Thus it will be necessary to evaluate and adapt existing models (Zanutto and Zaslavsky 2002; Houbiers 2004; Mulry et al. 2006).

If the second priority cannot be implemented, an alternative model to be developed for the 2021 Census could be the traditional method using administrative registers. With this method, all individuals are enumerated but, as explained in Subsection 3.3.1., the use of register data increases the efficiency in field operations: mailout of the questionnaires to all households in the list of households and addresses (obtained from administrative registers) and multichannel collection of responses (web, mail back, municipal office of collection). This approach may help to improve the coverage and quality of the registers and, as a result, it is often selected as the first step from a traditional census towards a register-based one (Valente 2011).

In order to implement a census based on administrative data, some of the problems identified in Section 5 need to be solved: incomplete coverage of the housing stock, excess of population in the civil register, lack of coverage of the illegal and foreign population, variables with small coverage of the population, incompatible identifiers and failure to access the complete income-tax records. Some proposals are presented to solve these problems. One fundamental element of a system for integrating administrative registers is the availability of a definitive National Address Register, providing a list of all housing addresses. To avoid the difficulties in accessing the global stock of housing properties, the housing data from Census 2011 could be the basis for the National Address Register. Additionally, it will be possible to check the coverage in the period 2003–2010 by comparing census housing data with the real estate register. If the results present a good coverage, the Address Register will be updated by the real estate register.

The problem of the excess of population in the civil register could be solved in the same way. The population data from Census 2011 could be the basis for the National Population

Register. If we compare the individual Census population data with the civil register, it will be easier to understand the differences between the two files in recent years. The results of this analysis may help to improve the quality of civil register data and evaluate the use of the civil register to update the National Population Register. However, the civil register update has more weaknesses than the real estate register updates. It will always be a complex challenge to detect people who are legally resident in Portugal and have a Citizen Card (included in the civil register) but who actually reside abroad all year and just spend vacations in their dwellings in Portugal.

The lack of coverage of the illegal and foreign population and variables with small coverage of the population could be overcome by implementing ad hoc sample surveys specifically designed for this purpose. The problem of incompatible identifiers between administrative registers could be minimised using the address as a key connection, since this field is present in almost all registers. Despite the difficulties of using this field for linking registers, new matching techniques developed in recent years have produced very robust results (Maldonado et al. 2010). In addition to solving the failure to access the complete income-tax records, it will be necessary to revise the National Statistical Act. The new legislation should be strong, clear and unambiguous, giving Statistics Portugal unrestricted access, for statistical purposes, to administrative data on unit level with identification data and the possibility to link them with other administrative registers. In order to accomplish the census information system for the transition, it will be necessary to gain access to more administrative sources, particularly in the areas not yet covered: education and unemployment.

7.   Conclusions

Decennial census operations are important and require large human, financial and material resources. Given these constraints and bearing in mind that statistical information is essential, the implementation of the census transition in Portugal is focused on three goals: to decrease the burden on citizens, to allow for a greater frequency of census data (annually if possible) and to reduce the high costs associated with census operations. This article presents a systematic critical review of alternative methodologies to traditional censuses, identifying their advantages and disadvantages as well as the countries that use them. Comparing several methods, it appears that those that best fit the objectives defined for the Portuguese case are the methods that rely on administrative data. However, it is also clear that the present legal framework and the nature and quality of available administrative registers still require changes or improvements in order to enable such methodologies.

Under the current Portuguese Statistics Act, Statistics Portugal has access to some administrative registers with individual data that have potential for obtaining census variables related to housing and population. With regard to housing topics, a core register has been identified – the real estate register file (municipal property tax). The real estate register has great potential for use, although there are differences in concepts, limitations in terms of harmonisation of fields and low rates of completion of some variables. The real estate register contains information flows and does not include the stock of buildings/properties. The data analysed between 2003 and 2010 represents around 36% of total dwellings obtained in the provisional results of the 2011 Census.

As for population variables, around 92% of the mandatory census variables are represented in some of the existing registers. However, the administrative registers identified present important limitations in terms of content (suitable concepts) and coverage. The civil register should be the core register for the population statistical units. Although they belong to different universes, we compared the total resident population in Portugal, obtained in the provisional results of the 2011 Census, with the civil register file (on 31 December 2010). The civil register file presents 7.1% more of the resident population than the 2011 Census. The social security and employment administrative registers (referring to 2010) were analysed and showed good quality for use in producing census statistical information. On the other hand, the analysis of the results of the foreigner register (referring to 2009) presented significant limitations in their use for statistical purposes.

The different approaches to census taking based on administrative data involve trade-offs between the overall quality of census information, the cost, complexity, and frequency of census data, and the burden on citizens required for change. The register-based census is probably the first priority model to develop for the post-2011 Census in Portugal. However, there may be difficulties in moving directly from a traditional to a complete register-based census as a result of the problems identified in the preliminary analysis of existing registers. The second priority could be the combined method of administrative registers and sample surveys. Nevertheless, the implementation of this model implies a complex process to integrate administrative data and sample survey data. An alternative model could be the traditional method using administrative registers. This approach enables efficiencies in field operations and may help to improve the coverage and quality of the registers. Therefore, this could be considered a first step in the right direction, contributing towards a future register-based census.

It was also possible to identify some key points in evaluating administrative registers that constitute challenges for the continuity of the work and for the strategy to be defined: incomplete coverage of the housing stock, excess of population on the civil register, lack of coverage of the illegal and foreign population, variables with small coverage of the population, incompatible identifiers and failure to access the complete income-tax records. In the article, some proposals have been presented that may contribute towards solving or minimising these problems.

Although the work done so far is only the first step towards using administrative registers for statistical purposes, a number of important lessons have already been learned that might be useful for other statistical agencies at the same transitional stage. It is crucial to have a strong legal basis that will provide the national statistics agency with the right to access administrative data at unit level with identifiers, and with the right to link them with other administrative registers for statistical purposes. On the other hand, there must be a high level of coordination and cooperation with the register managers to improve the quality of information that is collected administratively and when introducing potential adjustments to the collection forms. In the potential files of interest, the records must be evaluated in terms of coverage, content, quality, and identifiers. Identifiers play a considerable role in linking information from various sources. A potential line of work in overcoming the problem of incompatible identifiers between administrative registers is the use of the address as a key connection.

8.   References

Andersen, E. and H. Utne. 2011. "Censuses in a Register-Based Statistical System: Norwegian Experiences." Paper presented at the 58th World Statistics Congress ISI 2011, IP064.01, 21–26 August, Dublin, Ireland.

Ballano, C. 2008. "A Census of Population Based on an Administrative Register." Paper presented at the 24th International Methodology Symposium, Statistics Canada, 28–31 October, Ottawa, Canada.

Berka, C., S. Humer, M. Lenk, M. Moser, H. Rechta, and E. Schwerer. 2010. "A Quality Framework for Statistics Based on Administrative Data Sources Using the Example of the Austrian Census 2011." *Austrian Journal of Statistics* 39: 299–308.

Borchsenius, L. 2000. "From a Conventional to a Register-Based Census of Population." Paper presented at the INSEE/Eurostat Seminar on the Censuses after 2001, 20–21 November, Paris, France.

Conti, P., D. Marella, and M. Scanu. 2012. "Uncertainty Analysis in Statistical Matching." *Journal of Official Statistics* 28: 69–88.

Daas, P., S. Ossen, and J. Arends-Tóth. 2009. "Framework of Quality Assurance for Administrative Data Sources." Paper presented at the 57th World Statistics Congress ISI 2009, 16–22 August, Durban, South Africa.

Dolenc, D. 2010. "Quality Assessment in a Register-Based Census – Administrative Versus Statistical Concepts in the Case of Households." Paper presented at the European Conference on Quality in Official Statistics, 4–6 May, Helsinki, Finland.

Dugmore, K., P. Furness, B. Leventhal, and C. Moy. 2011. "Beyond the 2011 Census in the United Kingdom – With an International Perspective." *The Market Research Society* 53: 619–650.

Durr, J. and J. Dumais. 2002. "Redesign of the French Census of Population." *Survey Methodology* 28: 43–49.

Eppmann, H., S. Krügener, and J. Schäfer. 2006. "First German Register Based Census in 2011." *Allgemeines Statistisches Archiv* 90: 465–482. Doi: http://dx.doi.org/10.1007/s10182-006-0246-9.

Eurostat. 2011. *EU legislation on the 2011 Population and Housing Censuses – Explanatory Notes*. Luxembourg: Publications Office of the European Union.

Herman, E. 2008. "The American Community Survey: An Introduction to the Basics." *Government Information Quarterly* 25: 504–519. Doi: http://dx.doi.org/10.1016/j.giq.2007.08.006.

Houbiers, M. 2004. "Towards a Social Statistical Database and Unified Estimates at Statistics Netherlands." *Journal of Official Statistics* 20: 55–75.

Instituto Nacional de Estatística (INE). 2010. *Programa de Acção para os Censos 2011*. Lisbon, Portugal.

Kish, L. 1986. "Complete Censuses and Sample." *Journal of Official Statistics* 2: 381–395.

Kish, L. 1990. "Rolling Samples and Censuses." *Survey Methodology* 16: 63–93.

Linder, F. 2004. "The Dutch Virtual Census 2001: A New Approach by Combining Administrative Registers and Household Sample Surveys." *Austrian Journal of Statistics* 33: 69–88.

Maldonado, A., D. Scheuregger, and K. Ziprik. 2010. "Setting Up the Central Register of Addresses and Buildings of the German Census 2011." Paper presented at the European Conference on Quality in Official Statistics, 4–6 May, Helsinki, Finland.

Mulry, M., S. Bean, D. Bauder, D. Wagner, T. Mule, and R. Petroni. 2006. "Evaluation of Estimates of Census Duplication Using Administrative Records Information." *Journal of Official Statistics* 22: 655–679.

Myrskylä, P. 1991. "Census by Questionnaire – Census by Registers and Administrative Records: The Experience of Finland." *Journal of Official Statistics* 7: 457–474.

Nordholt, E. 2005. "The Dutch Virtual Census 2001: A New Approach by Combining Different Sources." *Statistical Journal of the United Nations Commission for Europe* 22: 25–37.

Nordholt, E. and F. Linder. 2007. "Record Matching for Census Purposes in the Netherlands." *Statistical Journal of the IAOS* 24: 163–171.

Nordholt, E., S. Ossen, and P. Daas. 2011. "Research on the Quality of Registers to Make Data Decisions in the Dutch Virtual Census." Paper presented at the 58th World Statistics Congress ISI 2011, STS050.01, 21–26 August, Dublin, Ireland. Available at: http://2011.isiproceedings.org/Abstracts/STS050.html (accessed April 2012).

Ralphs, M. and P. Tutton. 2011. "Beyond 2011: International Models for Census Taking: Current Processes and Future Developments." Working Paper: Beyond 2011 Project - Office for National Statistics. Available at: http://www.ons.gov.uk/ons/about-ons/what-we-do/programmes—projects/beyond-2011/news/reports-and-publications/early-reports-and-research-papers/international-models-for-census-taking.pdf (accessed April 2012).

Redfern, P. 1986. "Which Countries Will Follow the Scandinavian Lead in Taking a Register-Based Census of Population?" *Journal of Official Statistics* 2: 415–424.

Redfern, P. 1989. "Population Registers: Some Administrative and Statistical Pros and Cons." *Journal of the Royal Statistical Society, Series A* 152: 1–41. Doi: http://dx.doi.org/10.2307/2982819.

Scheuren, F. 1999. "Administrative Records and Census Taking." *Survey Methodology* 25: 151–160.

Statistics Denmark. 1995. *Statistics on Persons in Denmark – A Register-Based Statistical System*. Luxembourg: Eurostat and Denmark Statistics.

Statistics Finland. 2004. *Use of Registers and Administrative Data Sources for Statistical Purposes – Best Practices of Statistics Finland*. Handbook 45, Helsinki, Finland.

Statistics Portugal. 2010. "Implementing a Register-Based Census in Portugal: Changing the Paradigm." Paper presented at Joint UNECE/Eurostat Expert Group Meeting on Register-Based Censuses, 10–11 May, The Hague, The Netherlands.

Szenzenstein, J. 2005. "The New Method of the Next German Population Census." *Statistical Journal of the United Nations Commission for Europe* 22: 59–71.

Tönder, J.-K. 2008. "The Register-Based Statistical System: Preconditions and Processes." Paper presented at the International Association for Official Statistics Conference, 14–18 October, Shanghai, China.

United Nations Economic Commission for Europe (UNECE). 2006. *Conference of European Statisticians Recommendations for the 2010 Censuses of Population and*

Housing, Prepared in cooperation with the Statistical Office of the European Communities (Eurostat). New York and Geneva: United Nations Publications.

United Nations Economic Commission for Europe (UNECE). 2007. *Register-Based Statistics in the Nordic Countries: Review of Best Practices with Focus on Population and Social Statistics*. Geneva, Switzerland: United Nations Publications.

United Nations Economic Commission for Europe (UNECE). 2013. "Census Methodology: Key Results of the UNECE Survey on National Census Practices, and First Proposals about the CES Recommendations for the 2020 Census Round." Paper presented at the Conference of European Statisticians – Group of Experts on Population and Housing Censuses, 30 September – 3 October, Geneva, Switzerland. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2013/census_meeting/3_E_x_15_Aug_WEB_revised_map.pdf (accessed September 2014).

United Nations Statistic Division (UNSD). 2010. *Report on the Results of a Survey on Census Methods Used by Countries in the 2010 Census Round*. Working Paper: UNSD/DSSB/1. Available at: http://unstats.un.org/unsd/demographic/sources/census/2010_phc/docs/ReportOnSurveyFor2010Census.pdf (accessed April 2012).

Valente, P. 2010a. "Main Results of the UNECE/UNSD Survey on the 2010/2011 Round of Census in the UNECE Region". Paper presented at the Working Group on Demography and Census - Eurostat, 19–20 April, Luxembourg.

Valente, P. 2010b. "Census Taking in Europe: How Are Populations Counted in 2010?" *Population & Societies* 467: 1–4.

Valente, P. 2011. "Innovative Approaches to Census-Taking: Overview of the 2011 Census Round in Europe." Paper presented at the conference "Statistics in the 150 years from Italian Unification", June 8–10, Bologna, Italy.

Wallgren, A. and B. Wallgren. 2007. *Register-Based Statistics – Administrative Data for Statistical Purposes*. Chichester, UK: John Wiley & Sons.

Winkler, W. 2011. "Machine Learning and Record Linkage." Paper presented at the 58th World Statistics Congress ISI 2011, IPS057.01, 21–26 August, Dublin, Ireland. Available at: http://2011.isiproceedings.org/papers/450070.pdf (accessed January 2016).

Woods, S. 2009. "Evaluating Population Estimates in the United States: Counting the Population Between the Censuses." *Government Information Quarterly* 26: 144–147.

Zanutto, E. and A. Zaslavsky. 2002. "Using Administrative Records to Improve Small Area Estimation: An Example from the U.S. Decennial Census." *Journal of Official Statistics* 18: 559–576.

Zhang, L. 2011. "A United-Error Theory for Register-Based Household Statistics." *Journal of Official Statistics* 27: 415–432.

Zhang, L. 2012. "Topics of Statistical Theory for Register-Based Statistics and Data Integration." *Statistica Neerlandica* 66: 41–63. Doi: http://dx.doi.org/10.1111/j.1467-9574.2011.00508.x.