The 2021 population and housing Census in Spain: challenges and findings

Jorge L. Vega Valle*
National Statistics Institute, Madrid, Spain – jorgeluis.vega.valle@ine.es

Antonio Argüeso Jiménez
National Statistics Institute, Madrid, Spain – antonio.argueso.jimenez@ine.es

Marina Pérez Julián
National Statistics Institute, Madrid, Spain – marina.perez.julian@ine.es

Abstract

From the methodological point of view Spain has carried out the Census on a different way in each one of the latest three editions. 2021 Census can be considered as the end of a long process that started more than 20 years ago with the inception of the Population Register, but also as the beginning of a new system of social information mainly based on registers.

The decision to move towards a register-based Census is based on a simple principle: it will provide more information and with higher quality than any other alternative. Moreover this strategy has several collateral advantages that must not be underestimated: reduction of burden, avoid the logistic challenge that traditional censuses imply, possibility of including new variables, save in the budget, improvement of the amount and quality of statistical information available for other statistics…

Keywords: Registers; quality; data-linkage, administrative data.

1. Introduction

In the latest editions Spain carried out the Census using different methodologies: traditional (1991), combined: traditional + registers (2001) and combined: survey + registers (2011).

Spain was by no means the only country that gave up collecting questionnaires from all households in 2011. The tendency to build census information without the need to send questionnaires to the entire population, initiated by some pioneer countries in the 1980s, has been growing gradually so that in 2011 there were already about twenty countries in the world that abandoned the traditional method. The exhaustive collection is replaced by other more efficient and higher quality alternatives.

The number of countries that do not expect to carry out a traditional census in the next round continues to grow[1]. And Spain (46.5 million inhabitants on 1st July 2016) will be among them, carrying out perhaps the largest Register-based census in the world, as the most populated country that developed it in the 2011 round was the Netherlands (with near 17 million inhabitants).

There are three important elements that justify the changes in the Spanish census methodology. First, and more important from the point of view of the production of demographic information, was the

[1] According to a survey carried out by UNECE in April 2015, 13 countries plan to have a register based Census in 2021 and 9 countries plan a combined Census.

creation of the Population Register in 1996. Secondly, the enormous variety of administrative data in Spain that contain key information and the improvement of these registers during the last years. And last, INE[2] has access to these data according to Spanish Statistical Law[3].

After finalising the works of 2011 Census, a deep quality analysis of the information that the registers contain was carried out and we pleasingly discovered that most of the foreseen drawbacks[4] can be solved. We can conclude that we are in the best situation ever to carry out our first register based Census in history.

2. The Spanish Population Register (*Padron*) is different.

In order to carry out a register based Census, it is key to have a high quality Population Register (PR) as the PR is the backbone of any register-based census.

The Spanish PR (*Padrón*) is very different from others. Unlike other countries where the Police or other administrative bodies are in charge of population registers, in the case of Spain, INE is the national institution that coordinates this single national population register.

There are more than 8,000 municipalities in Spain. Every person residing in Spain must be registered in the municipality where they usually[5] live. Thus, there are around 8,000 PRs in Spain. The creation, maintenance, revision and safekeeping of every population register is carried out by each municipality.

According to Law, INE is responsible for the coordination and integration of all these municipal PRs into a single national population database. This is what we actually call "the *Padrón*". In order to carry out this task INE receives every month all changes produced in every municipality, performs validations and forward these results to all the municipalities, to avoid duplications, but also to include deaths, births or acquisition of Spanish citizenship that INE receives on a monthly basis (or more frequently) from the Civil Register.

The fact that INE has direct access to the PR makes the Spanish situation very different from other countries. It makes things easier to produce other statistical operations such as the Census. The *Padrón* contains only a few variables for each person: name and surname, identification number (ID card number[6] or passport), sex, date and place of birth, citizenship, current address, and educational attainment. Contrary to other countries[7] in Spain we have direct access to personal information which allows to improve data-linkage percentages among data sources.

This system has been working since 1996 and continuous improvements have been made during the last years. For example, coordination and computer developments among municipalities have been included,

---

[2] INE: Instituto Nacional de Estadística (National Statistics Institute of Spain)

[3] Law 12/1989, of 9 May on the Public Statistical Function (LFEP)
https://www.boe.es/boe/dias/1989/05/11/pdfs/A14026-14035.pdf
In this law it is stated that INE can have access to information from other data-sources for statistical purposes. Apart from that, the fact that the Census is an operation with international regulation reinforce this permission in order to access to external information.

[4] Described more in depth in the beginning of section 3.

[5] If a person lives in several municipalities, he/she must register in that one where they live most days during a year.

[6] Compulsory for all the people with Spanish nationality and aged over 14.

[7] The most common situation in other NSIs is having access only to a PIN number but not to personal information such as name and surname.

some expiration mechanisms to delete registration of foreign people that are no longer living in Spain have been considered, special revisions and personal checks carried out by enumerators in municipalities with more doubtful information…

3. Building a Census file around a Population Register: adding variables

In 2014, once the 2011 Census production was over, we started the analysis of information available to build a register-based Census in 2021. During this period of time in which we have been conducting several pilot studies, for each obstacle we faced, we found its correspondent solution. For example, information about legal marital status, not available in the *Padrón*, was found to be, and with high quality, in the Tax Agency registers; we also discovered that information about current occupation is included in some Employment Registers and so on.

Since 2014 we have been checking the quality of the available administrative sources focusing on some small pilot projects at NUTS3 level[8] and comparing 2011 Census information with the new one based in administrative data.

The analysis of sources is not immediate because there are many variables collected in a census like this one, which is not only collecting information on population but also on buildings, housing and dwellings. Some of the variables are very easy to obtain from records, but in other cases the challenge is much greater.

Today the Census Project is organized in two main strands: population and dwellings. The first project is trying to integrate data coming from other sources available to produce census-like information about the population. The goal today is to make a kind of first pre-census at the beginning of 2018 with all the available sources at 1st January 2016 as reference date.

It is important to enlighten that each variable or group of variables needs to be studied in depth and sets out specific problems that must be tackled individually.

If we group by type of variables, these are the data sources that are under study today to produce the Census information:

- Migration variables: The *Padrón* is the main source but, as it started in 1996, it has been linked to 2001 (and 2011) Census to provide data on movements that happened before 1996.

- Legal marital status and household type: the Padrón only provides information about people living in the same address but it offers no information about relationship among them. We can use data from the Tax Agency, Social Security, marriages bulletins, civil register information and Register of Foreign Residents (residence permits), or even Identity Cards (where name of father and mother is included) to build the household type.

- Economic variables (labour status, occupation…): Social Security registers, Contracts Registers, Unemployment registers, Public Aids Register, Specific groups mutualities (i.e: Armed forces, civil servants…) and Tax Agency registers are used.

[8] NUTS is a European hierarchical and geographical classification. Spain is divided into 7 NUTS1 regions, 19 NUTS2 regions and 59 NUTS3 regions.

- Educational attainment: The *Padrón* contains basic information for all the people as the registration form contains a variable with 9 categories but it has to be improved and updated combining with 2001 and 2011 Censuses, Unemployment register or Ministry of Education registers.

It is obvious that it will not be possible to obtain all the information exclusively from administrative sources and some missing information will have to be imputed, normally using external information from existing surveys. For example, in those dwellings occupied by two persons (of the same sex or of different sex) is not always clear whether they are a couple or not, especially if registers do not contain information about it.

The second strand is dwellings (and buildings). It is planned to obtain information by integrating various administrative sources, namely: *Padrón*, Census 2011, the Cadaster and electricity consumption registers. The idea is to make use of existing information in all these sources to determine the total number of dwellings within each population area, the most basic being the census section. This is actually the most challenging part of the 2021 Census from the technical point of view.

4. Measuring quality in a register-based Census

The availability of administrative records provides, in general, a qualitative step forward in the quality compared to the collection of the information through households questionnaires.

For example, if we were able to obtain from administrative data the educational attainment for the entire population, it is clear that this data would offer much more information and it would be more reliable than the one we could collect from interviewees.

But in order to allocate the educational attainment to the entire population, it will be necessary to find it among the different sources (that of course have different quality). It will not always be possible for all administrative files to find data for the entire population, and of course the educational attainment is one of them.

The quality of the information provided in a register-based census can be quantified, to some extent, by means of a very simple method, which can be refined. Although the detailed way of assessing quality is still not fully decided, the strategy that is intended to follow will be based on these general lines:

We can think of the population census as a matrix of approximately 47 million rows (people) and around 100 columns (variables). The population Census is then a matrix containing several billion cells. In order to evaluate data quality of different sources, we are planning to include at least some metadata associated to each one of the Census cells[9]. The idea is to include information like which specific administrative register is the source for every data provided. At least it would be an indication, for every cell, of whether this data comes from direct (good quality) sources or even if this information has been imputed. For example, in the particular case of educational attainment it would be possible to assign a different value for every cell if the data comes from direct sources (Ministry of Education), indirect (previous censuses, the *Padrón*) or imputation.

Although this approach may seem relatively simple, it provides a powerful tool for the quality measurement of the Census and it helps us to refine its methodology.

[9] Although in this paper we only concentrate on the matrix with information about people, another matrix with information about dwellings can also be built.

An analysis by columns (variables) across people, allows us to detect for every variable involved what is the percentage of records provided by different data sources and the percentage of imputed records. This information helps us to detect the quality of sources.

If we focus our research on rows (people) we can identify those records with the poorest quality level: those that have missing values or imputed information in most of variables. It will not be surprising that these records have a common pattern such as being foreigners or living in particular (maybe deprived) areas. This information could be used, to conduct a specific survey in those geographic areas or addressed to these specific population groups with lower data quality and can help to improve Census data quality.
This idea of using specific, ad-hoc surveys, oriented to collect particular information in particular places is something still not decided. The same goes with dwellings and buildings. Specific fieldwork operations could be planned to solve situations where administrative registers do not provide enough information. But they would be, in any case, very limited operations in terms of people or territory (less than 1% of people or census sections).

5. Main advantages of a register-based Census

The reasons[10] that justify the change to a register-based Census are many:

- The response burden disappears or is reduced dramatically. This element is especially important in the current context where non-response rates in social surveys are continually increasing in all the countries.

- Logistics. A traditional census is a huge challenge for NSIs. Compared to other alternatives, a register-based census does not involve staff recruitment, pyramidal training to transmit knowledge, outsourcing of different IT developments…

- Budget reduction. Even though is not easy to evaluate all the costs related to the Census, in general register-based Census is almost zero compared to other alternatives. In Spain we moved from a 200M €[11] budget in 2001 to 85M € in 2011, and although 2021 information is not known yet, even considering ad-hoc surveys it would be surely below 5 million €.

- Completeness and coverage improvement. From previous censuses we know that the PR does not have under-coverage problems. This is justified because of all the rights that every person receive once they are registered (access to public education, public health system...). Over-coverage problems can be tackled focusing in those records with less amount of information from external data sources and if it is the case to carry out a very specific survey.
  According to coverage, in some types of dwellings (foreign people, young mobile people, homeless, linguistic minorities, people that live in private residential areas…) it is more and more difficult to achieve an answer using traditional techniques and this causes bias in the information.

- Administrative data open the door to new possibilities. For instance, new variables could be added without considering the physical constraint of the paper questionnaire and the questions that households can reply (health, disabilities, labour…).

[10] Although based on the Spanish situation, they can be applied somehow to all register-based censuses.
[11] At prices in 2001 (an estimation of the cost at 2016 prices and labour costs would be between 400 and 500 million euros)

- Census-like operations could be carried out more often, without having to wait 10 years until new updates are available, which is very important to make decisions.

- Improvement of existing social surveys. Several questions that are included in most of the surveys could then be deleted from questionnaires and be obtained directly from the census file in a homogeneous way. Besides, it enables further research by linking the Census information with other types of sources.

It should also be outlined that a register-based approach does not necessary provide the same type of information than a traditional census. A very illustrative example could be to compare figures of unemployed people obtained directly from registers and those ones calculated using a traditional questionnaire.

There is also information gathered in previous censuses that surely cannot be found in administrative records, because it does not exist or because its elaboration is too complex and of low quality (for example, concerning the equipment of some houses such as the availability of a toilet room and bathroom or shower, heating, Internet access, water supply system, number of rooms, elevator in the building, Gas, telephone line, central hot water and sewage disposal).

For some of them (such as the availability of a toilet room), it is possible to assume that they are present in 100% of dwellings or buildings. In general, it does not seem too risky to stop investigating these variables (the percentages collected in the 2011 Census were close to 99.5% in almost all territorial areas).

6. Conclusions

It takes a long period of time (even decades) to move from a traditional to a register-based Census. Anyway, although a totally perfect solution is impossible, it is worthy making the effort of this methodological change due to the numberless advantages that it entails. These advantages concern not only to the Census itself but also to the statistical system as a whole.

However, an ambitious project like this can only succeed if a set of circumstances are fulfilled: availability of human and technical resources, develop of record-linkage techniques to match information, legal and efficient access to data sources that belong to different organizations and management support in the project.

Spain, after more than 20 years of continuous work, is ready to join this small list of countries and complete its first register-based Census in 2021. And perhaps it will even be the largest register-based census in the next world round.