

# 2021 population census and migration statistics in Spain. Why exchanging microdata?

Mr. Antonio ARGÜESO

*Director, Socio-Demographic Statistics. INE (Spain)*

**Abstract.** In this paper we briefly present some key elements and recent developments of the 2021 population and housing census in Spain, which will be very largely based on administrative data. The availability of many new sources is already helping to improve population and migration figures today. However, some of the limitations we face in trying to calculate population figures using only national sources could be reduced by exchanging microdata for statistical purposes with other NSIs. Concrete examples are shown to illustrate it.

## 1. “Padrón”, the Spanish population register, is different...

Unlike other countries, where the Police, the Tax Collection Agency or other bodies are in charge of the population register (PR), in the case of Spain INE is the national institution that coordinates this database, called “*Padrón*”.

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

According to the Law, INE is responsible for the coordination and integration of all these municipal PRs into a single national population database. INE receives every month all the changes produced in every municipality, performs validations and forwards these results to all the municipalities, to avoid duplications, but also to include deaths, births or acquisitions of Spanish citizenship that INE receives on a monthly basis (or more frequently) from the Civil Register. This system has been working since 1996 and continuous improvements have been made during the last years, mainly as regards updating and de-registration procedures.

People living, or willing to live, in a municipality of Spain have the right (this is actually an obligation) to be registered in their municipality of residence. The only thing that they have to do is going to the municipality offices, filling a declaration and providing a proof of residence. An identity card or a passport together with a water/electricity/telephone receipt where the address is

shown is enough. The benefits of doing so are many: free access to public health system, public schools, and other services and rights are granted by the national or regional government; also the municipality provides other services like parking license among many others.

All residents in legal situation must have an Identity Card number that is stored in this database. In the case of foreign people in irregular situation, as they do not have this number, the passport number is recorded.

Managing an administrative register like this is not a common situation for NSIs. As statistical office, INE is a (special, but only a) user of this population register for statistical purposes but, on its role as manager of this administrative register, it has full access to it. Actually, the Directorate for Socio-Demographic Statistics has only a limited access to the register.

Because of this double role of INE, clearly defined by Law (statistical office on the one hand and manager of the population register on the other) there is a different administrative unit within INE coordinating the register, the “*Padrón*” Unit.

## **2. 2021 Population Census**

As regards population statistics, the main source in Spain is this population register. Nevertheless, one of the main drawbacks of *Padrón* is that it includes, for each person, only a few variables: sex, date of birth, place of birth (country, in the case of foreigners), citizenship, educational attainment and passport or national identity number. All previous places of residence (within Spain) are also stored in the database, thus allowing longitudinal analysis and a very close and precise monitoring of internal migration. The census 2011 was a combination between *Padrón* and a survey to 10% of population, which was needed to get the rest of census variables.

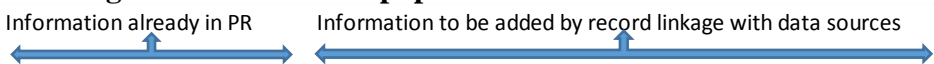
In 2014, once the 2011 Census production was over, we started the analysis of information available to build these census variables for 2021 from administrative sources instead of using a survey.

Since then, we have been getting access to many different administrative registers. The analysis of sources is not immediate because there are many variables collected in a population census, which is not only producing information on population but also on buildings, households and dwellings. Some of the variables are very easy to obtain from registers, 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 with *Padrón* data coming from other sources available to produce census-like information about the population. The goal today is to make a first pre-census file at the beginning of 2018 with all the available sources at 1st January 2016 as reference date.

The population census' final product (see figure 1) can be considered as a matrix of (approx.) 47 million rows (people) with a few columns already in the PR and many other additional columns (the rest of variables) obtained from the integration of other sources.

**Figure 1. Scheme of a population census file**



The diagram illustrates the structure of a population census file as a matrix. It consists of 11 columns and 12 rows. The first four columns are labeled 'identification', 'sex', 'age', and 'nationality'. The remaining seven columns are empty. Above the matrix, two blue arrows indicate the source of the data: a double-headed arrow labeled 'Information already in PR' spans the first four columns, and a single-headed arrow labeled 'Information to be added by record linkage with data sources' points to the remaining seven columns. The first four rows are numbered 1, 2, 3, and 4. The last row is labeled '47M'. A large black arrow points from the 'nationality' column to the first empty column, indicating the integration of data from other sources.

	identification	sex	age	nationality						
1		...	...	...	...	...	...	...	...	...
2		...	...	...	...	...	...	...	...	...
3		...	...	...	...	...	...	...	...	...
4		...	...	...	...	...	...	...	...	...
*		...	...	...	...	...	...	...	...	...
*		...	...	...	...	...	...	...	...	...
*		...	...	...	...	...	...	...	...	...
*		...	...	...	...	...	...	...	...	...
*		...	...	...	...	...	...	...	...	...
*		...	...	...	...	...	...	...	...	...
*		...	...	...	...	...	...	...	...	...
47M		...	...	...	...	...	...	...	...	...

One of the main difficulties faced by any census is to be able to accurately count the resident population. INE already produces annual population and migration figures from *Padrón*. These figures are not obtained by simply counting the number of registered people. Some statistical processes are applied to raw data in order to better measuring the number of residents.

On the one hand there are administrative procedures for de-registration in *Padrón*. Since 2005, non-EU nationals without permanent residence permit have to renew their registration every two years. Otherwise their inscriptions expire and they are not counted in the population figures (they must be notified from authorities when the registration is about to expire and the administrative procedure that eventually would lead to de-registration may take some months).

In the case of EU citizens and non-EU with permanent residence permit, since 2009 a procedure of confirmation of residence applies. Municipalities must check the residence for all these people at least once every five years (it is only two years in some particular cases). When they are not found, their registration expire as well, following an administrative procedure that is similar to the

one of non-EU citizens. Roughly, 80% of the de-registrations are made by this administrative process and the remaining 20% by citizens communicating their departure to the authorities.

In order to produce statistical figures some assumptions are made as regards presence in Spain of foreign nationals whose registrations are expired or about to expire. On the other hand, population figures are statistically corrected for better approaching to the usual resident definition, for instance, applying the twelve-month residence concept. In short, population figures are obtained from *Padrón* but they are not exactly the result of counting the registered population.

The strategy for counting population in the 2021 census will improve this method applying, as other countries are already doing, what is known as "signs of residence" or "signs of life": a statistical mechanism, for example, a logistic regression, which allows us to decide whether a person should be counted or not as resident according to their presence in different administrative records.

In the case of Spaniards it is relatively easy to count people using this approach. Although the linkage level between *Padrón* and the admin data is very high, for foreigners it is not so good.

### 3. Foreign residents in Spain

Looking at the foreign population in Spain (table 1), these are the most recent (still provisional) figures:

**Table 1. Foreign Population by citizenship  
(January 1st, 2017)**

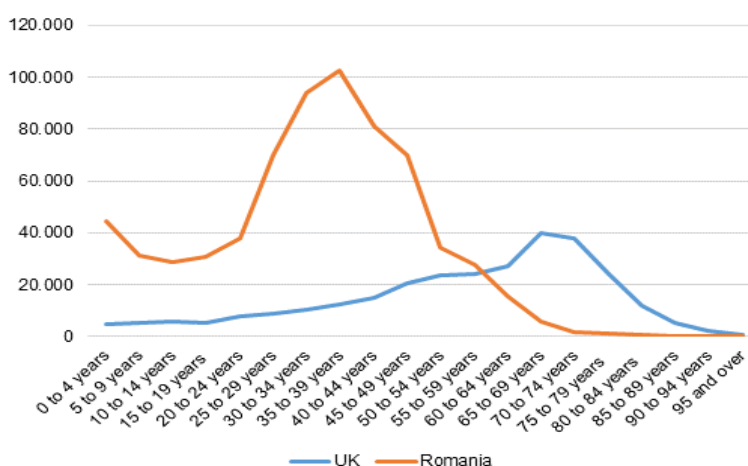
Total Population	46,528,966
Foreign Population	4,424,409
Romania	678,098
Morocco	667,189
United Kingdom	294,295
Italy	203,118
China	177,738
Ecuador	145,879
Germany	141,523
Colombia	139,213
Bulgaria	126,436
France	103,062
Portugal	100,822
Ukraine	94,770
Bolivia	76,060
Russia	71,959
Argentina	71,622

Source: Population Figures. INE

Considering only the EU, the two main groups are citizens from Romania and UK.

The reasons for migrating into Spain for these two nationalities are somehow implicit in their distribution by age (see figure 2). Romanians in Spain are mainly concentrated in working ages while Britons are mainly older people.

**Figure 2. Romanian and British Nationals in Spain by age**



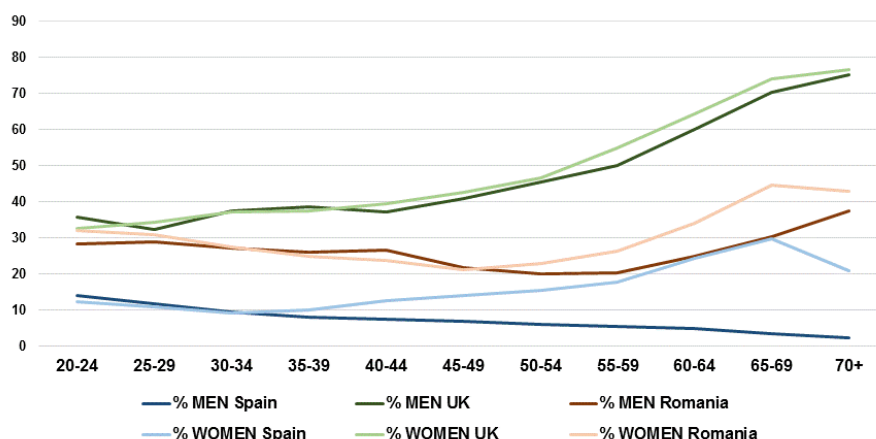
Source: Population Figures. INE

A very interesting example of the effects of linking *Padrón* with administrative data can be found in the variable “Economic Activity Status”. As a part of the current developments for the census 2021, a first test has been carried out linking *Padrón* at January 1<sup>st</sup>, 2016, with many administrative sources (around 10) providing information on employment, unemployment, educational system, pensions, etc.

The data below correspond to a NUTS-3 region (Province of Alicante, 1.84 Million inhabitants). This province has been selected because it shows the highest percentage of foreign population in Spain, with many EU citizens both as working migrants and retired people (sometimes referred to as *sun-seekers*).

Figure 3 shows the percentage of people not found in any register related to economic activity status in Alicante.

**Figure 3. Economic Activity Status by age. Unclassified residents (%)**  
(Without any information in administrative activity-related registers)



Source: census 2021 project. Pre-census file at January 1st 2016

In the case of Spanish nationals, it can be observed that the quality of the linkage is very good. People not found in unemployment records, neither in the educational system as students nor as employed in Social Security system, is around 5% for central ages. In the case of women it increases for later ages.

It can be considered that this “unclassified population” are mainly inactive people and not finding them in those registers can be logical to some extent. Actually, if we include this unclassified population into inactive population and compare it with the LFS, the results are quite consistent for Spanish population. The increasing percentage of unclassified women by age is particularly consistent with the figures of inactive women by age in Spain.

In contrast, the percentage of foreign residents according to *Padrón* that are not found in any register is much higher. Considering Romanians and Britons and comparing them to Spanish nationals, we find that less foreign people are found in the administrative data. What are the reasons? Are these foreign people in central ages (say 30 to 60) more likely concentrated in the informal economy and consequently out of reach using these administrative data? Have (at least part of) these people left the country?

### **Romanians in the labour market**

Looking at these figures we could consider that the procedures used for de-registering or not counting people in *Padrón* may be overestimating the number of Romanians in Spain.

If we had access to a Romanian administrative source we could detect, at least partially, those people that are registered as residents in the Spanish *Padrón* but also, for example, recorded as employees in the Social Security in Romania. Is it too bold to think of this kind of exchange for statistical purposes?

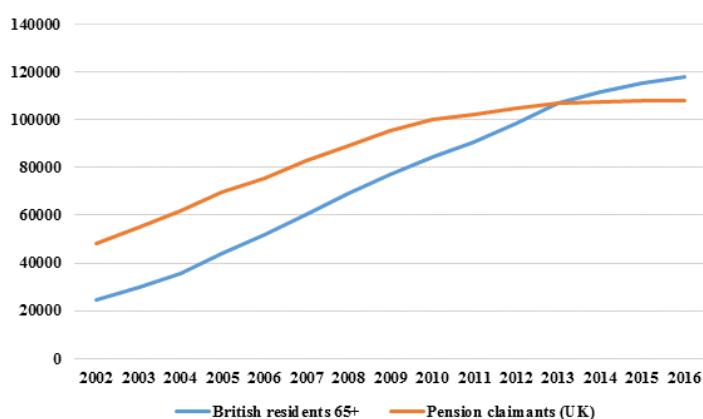
### British Nationals in Spain and their impossible death rates

As for Romanians, in the case of Britons the same doubts about people in working ages arise. There are 33,000 Britons aged 20 to 60 registered as residents in Alicante and the number of unassigned records (30-50%) looks very high.

It seems logical that people aged 65 and over cannot be assigned any economic activity status since many of them might be receiving a pension from UK and consequently cannot be found in Spanish pensions or benefits administrative sources. But, how many retired people are actually living in Spain most of the year?

Actually we can make a rough comparison between British pensioners living in Spain and British people registered in *Padrón* (see figure 4). On the one hand we have UK State Pensions claimed by British citizens living in Spain published by The Department for Work and Pensions (DWP)<sup>1</sup> of UK. It shows that 108,442 UK State Pension claimants have an address in Spain in May 2016. The second series is British Nationals aged 65 and over according to the Spanish Population Register (*Padrón*).

**Figure 4. Pension claimants (UK) vs British residents aged 65 and over**



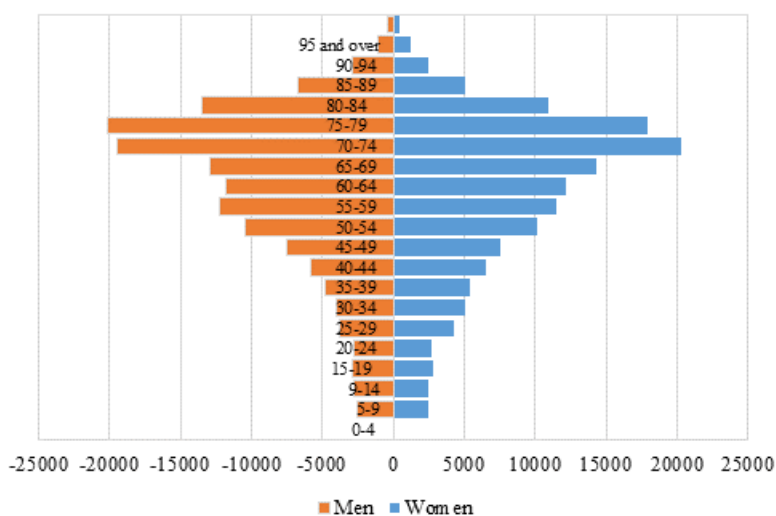
Sources: UK: Department of Work and Pensions (DWP)  
Spain: Population Figures (INE)

<sup>1</sup> Available from: <https://stat-xplore.dwp.gov.uk/webapi/jsf/login.xhtml>

In general terms we can conclude that both series are rather consistent. But figures on UK nationals in Spain could be improved if these British administrative data were linked with *Padrón*. For example, we could detect pensioners going back to UK. These data would never be used for de-registering people (as this is an administrative process with its own rules) but to improve the statistics.

Again, looking at the data on deaths it is easy to see that something must be wrong.

**Figure 5. Population pyramid of British residents in Spain (January 2017)**



Source: Population Figures. INE

The number of deaths of UK Nationals (registered as residents in Spain) was 3,078 in 2016. The crude mortality rate obtained for UK Nationals (10.5 per thousand) is very small compared to the one that could be expected from a very old population like this.

The observed death rate for age groups older than 70 for UK Nationals is roughly only one third the one of Spanish nationals which is impossible. It means that either most of the deaths in this age group happened outside Spain (the numerator is wrong, deaths of residents are under-counted) or most of the people are no longer living in Spain (the denominator is wrong and residents are over-counted) or very probably both of them.

Consequently there is some uncertainty about the foreign population that cannot be resolved using only national sources.



#### 4. Spanish Nationals in Germany

We can find another very interesting example in Spanish residents abroad. How many skilled young Spaniards are leaving the country to find opportunities in Germany or UK? It has been an important matter of public debate in Spain in the recent years.

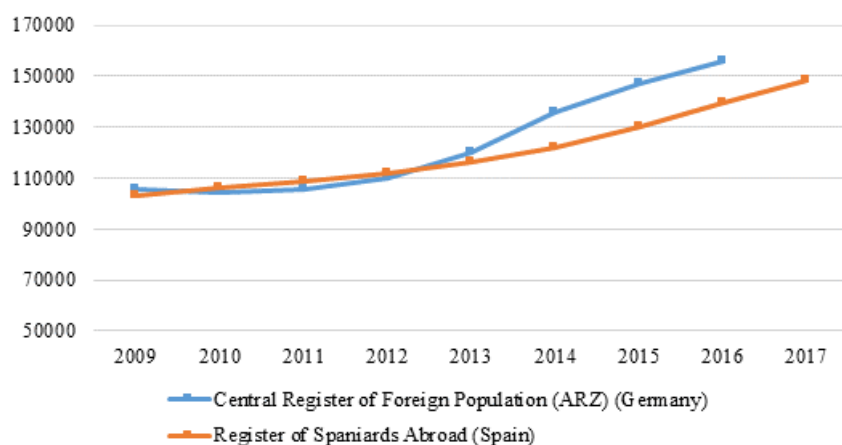
All consular offices of Spain throughout the world (around 250) are connected to *Padrón*. If a Spanish national from Barcelona moves to live in Berlin and he/she declares his/her new address in the closest consular office (there are 6 in Germany), this consular office sends this information to INE and INE forwards it into the monthly file to Barcelona so that automatically this person will be erased from the list of residents in Barcelona and included in the one of the consular Office in Berlin.

But there are not many benefits for Spanish people being included in this lists of residents abroad (the main one is having the right to vote) while there are some drawbacks when being de-registered as living in Spain, like losing the official consideration of resident in Spain. People might be afraid, for example, of the risk of not being properly attended by the Public Health system in case of need during holidays in Spain.

Moreover, the registration in consular offices must be signed in person and consequently the distance to the closest consular office is also an important element that plays against registration in most of the countries.

So we could conclude that the figures of Spanish citizens living abroad issued by INE might under-estimate the phenomenon. Let us compare the figures from German statistical Office (Destatis) based on the Central Register of Foreign Population in Germany named *Ausländerzentralregisters* (ARZ), with the Spanish Register of Spaniards Abroad.

**Figure 6. Spanish citizens in Germany (German vs Spanish sources)**



It seems that Spanish and German data were very similar (and stable) until 2012. Since then, an increase in Spanish Nationals living in Germany was detected in German Sources but not so much in the Spanish registers. The difference is 16,000 residents as of January 1<sup>st</sup> 2016.

These data would suggest that INE is very likely to be underestimating Spaniards abroad (at least in Germany). If we had access to these registers we could improve the population figures in Spain. But we could also link these data with administrative sources in Spain to picture these migrants statistically: their previous jobs in Spain, their educational attainment... This is beneficial for the European statistical system in general.

Circular migration, another field of growing interest would also be measured much better exchanging information. We (Spain and Germany) really do not know today to what extent young Spaniards move to Germany, spend two or three years there, and eventually come back. This kind of information could be bi-directional and beneficial for both sides.

## 5. Conclusions and some final remarks

Evidently, administrative data provide much more and better information today than ever before. Population figures would be undoubtedly better than the ones that could be obtained with a classic population count, based on visits to households and census questionnaires.

Using statistical procedures we can try to obtain a better approach to population figures on stocks and migration flows. But a further step can be taken in the quality of this information if we could use not only national sources in determining the number of foreigners and even nationals residing actually in Spain.

The proposal of linking foreign administrative data with national sources may sound very futuristic, even naive. There are legal, administrative, and technical barriers, of course. But it is also a matter of trust between NSIs that are ruled by similar legal provisions in terms of confidentiality and share common targets and values as members of the ESSC.

In our opinion there is room for exploring the cross-border exchange of administrative registers for statistical purposes. At least it would be helpful to identify technical limitations like the difficulties found linking without common identifiers, among many others.

This document is a first approach to the subject. Many details of the methodology of current demographic statistics in Spain, and about the possible dynamics of exchange have deliberately not been addressed here. For example, could these exchanges be implemented with Eurostat in between? Would it make things easier?

INE is ready to start a limited pilot research with other countries interested within the EU. The (very strict) Spanish Data Protection Agency has been consulted and, in principle, it poses no objections on these exchanges, provided that they are only used for statistical purposes.

This proposal is not bounded to countries having a population register but to those having non-anonymised administrative files that could be used for producing information on migration or population statistics.