

10 October 2019

United Nations Economic Commission for Europe

Conference of European Statisticians

Work Session on Migration Statistics

Geneva, Switzerland

29-31 October 2019

Item 3 of the provisional agenda

Data integration for measuring migration

Integration of population base statistical register with international migration statistical register

Note by National Administrative Department of Statistic – DANE, Colombia

Summary

In 2018, Colombia through the Statistical Institute of Colombia - DANE, carried out the National Population and Housing Census where people were asked about their place of residence 5 years ago and 12 months ago. These questions allow obtaining internal migration and international immigration at the sub-national level by age and sex. However, by the very nature from these it is not possible to determine international migration. On the other hand, the census includes the question for women over 10 who have children abroad, that makes possible to establish approximate of international emigration, but it has biases related to the lack of information of women who emigrated with their children and deceased women. This information is inconclusive.

This is how the idea of joining two research exercises arises, the first aimed at the construction of a Population Base Statistical Register from administrative records available in the country and the second for the construction of a statistical register of international migration. From the integration of the two registers, the location at the sub-national level of international emigration can be determined.

In Colombia, there is no official statistical base population register, and that is why an exercise like the one proposed is so important for the country's interests in population statistics.

Keywords: use of administrative registers, migration, integration of records, sub-national location of emigrants

I. Introduction

1. Throughout the article, the construction of the Population Base Statistical Register - PBSR - and the transformation of the administrative register of international entrances and exits of the Special Administrative Migration Unit Colombia – UAEMC- to determine international migration, will be shown. The results of the crossing of these two registries are described and how from this, it is possible to determine the origin of the international emigration.
2. The Population Base Statistical Register is composed by several administrative registries in the areas of health, education, vital statistics, tax registration and victim registration, among others. These registers allow to know the location in the Colombian territory and therefore it is possible to determine the origin of the emigration at this level.

II. Population Base Statistical Register (DANE, 2019)

3. The steps to form the PBSR are described below:
 - i. Make an inventory of administrative records.
 - ii. Define the variables.
 - iii. Standardize and validate administrative records.
 - iv. Integrate and improve administrative records.

A. Inventory of administrative records.

4. The administrative records available are:

Table 1

Inventory of administrative records that make up the PBSR

Area	Registers
Health	BDUA- Single Database of Health Affiliates – Ministry of Health
	EPS Health Promoting Entities
	SISBEN Beneficiary Selection System for Social Programs – National Planning Department
Education	SIMAT - Integrated Educational Enrollment System – Ministry of Education
	SNIES – National Higher Education Information System
	SIET – Work Education Information System
Births and Deaths	Vital Statistics – DANE
Vaccination	PAI - Expanded Immunization Program – Ministry of Health
Social Security Contributions	PILA Integrated Payments Settlement Form – Ministry of Health
Victims of the armed conflict	RUV – Single Register of Victims - Unit for Victims
Tax registration	RUT – Single Tax Register

B. Variable Definition

5. The variables that must be included in the PBSR must allow to identify, locate and determine demographic characteristics of the population, in this way the PBSR must be made up of the following variables.

Table 2
PBSR variables

Variable type	Variable
Identification	Identification Document number
	Type of document
	Complete name
	Birth date
	Sex
Demographics	Birth place
	Civil status
	Kinship
	Date of death
	Age
Location	Department
	Municipality
	Address
	Phone number
	Place of residence
Register	Status
	Provider
	Consistency
	Duplicate marking

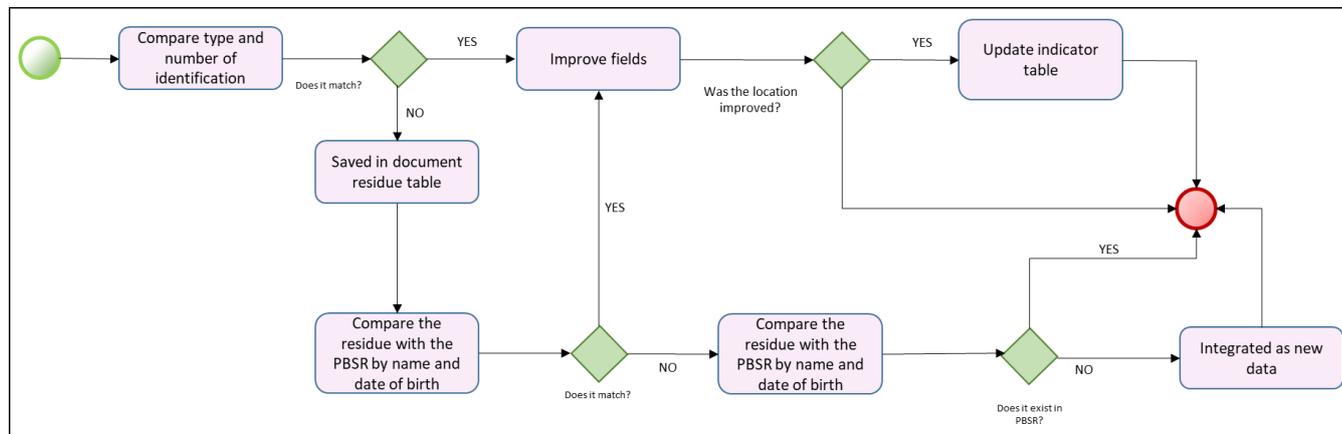
C. Standardization and validation of administrative records

6. Standardization: Adapt the definitions of the variables of each administrative record to the standardized definitions of the PBSR.
7. Validation:
- Inconsistency Analysis: determine if the information in each record is consistent with the person's information
 - Debugging and imputation: amend incorrect information and impute missing information
 - Duplicity analysis: identify duplicates to debug the register

D. Integration and improvement of administrative records

8. The following diagram shows the process by which the integration of administrative records is performed:

Illustration 1
Integration of administrative records



9. The record that contains the most information is the Unique Database of Health Affiliates, which includes all the people who, at some time, were or are affiliated with the Colombian health system. In total, this administrative record contains over 47 million records between active and suspended people. Inactive records are those that are not taken into account within the Colombian population or people who have a death mark or are duplicated.
10. Once all administrative records are integrated, a little more than 48 million records are obtained in the PBSR.

III. International Migration Statistical Register

11. This register is the result of the statistical transformation of the administrative register of international entrances and exits that is administered by the Special Administrative Unit Migration Colombia, a methodological proposal of the professor Joaquín Recaño (Recaño Valverde, 2011).
12. This register contains the information of all the entrances and exits of the country from January 1st, 2004 until December 31st, 2018. However, it does not contain the residence of the people at the sub-national level. Applying the definitions of the United Nations, of long term migrants and habitual residents; it is possible to determine the amount of international migrants between the years 2005 and 2018¹.
13. The register contains a very precise reference of the dates of entry and exit, and secondly it allows to use equivalent definitions for the flows of entry and exit of the country using exactly the same register in both cases.
14. The most important variables in the administrative record are shown below.

¹ https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.10/2016/mtg2_WS/4_Colombia_Copete_Martinez_ENG.pdf

Table 3
Variables of International Migration Statistical Register

Variable Type	Description of Information
Individual identifier	Type of document
	Id number
Administrative	Document type
	Country issuing passport
	Check Point
	Means of transport
Flow type travel	I: income
	E: outcome
Temporality	Date of trip
Demographics	Sex
	Birthday
	Country of birth
	Traveler's nationality
Geographic	Residence country
	Country Destination / Origin

15. The steps for transforming administrative records into statistical records, are shown below. Each of them is explained with the pertinent comments regarding the scope of the estimate:
16. Step 1. Trips are organized by anonymous identifier in alphabetical order and travel date from the furthest to the most recent date. This allows you to group the travel information of the same person and perform the subsequent steps.
17. Step 2. The number of trips per person is counted. First, those who have less than 30 trips in total will be taken into account for the estimate. It is assumed that those travellers with 31 and more trips form the population that travels frequently, and for which reason it is especially difficult to establish immigration status and habitual residence. Such group represents 28.5% of trips and only 3% of people. From this point, travellers who have 30 trips or less: 71.5% of trips, will be taken into account.
18. Step 3. Consistency analysis of the travel series. For each traveller, the logical principle followed is that to each exit (E) must correspond an entry (I) and vice versa. That is, to re-enter or exit, a traveller must have previously left or entered, not having entered or left twice in a row. Cases with inconsistent sequences are discarded. These illogical sequences are present in 11.8% of trips.
19. Step 4. Definition of initial residence status. Counting the number of days from the start of the registration (January 1, 2004) until the date of the first trip. If from the beginning of the registration the traveller has more than 365 days for his first trip, the initial residence of the same is established as Colombia (if the first trip was of departure) or abroad (if the first registration was of entry).
20. The conceptual definition that is used and applied here, in which a one-year period of stay is required to certify migration, means that since the database is temporarily truncated to the left (January 1, 2004), it is not possible to know the initial residence status for the movement of the initial year 2004: for example, an exit registered on that first day of the database could be preceded by a previous entry at any time of the year 2003, in previous years, or be the first. This implies that the residency status of 2004 is not detected, which should consider the integrity of the register and the estimate only from the movement of 2005.

21. Step 5. Changes in residence status and consequently migration events. Counting the number of days between the first trip and the second. If this is greater than 365 (one year) a change of residence is determined. Therefore, the residence status is updated, or else the residence will remain the same as the initial residence. If there is a change of residence, a migration is constituted with the date of the movement that initiated the 365-day account.
22. Step 6. Repetition and iteration of step 5 for the other travel chains of each person (second and third, third and fourth, etc. etc.). This allows to identify successive migrations.
23. Step 7. Verification of residence status until the end of the measurement. Counting the number of days from the last trip until the cut-off date (December 31, 2018) contemplating this date as if it were the last trip of the person and thus update the residence status. If there is a new residence status, a new migration is registered, analogously to steps 5 and 6.
24. Similar to step 5, at the end of the base there is a truncation of information, only, this time to the right. If the most recent record is that of the 2018 movement, only the estimation of the migrations can be made until 2017; since to certify migrations of the final year 2018 the information of the trips of 2019 would be needed, by definition not available.
25. Step 8. Perform the counting of immigration and emigration events in the analysis period and with the level of temporal detail required. At this point the information processing ends, generating the result in the relationship between people and migration events with their respective dates and thus generating a statistical record.
26. The transformation into a statistical record shows that, of the 137'870.943 trips, 3'773.990 constitute migration. Out of these, 56% are emigrations and the remaining 44% are immigration.

IV. Integration of registers

27. The integration of the two registries: population-based statistical register and international migration statistical register, is made from the type and number of the identity document of the persons, that is, national id, passport, foreigner id, etc. Once integrated, a total of 1,577,350 records that are in the two records, are obtained. In effect, 987,688 are emigrants.
28. From the same definition and construction of PBSR, the location in the Colombian territory at the departmental level is contained within the register information; as mentioned before, the international migration statistical record does not have this information.

V. Limitations of the methodology

29. The PBSR, was constructed from different administrative records that contain different variables, these variables are not homogeneous throughout the register, it is the case of the place of residence. In particular for this exercise, the municipality of residence of the people is being assumed as the place where they receive their health service; It is assumed that, although the municipality of residence may be different from the municipality in which it is registered in the health system, the department is the same, so we can obtain the location at the departmental scale.
30. All administrative records contain information of the type and number of identity document, however they can be wrong. This is because at the time of capturing this information it was entered the wrong way. In order to match the records that do not match, a phonetic match of

the names and surnames of the people is made. This pairing could be done in a probabilistic way, but the technological capacity of DANE does not allow it to be carried out.

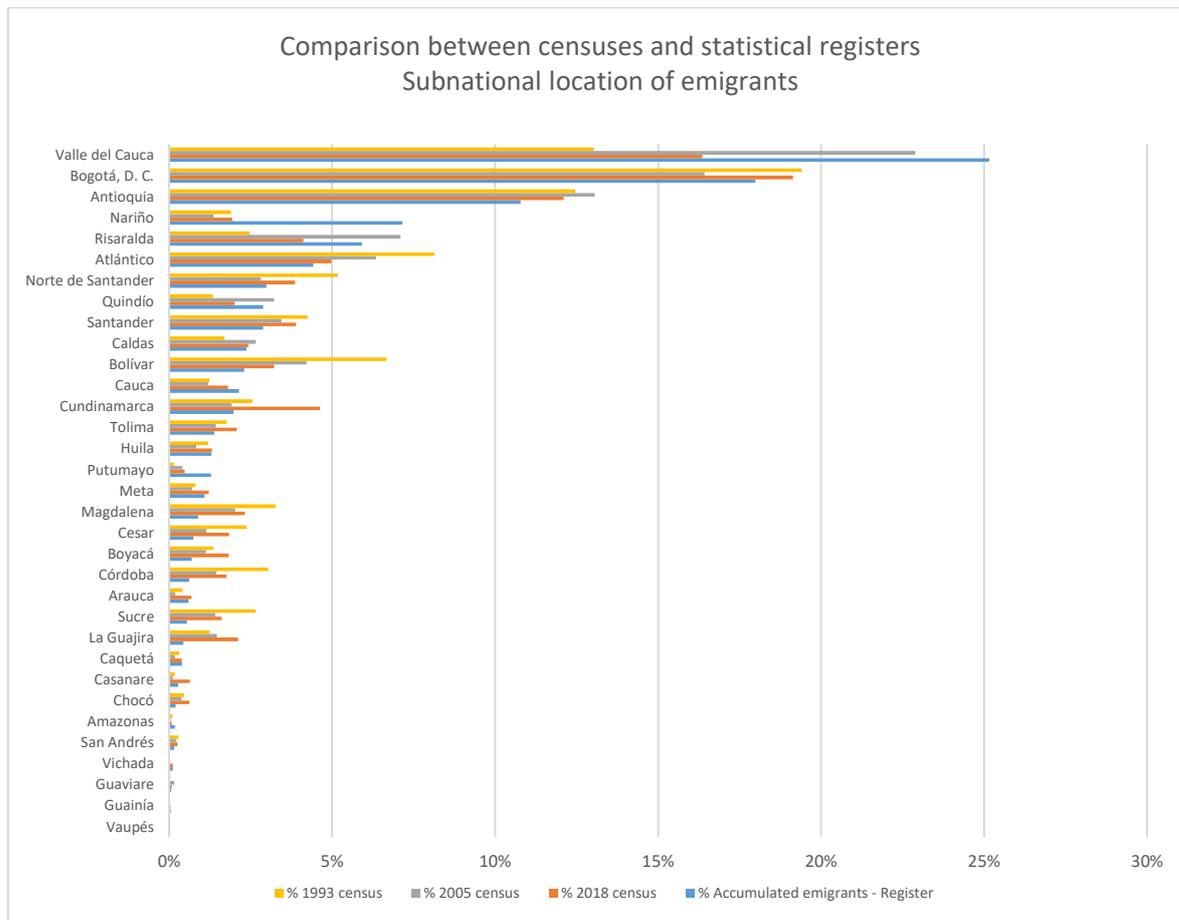
31. The border control register does not contain all the information of the crossings that can occur at the Colombian borders, in total there are 50 control points in the country. The bordering places where there is no point control and where border crossings happen, are not registered within the administrative record. Geographically, there are land border limits with five countries: Panama, Venezuela, Ecuador, Peru and Brazil. In each of them there are border control points, however, the borders are spaces of very extensive territory that prevent absolute control of travel. On the other hand, it was not possible to have the information of the document number in the administrative border control register of 2016².
32. In the population and housing censuses of 1993, 2005 and 2018, questions were asked about households with migratory experience. In the 1993 census, women over 12 were asked about surviving sons and daughters abroad. The 2005 census asked about people belonging to the household who lived abroad. Furthermore, in the 2018 census the question was again directed to women over 10 who had surviving children abroad. This information determines the territorial distribution of households with migratory experience. It is an approximate information of the place of origin of the migration since the source of information may have presented internal migration. In addition, there may be omission due to the death of the reporting mother or because the family/mother emigrated out of the country.

VI. Results

33. The objective is to obtain a distribution of international emigrants at a sub-national level. This document manages to determine that out of the total of international migrants present in the statistical register of international migration 3,775,783; 1,544,379 records (41%) were matched, of which 978,110 (63%) correspond to emigrations, and it was possible to establish that 958,663 (97%) have residence department information.
34. Many of the entries in the database do not have a document number or the document corresponds to the passport number that is not a unique number that can be found in the PBSR, so the coincidence between the population register and the migration register is less than half and many of which are foreigners. Despite of this, 97% of migrant information has information from the department.
35. The following graph shows the information obtained from the census by department (sub-national level) and the close relationship there is with the location obtained with the crossing of the two registers.

² The variable has already been requested from the Special Administrative Unit Migration Colombia

Illustration 2
Comparison between censuses and statistical registers Subnational location of emigrants



36. The departments that have emigrant tradition are very well reflected with the integration of the records, this relationship is shown in Table 4. There are some departments in which the origin of the migration seems unrelated. For example, in the department of Nariño, the census information shows that emigration is not so high but in the register it shows that it is the opposite (the fourth nationwide). This department is located on the border with Ecuador and there is a high flow of Venezuelan population that leaves the country towards the south of the continent.

Table 4
Correlation between censuses and the Statistical Registers

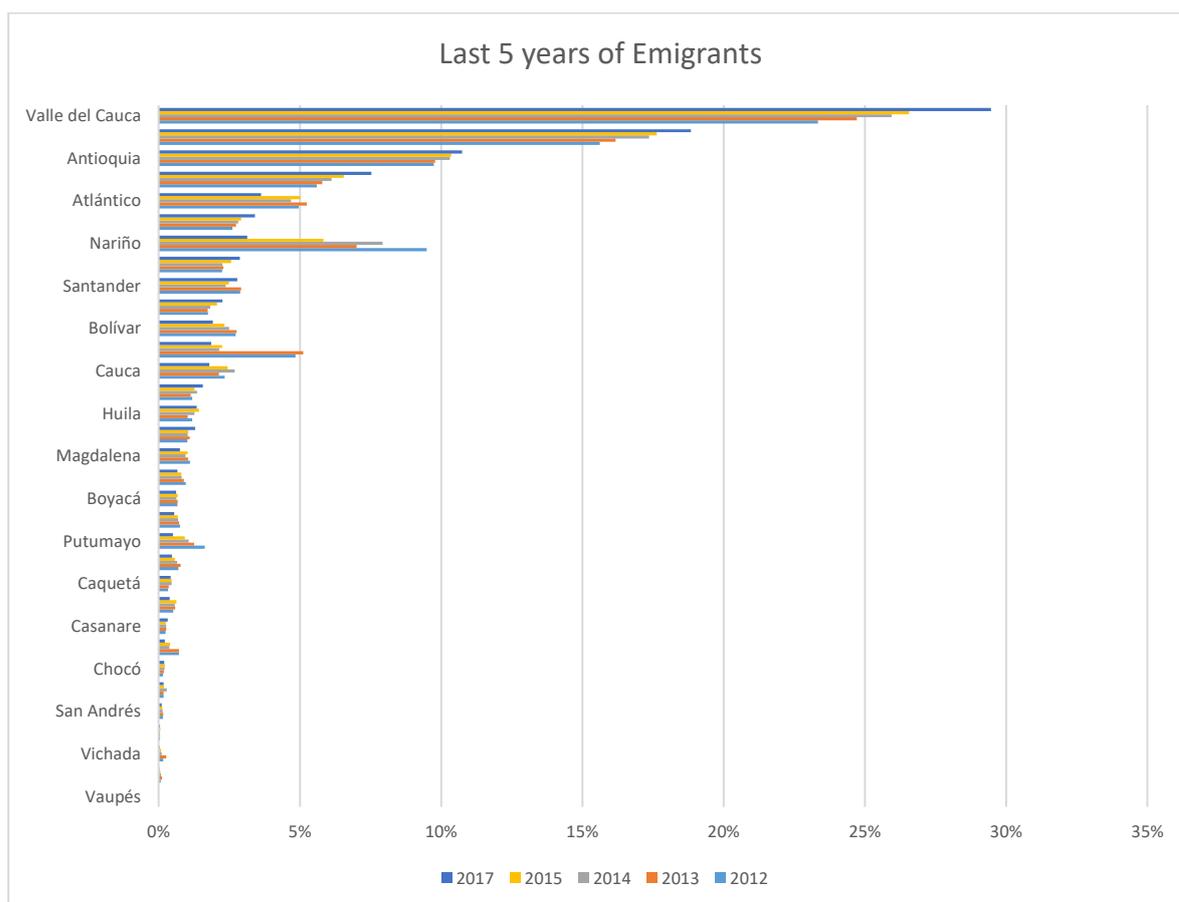
	Pearson correlation
1993 census	0.8518
2005 census	0.9662
2018 census	0.9321

37. Because the information from the 1993 census is so far away, the correlation is not as strong with the emigrant register due to changes in migration over time, that is, behavioural

changes in migration. The information in the register is accumulated from 2005 to 2017, so that the correlation with the censuses is very high.

38. The distribution of emigration over the years is concentrated in the same departments as shown in the previous graph. Evidently, it is a different distribution over the years, but it retains the pattern of the departments with the greatest presence of emigrants. The following graph shows the distribution of the last 5 years with the exception of 2016 for which DANE has not got information.

Illustration 3
Last 5 years of Emigrants



39. This finding is very important since Colombia did not have a way to establish the sub-national origin of international emigration and also because it succeeds in integrating the population register with the statistical migration register.

VII. Future work

40. The integration of population census information, household survey (migration module) and the international migration statistical register will allow to find social, economic and demographic characteristics of migrants. In addition, with the census questions (about the place of residence between the last 5 years and 12 months), it is possible to establish the migration history of people before leaving the country; and vice versa after entering the country in the case of immigrants.

41. From the integration it is also possible to determine the origin of the emigration at the municipal level, but caution should be taken with the distribution since it is assumed that the place of residence of the person is conditioned to the place of affiliation to their health system.

VIII. Conclusions

42. One of the objectives of this exercise was to integrate the population-based statistical register and the international migration statistical register based on the type and number of identity document. Of the total international migrants present in the 3,775,783 international migration statistical register, 1,544,379 records (41%) were matched, of which 978,110 (63%) correspond to emigrations.
43. The other objective, perhaps the most important one, was to determine the sub-national origin of international emigration. From the total number of emigrants present in both registries, 958,663 (97%) have information on the residence department.
44. The departments that have an emigrant tradition, are very well reflected with the integration of the records. The distribution of emigration measured from the statistical record of international migration over the years is concentrated in the same departments as the one measured based on census data.
45. This finding becomes one of the inputs of which there was not enough information, since the census data may have biased and omitted information. It is the first time that, in Colombia, it is possible to determine the origin of emigration at the sub-national level.

IX. Glossary

46. In this section the definitions of register, administrative register, population base register and population base statistical register are written³.
- a) *“Register:* Corresponds to a list of objects that seeks to contain the total elements of a specific population and individualized data on its main characteristics. Additionally, they allow unequivocal identification of each object in such a way that the record can be updated and expanded with new variables for each object.
 - b) *Administrative register:* The record used for administrative purposes in an administrative information system. It will contain the objects to be managed, including the identification and variables used for administrative purposes. A simple typology of data sources is used to consider how administrative sources should be defined. Firstly, all data sources are divided into primary sources (data collected for statistical purposes) and secondary sources (all other data).
 - c) *Population base register:* Refers to the register that establishes the foundations for basic demographic statistics, which are intended to describe the structure of the population and its changes over time. The important variables in a population base register are those that identify a person and can be used to link that person with other base registers.

³ Sections 5.2.1 to 5.2.5 consist largely of material taken verbatim from the UNECE handbook on administrative sources (UNECE 2011).

- d) *Statistical base population register*: This is created from the transformation of one or several administrative registers, in such a way that the objects and the variables are adjusted to satisfy the statistical needs. It is also possible to create a statistical record from the processing of administrative records together with other statistical records.

References

DANE - DIRPEN. (2010). *Informe final de evaluación de la calidad estadística del "Registro de entradas y salidas de"*. Bogotá: DANE.

DANE. (2007). *Aproximación a la migración internacional en Colombia a partir del Censo General 2005*. Bogotá: DANE.

DANE. (2008). "Estimación de la migración 1973-2005", *Estudios post-censales número 6*. Bogotá: DANE.

DANE. (2019). *Metodología General para la Construcción y Actualización del Registro Estadístico Base de Población – REBP*. Bogotá: DANE.

Martínez, J. (2009). "Medición e información sobre la migración internacional a partir de los censos: lecciones desafíos y oportunidades". En *Notas de Población N°88* (págs. 97-133).

Recaño Valverde, J. (2011). *La emigración exterior de Colombia. Informe técnico del proyecto n° 1121*. Bogotá: DANE.
