

A new database for studying the international migration in Colombia: methodology and comparative results

Joaquín Recaño Valverde, *Universitat Autònoma de Barcelona*

Myriam Cifuentes Noyes, *Departamento Administrativo Nacional de Estadística (DANE)*

Jacqueline Martínez García, *Departamento Administrativo Nacional de Estadística (DANE)*

Marta Roig, *United Nations*

Carolina Sánchez Barriga, *Departamento Administrativo Nacional de Estadística (DANE)*

Abstract:

International migration constitutes the weakest point with regards to demographic information. In a great deal of countries these do not have statistical records on international migration, and the data they use generally comes from censuses that are carried out in ten-year intervals which are unable to capture cyclical changes of international migration.

The purpose of this paper is to present the methodology and results of a new source of information on international migration flows in Colombia which is being developed by the *Departamento Administrativo Nacional de Estadística (DANE)* under the agreement with the *Unidad Administrativa Especial Migración Colombia (MIGRACIÓN COLOMBIA)*. This new source consists of individual records collected by the *MIGRACIÓN COLOMBIA* at border controls by air, land and sea. With this information, the DANE produces an estimate of the country's migratory inflows and outflows. The new database contains demographic information (sex, and date of birth), geographic information (country of nationality and birth, and country of residence), socioeconomic data (occupation, purpose of travel), and the legal status of travellers (type of visa). Due to the special geographical location of Colombia, nearly 90 percent of the entries and departures with origin or destination to other countries are carried out through air transport, which ensures a high degree of coverage of this source, especially when trying to identify those return movements that are not detected by the statistical systems of the countries of origin.

The data used in this study comes from the 41,683,460 international trips flows in the period 2004-2010. An anonymization process has undergone so as to allow the reconstruction of individual information on the travel itineraries of 10,461,355 people for the same time period. The development of the methodology and the reconstruction of the migration flows (comparing them with those registered in countries like Spain and Colombia) are some of the aspects discussed in this paper.

Data and methods:

The database of the Unidad Administrativa Especial Migración Colombia is a register of travelers / trip and the final goal of the *Departamento Administrativo Nacional de Estadística* (DANE) is to generate a record of migration and migrants with certain demographic and territorial characteristics;

When analyzing this data, an essential question arises: what do we define as migration and migrant? To accomplish this we use a spatial scale: we cross the border of Colombia and, on the other hand, a time scale resulting from the recommendations of the UN, which consists of the stay of over 1 year of residence abroad so as to consider an individual as a migrant. From the previous definition, the need to close the information on migration with one year delay in relation to the records of the MIGRACIÓN COLOMBIA is deduced.

Moreover, the complexity of the concept of residence abroad that involves estimating the duration of the episode and, therefore, the cumulative length of stay in Colombia and outside Colombia. We have already defined the "problem" or "object of study" that happens in this case from the reference object on the database of MIGRACIÓN COLOMBIA, the travel / traveler, to the people with those characteristics of residence and temporary residence that will permit to distinguish their character as migrant or non-migrant. The basis of information received by the DANE from the MIGRACIÓN COLOMBIA is a database subject to a prior process of anonymization: trips belonging to a "same" traveler are connected through an algorithm link that employs a number of information fields from the database, resulting in a univocal identification key for locating an individual with his/her trips.

One of the essential aspects for the construction of international migration statistics is the realization of a classification/typology of the number of episodes (trips) for each person registered in the MIGRACIÓN COLOMBIA' database. Therefore, we need to convert the trips' database to a travelers' database. In this way we can evaluate the consistency of the trips' database from the coherence of these routes: a departure is followed by an entry and vice versa. This type of consistency analysis is unprecedented in terms of the MIGRACIÓN COLOMBIA database given that it is a conceptual analysis from the perspective of migration studies and it is opposed to the mere

statistical analysis which pursues treatment and independent coherence of variables. In this case we consider the entire travel itinerary of the same person and their suitability or coherence.

The analysis of the logic sequencing of traveler itineraries aims to estimate potential errors in the register of entries and exits from a demographic and migratory perspective. On the other hand, it seeks to verify whether land entries and departures are the origin of some truncated sequences of the type I - I or E - E. To obtain this information it is necessary to transform the travel database (trips' database) from the MIGRACIÓN COLOMBIA to a travelers' database in which the n trips corresponding to an individual are transformed to a persons' register with n fields (each one made in reference to the trip undergone) as shown in Figure 1.

Figure 1: Transformation of the database from trips to travelers/people data

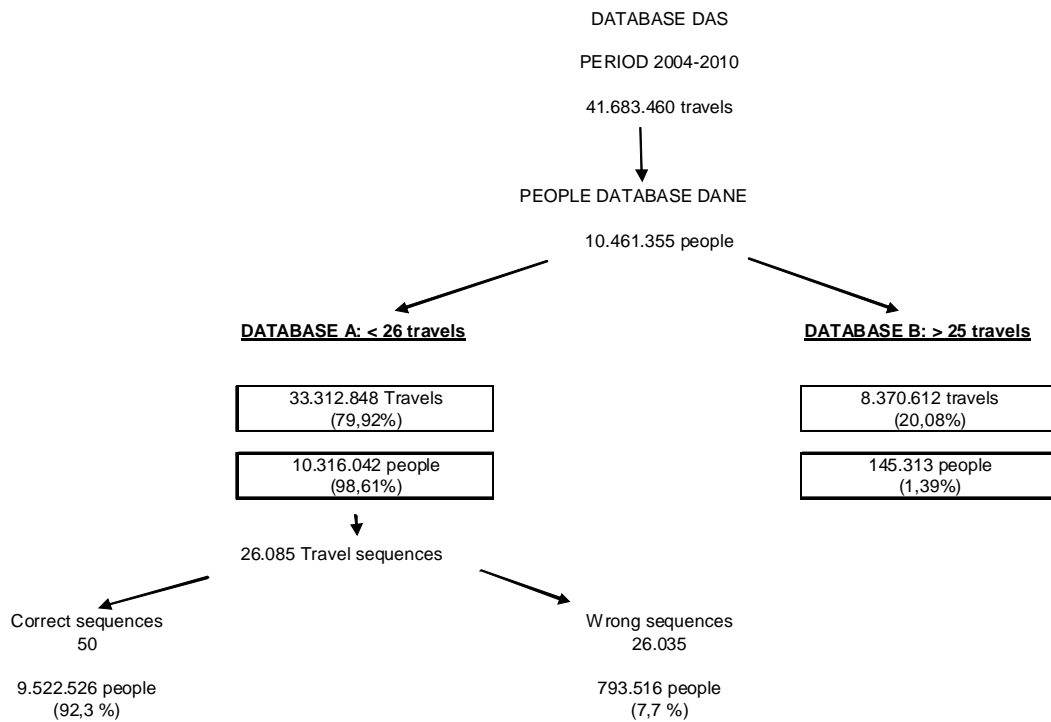


The persons' database is the one that DANE will use so as to carry out the estimation of international migrations, as it provides, as mentioned above, the variables that determine the migrant nature of people (see table 1).

The database of MIGRACIÓN COLOMBIA for the period 2004-2010 consists of 41,683,460 movements that correspond to 10,461,355 people (see Figure 3). Once the conversion of trips to travelers has been accomplished, people who have made more than 500 trips on this database of MIGRACIÓN COLOMBIA have been found which, probably, correspond to travelers who are not sufficiently identified in this database and constitute one of the MIGRACIÓN COLOMBIA registry problems. To

address an efficient analysis of this information the persons' DB has been divided into two files (Figure 2).

Figure 2: Distribution of sequences at the People Database of DANE



Source: DANE and Migración Colombia Database. Own Calculations.

Table 4: Correct sequences started with an emigration

Rank	Type of sequence	People	%	Trips	%
1	E	1.328.054	34,9%	1.328.054	10,9%
2	EI	1.202.025	31,6%	2.404.050	19,7%
3	EIE	207.630	5,5%	622.890	5,1%
4	EIEI	369.926	9,7%	1.479.704	12,1%
5	EIEIE	91.297	2,4%	456.485	3,7%
6	EIEIEI	176.521	4,6%	1.059.126	8,7%
7	EIEIEIE	49.140	1,3%	343.980	2,8%
8	EIEIEIEI	100.368	2,6%	802.944	6,6%
9	EIEIEIEIE	29.235	0,8%	263.115	2,2%
10	EIEIEIEIEI	61.984	1,6%	619.840	5,1%
11	EIEIEIEIEIE	18.490	0,5%	203.390	1,7%
12	EIEIEIEIEIEI	40.631	1,1%	487.572	4,0%
13	EIEIEIEIEIEIE	12.572	0,3%	163.436	1,3%
14	EIEIEIEIEIEIEI	27.756	0,7%	388.584	3,2%
15	EIEIEIEIEIEIEIE	8.506	0,2%	127.590	1,0%
16	EIEIEIEIEIEIEIEI	19.445	0,5%	311.120	2,5%
17	EIEIEIEIEIEIEIEIE	5.895	0,2%	100.215	0,8%
18	EIEIEIEIEIEIEIEIEI	14.140	0,4%	254.520	2,1%
19	EIEIEIEIEIEIEIEIEIE	4.359	0,1%	82.821	0,7%
20	EIEIEIEIEIEIEIEIEIEI	10.669	0,3%	213.380	1,7%
21	EIEIEIEIEIEIEIEIEIEIE	3.253	0,1%	68.313	0,6%
22	EIEIEIEIEIEIEIEIEIEIEI	8.238	0,2%	181.236	1,5%
23	EIEIEIEIEIEIEIEIEIEIEIE	2.484	0,1%	57.132	0,5%
24	EIEIEIEIEIEIEIEIEIEIEIEI	6.470	0,2%	155.280	1,3%
25	EIEIEIEIEIEIEIEIEIEIEIEIE	2.000	0,1%	50.000	0,4%
Total		3.801.088	100,0%	12.224.777	100,0%

Source: DANE and Migración Colombia Database. Own Calculations.

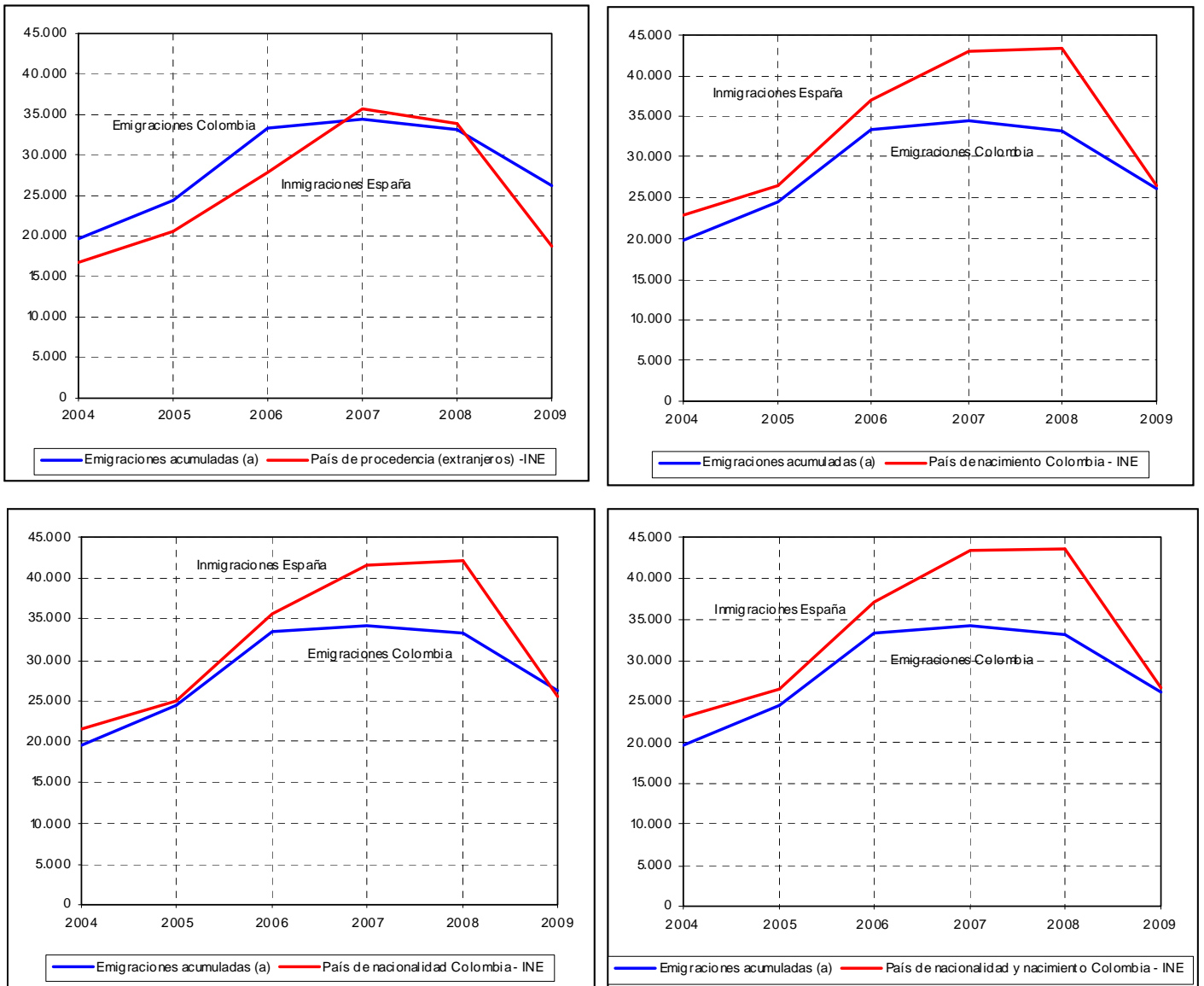
Table 5: Correct sequences started with an immigration

Rank	Type of sequence	People	%	Trips	%
1	I	1.130.579	19,8%	1.130.579	7,3%
2	IE	3.273.020	57,2%	6.546.040	42,0%
3	IEI	146.638	2,6%	439.914	2,8%
4	IEIE	574.050	10,0%	2.296.200	14,7%
5	IEIEI	62.297	1,1%	311.485	2,0%
6	IEIEIE	203.375	3,6%	1.220.250	7,8%
7	IEIEIEI	34.646	0,6%	242.522	1,6%
8	IEIEIEIE	96.391	1,7%	771.128	4,9%
9	IEIEIEIEI	21.247	0,4%	191.223	1,2%
10	IEIEIEIEIE	51.640	0,9%	516.400	3,3%
11	IEIEIEIEIEI	13.938	0,2%	153.318	1,0%
12	IEIEIEIEIEIE	30.282	0,5%	363.384	2,3%
13	IEIEIEIEIEIEI	9.387	0,2%	122.031	0,8%
14	IEIEIEIEIEIEIE	19.389	0,3%	271.446	1,7%
15	IEIEIEIEIEIEIEI	6.527	0,1%	97.905	0,6%
16	IEIEIEIEIEIEIEIE	12.324	0,2%	197.184	1,3%
17	IEIEIEIEIEIEIEIEI	4.588	0,1%	77.996	0,5%
18	IEIEIEIEIEIEIEIEIE	8.453	0,1%	152.154	1,0%
19	IEIEIEIEIEIEIEIEIEI	3.418	0,1%	64.942	0,4%
20	IEIEIEIEIEIEIEIEIEIE	5.884	0,1%	117.680	0,8%
21	IEIEIEIEIEIEIEIEIEIEI	2.557	0,0%	53.697	0,3%
22	IEIEIEIEIEIEIEIEIEIEIE	4.310	0,1%	94.820	0,6%
23	IEIEIEIEIEIEIEIEIEIEIEI	1.900	0,0%	43.700	0,3%
24	IEIEIEIEIEIEIEIEIEIEIEIE	3.149	0,1%	75.576	0,5%
25	IEIEIEIEIEIEIEIEIEIEIEIEI	1.449	0,0%	36.225	0,2%
Total		5.721.438	100,0%	15.587.799	100,0%

Source: DANE and Migración Colombia Database. Own Calculations.

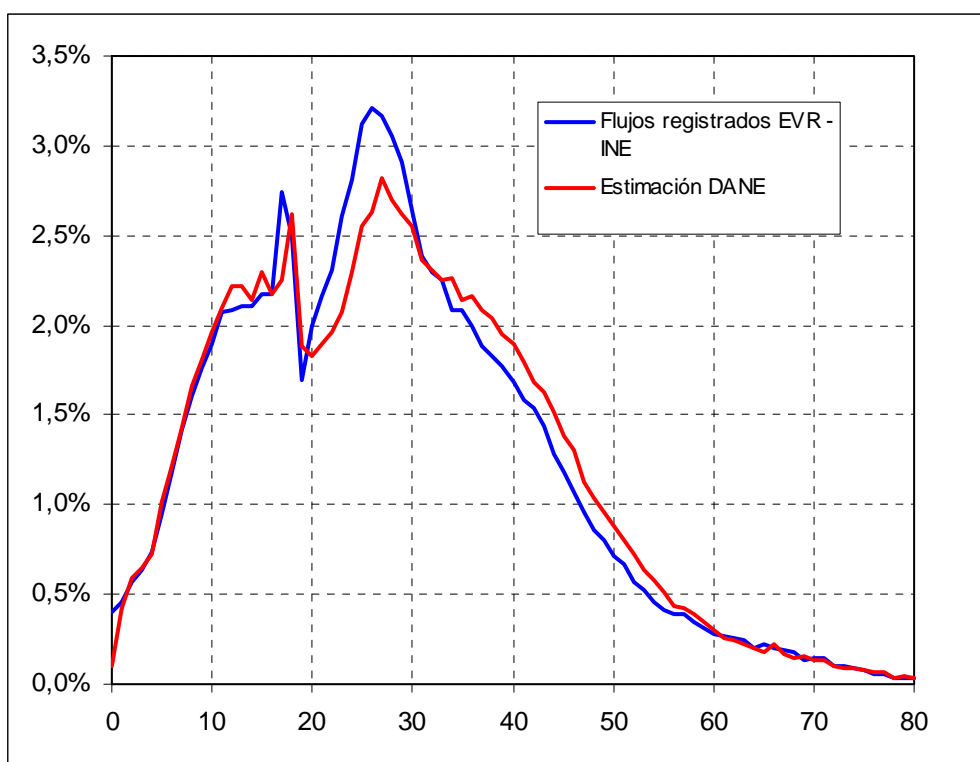
Preliminary Results

Figure 3
The estimation of emigration from Colombia to Spain with MIGRACIÓN COLOMBIA database and immigration data recorded in Spain (EVR)



Source Source: DANE and Migración Colombia Database. Own Calculations.
 and Micro-data EVR (Spain). Own Calculations.

Figure 4: Estimation of Colombian emigration to Spain by age and type of migratory sequence (rank <6) in the period 2004-2009



Source: DANE and Migración Colombia Database. Own Calculations.