UNFPA Population Data Fellowship

Subnational Differentials in Death Registration Completeness in Colombia, 2005 – 2018

Marco Fayet

June 2020

Abstract

We apply a hybrid of the Generalized Growth Balance method (GGB) and the Synthetic Extinct Generation method (SEG) to disaggregated census and mortality registry data, in order to estimate the rates of death registration completeness in Colombian departments. We find that the departments with the lowest rates of completeness tend to be the ones with the highest proportion of indigenous and afrodescendant populations.

Table of contents

I. Introduction 3

II. Methodology 3

a. Data 3

b. Death Distribution Methods 5

III. Results 5

a. Data Quality Assessment 6

b. Completeness estimates 8

IV. Limitation 11

V. Conclusion 12

VI. References 12

VII. Supporting Documentation 12

Figure 1. Extract of the analysis-ready dataframe 3

Figure 2. Age heaping in 2005 and 2018, national level 6

Figure 3. Sex ratios in 2005 and 2018, national level 6

Figure 4. Age ratios in 2005 and 2018, national level 6

Figure 5. Age heaping in 2005 and 2018, department level 7

Figure 6. Sex ratios in 2005 and 2018, department level 7

Figure 7. Age ratios in 2005 and 2018, department level 8

Figure 8. Sex differences in death registration completeness, department level 8

Figure 9. Comparison of death registration completeness estimates with Adair and Lopez, department level 9

Figure 10. Maps of death registration completeness, by gender and department 10

Figure 11. Age range sensitivity and its impact on results 11

Table 1. Description of variables 4

Table 2. Department names and code equivalencies 4

Table 3. Death registration completeness estimates and population counts, department level 10

I. Introduction

Death registration is an essential source of data for evidence-based policy making and public health surveillance. Estimating the rates of death registration completeness, and subsequently addressing them, are an integral part of strengthening Civil Registration and Vital Statistics (CRVS) systems.

The goal of this exercise is to apply demographic statistical methods in order to estimate death registration completeness in Colombia at the subnational (departmental) level.

This project is part of the ConVERGE initiative, jointly implemented by the United Nations Population Fund (UNFPA) and the Centre of Excellence for CRVS.

II. Methodology

a. Data

The input data consists of:

- 2005 census microdata (n = 41,468,384)
- 2018 census microdata (n = 44,164,417)
- 2005 to 2018 mortality registry data (n = 2,886,660)

The datasets were downloaded from the Colombian National Administrative Department of Statistics (DANE)'s website in the first quarter of 2020.

The variables of interest (population counts and intercensal mortality counts) were then extracted, loaded into RStudio, harmonized, and disaggregated by department, gender, and 5 year age-groups.

Although Colombia is comprised of 33 departments (including the Capital District of Bogota), the 2005 census "regionalizes" territorial units with less than 20,000 people for confidentiality reasons. As such, the Guaviare department also accounts for the departments of Amazonas, Guainía, Vaupes and Vichada, which are all located in south-eastern Colombia. This results in a department count of 29 for the purposes of this analysis. Multiplied by 2 genders and 17 age categories (0 to 80+ in 5 year increments), we obtain a **consolidated dataframe of 986 rows and 10 columns**, as illustrated below:

	dept_code1	dept_code2	dept_name	sex	age	pop1	pop2	deaths	date1	date2
1	1	5	Antioquia	m	0	261523	176817	8382	2005-05-22	2018-01-18
2	1	5	Antioquia	m	5	279146	195622	1063	2005-05-22	2018-01-18
3	1	5	Antioquia	m	10	291286	230481	1518	2005-05-22	2018-01-18
4	1	5	Antioquia	m	15	265600	254659	7549	2005-05-22	2018-01-18
5	1	5	Antioquia	m	20	233467	272044	12272	2005-05-22	2018-01-18

Fig 1. Extract of analysis-ready dataframe

The variable names and department notations are explicited in the tables below:

Variable name	Description
dept_code1	Department code (first notation)
dept_code2	Department code (second notation)
dept_name	Department name
sex	Sex (m = male; f = female)
age	Age
pop1	2005 population count for a given department, sex, and age group
pop2	2018 population count for a given department, sex, and age group
deaths	Intercensal (2005-2018) mortality count
date1	2005 census collection date
date2	2018 census collection date

Table 1: Description of variables

dept_code1	dept_code2	Department Name
01	05	Antioquia
02	08	Atlántico
03	11	Bogotá
04	13	Bolivar
05	15	Boyacá
06	17	Caldas
07	18	Caquetá
08	19	Cauca
09	20	Cesar
10	23	Cordoba
11	25	Cundinamarca
12	27	Choco
13	41	Huila
14	44	La guajira
15	47	Magdalena
16	50	Meta
17	52	Nariño
18	54	Norte de Santander
19	63	Quindío
20	66	Risaralda
21	68	Santander
22	70	Sucre
23	73	Tolima
24	76	Valle del Cauca
25	81	Arauca
26	85	Casanare
27	86	Putumayo
28	88	Archipiélago de San Andrés, Providencia y Santa Catalina
31	95	Guaviare, Amazonas, Guainía, Vaupés and Vichada

Table 2: Department names and code equivalencies

b. Death Distribution Methods

Death Distribution Methods (DDMs) compare the distribution of deaths by age groups with the age distribution of the living, and provide the age pattern of mortality in a defined reference period (UN DESA, 2002). The age distributions of the living and their evolution over time are typically extracted from two population censuses, while the age pattern of deaths during the intercensal period can be extracted from mortality registry data.

The General Growth Balance (GGB) and Synthetic Extinct Generations (SEG) are two DDMs that have been widely used to evaluate the coverage of registered deaths in developing countries.

While both methods rely on the following three major assumptions:

1) A closed population
2) Invariant coverage of population and deaths by age within, but not across sources, and
3) Accurate recording of age for both population and deaths.

The SEG method also relies on the following assumption:

4) Invariant coverage of population across time.

When there is a suspicion that some of these key assumptions are not fulfilled (notably regarding migration and death coverage variations by age), Hill, You and Choi (2009) suggest that combining both approaches – first by estimating the change in census coverage with the GGB method, then by subsequently applying the SEG method to the adjusted census data – produces the most reliable results. For the purpose of our mortality completeness analysis, and in an effort to optimize the robustness of our estimates, **we choose to focus on this combined approach henceforth referred to as the GGB-SEG method.**

III. Results

In order to apply the GGB-SEG method to our dataset we use the set of functions available in the SubnationalCRVS package developed by UNFPA Population Data Fellow Jeremy Roth, itself building on Tim Riffe's DemoTools package.

In addition to implementing a DDM-based analysis, the SubnationalCRVS package also allows for a streamlined assessment of demographic data quality, notably by converting tabulated data into visual products (plots) seamlessly. We illustrate some of these features below (a more detailed corpus of plots and results can be found in Annex 1 and 2).

a. Data Quality Assessment

First, we take a look at the age heaping, sex ratios, and age ratios at the national level - and how these have evolved between the 2005 and 2018 censuses:

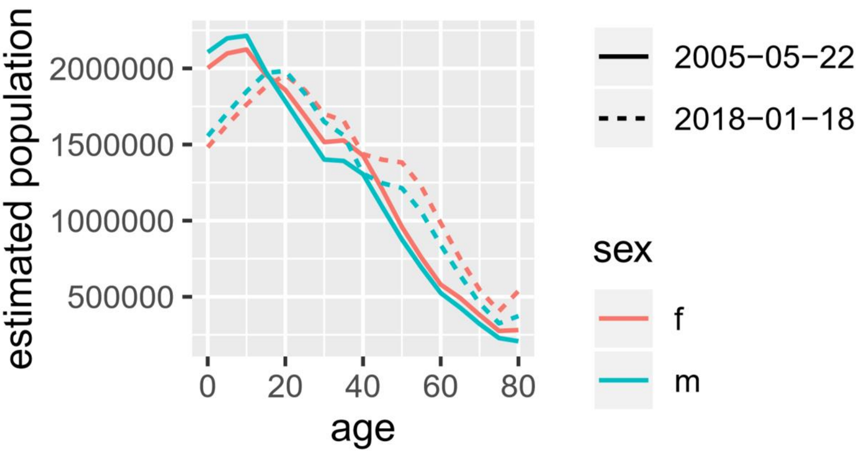

Fig 2. Age heaping in 2005 and 2018, national level

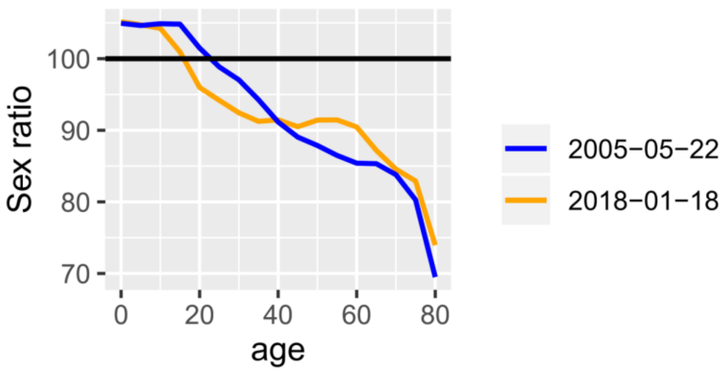

Fig 3. Sex ratios in 2005 and 2018, national level

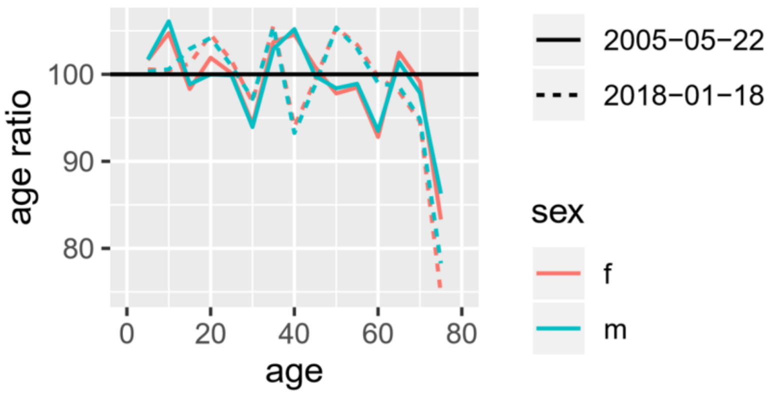

Fig 4. Age ratios in 2005 and 2018, national level

In an attempt to better understand the explanatory factors underlying the population distribution variations, we disaggregate these plots by department:

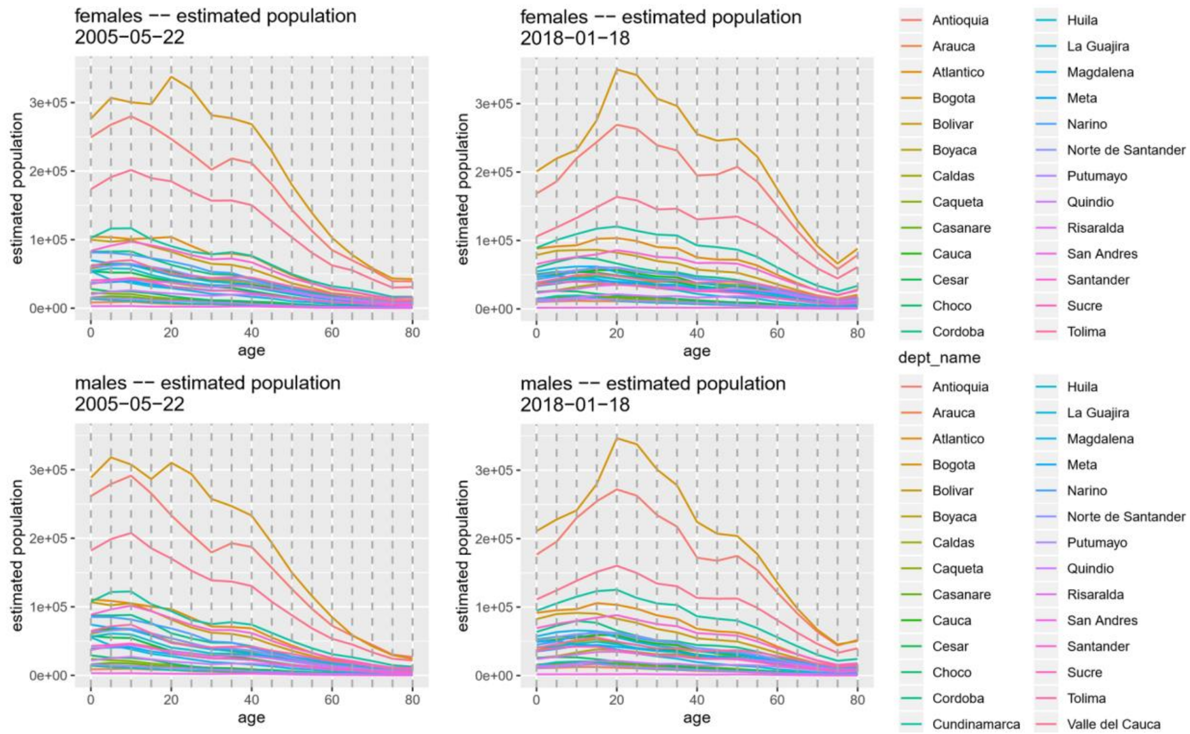

Fig 5. Age heaping in 2005 and 2018, department level

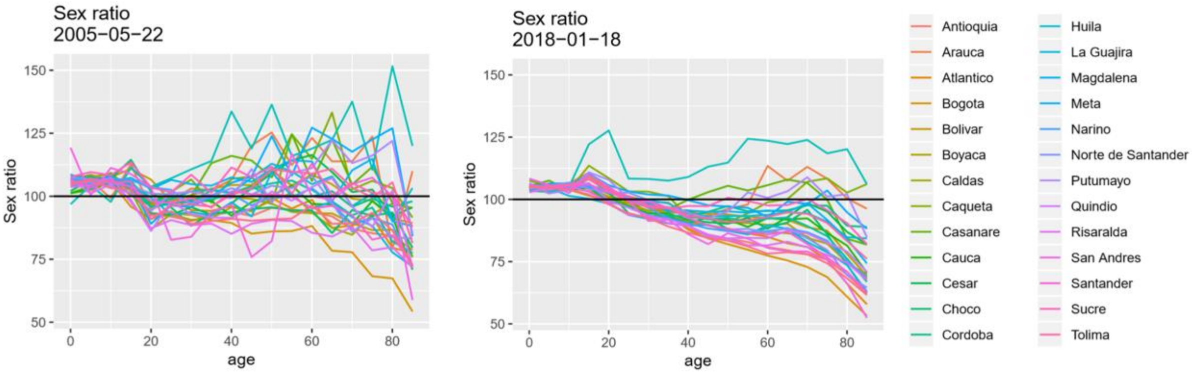

Fig 6. Sex ratios in 2005 and 2018, department level

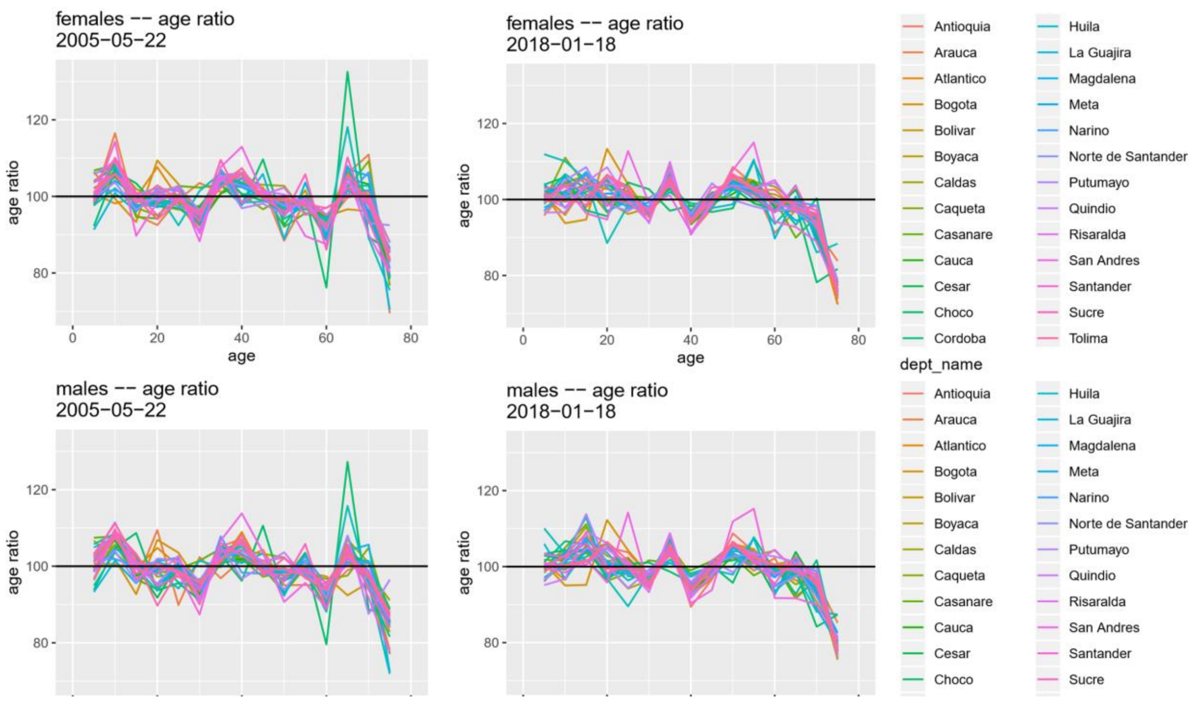

Fig 7. Age ratios in 2005 and 2018, department level

We see that the more populated departments, such as the District of Bogota, tend to behave in a more expected manner compared to less populated departments that exhibit more erratic patterns (see table 3 for population count by department). Individual plots for each department can be found in Annex 2.

b. Completeness estimates

The overall variability shown in the previous section reinforces the case for choosing to use the GGB-SEG method over other DDMs. We apply this technique to our main dataframe and, after ordering the departments by female completeness rates, obtain the following results:

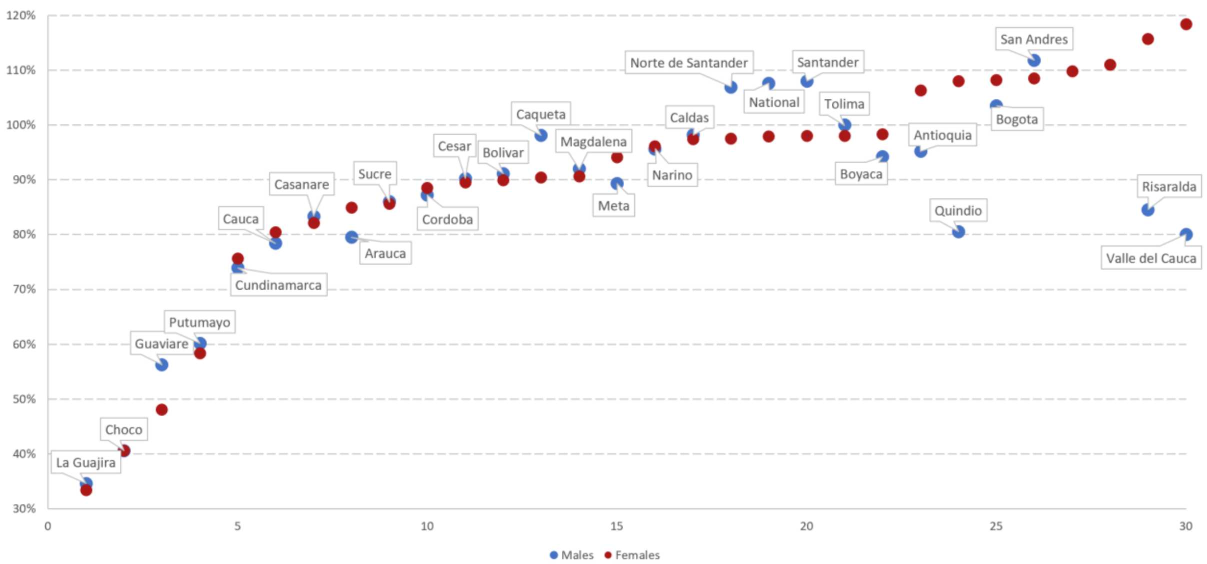

Fig 8. Sex differences in death registration completeness, department level

We then compare these results with those presented by Adair and Lopez (2018). Their first model (M1) predicts completeness based on the relationship between the population age structure and the true level of mortality, the latter relying on several variables including registered deaths, the under-five mortality rate, and the completeness of under-five death registration. The second model (M2) features the same set of variables with the exception of the under-five registration completeness measure. Similarly, we plot the results ordered by female completion rates as seen below:

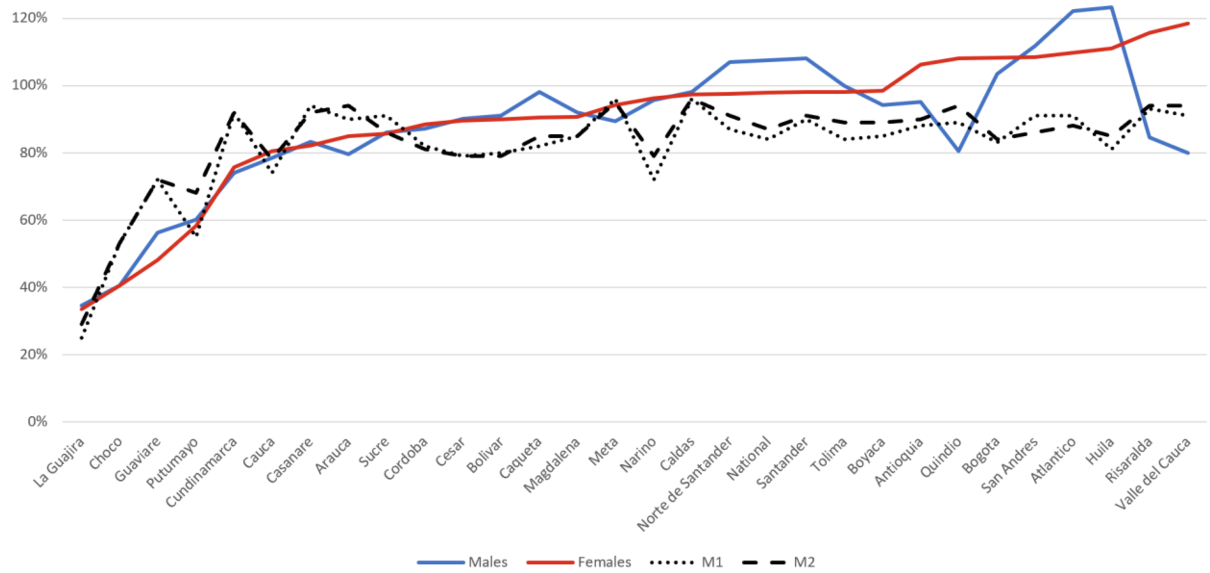

Fig 9. Comparison of death registration completeness estimates with Adair and Lopez models (M1 & M2), department level

Although the completion estimates obtained with the GGB-SEG method are on average slightly higher than those produced by Adair and Lopez, they are generally consistent as seen in the following table:

Department	GGB-SEG sex differential (M-F)	2018 population	GGB-SEG Males	GGB-SEG Females	Adair, Lopez M1	Adair, Lopez M2
La Guajira	1%	825,364	35%	33%	25%	29%
Choco	0%	457,412	41%	41%	53%	53%
Guaviare	8%	297,900	56%	48%	72%	72%
Putumayo	2%	283,197	60%	58%	55%	68%
Cundinamarca	-2%	2,792,877	74%	76%	92%	92%
Cauca	-2%	1,243,503	78%	80%	74%	78%
Casanare	1%	379,892	83%	82%	94%	92%
Arauca	-5%	239,503	80%	85%	90%	94%
Sucre	0%	864,036	86%	86%	91%	86%
Cordoba	-1%	1,555,596	87%	89%	82%	81%
Cesar	1%	1,098,577	90%	89%	79%	79%
Bolivar	1%	1,909,460	91%	90%	80%	79%
Caqueta	8%	359,602	98%	90%	82%	85%
Magdalena	1%	1,263,788	92%	91%	85%	85%
Meta	-5%	919,129	89%	94%	96%	95%
Narino	-1%	1,335,521	96%	96%	72%	79%
Caldas	1%	923,472	98%	97%	96%	96%
Norte de Santander	9%	1,346,806	107%	97%	87%	91%
National	**10%**	**44,164,417**	**108%**	**98%**	**84%**	**87%**
Santander	10%	2,008,841	108%	98%	90%	91%
Tolima	2%	1,228,763	100%	98%	84%	89%
Boyaca	-4%	1,135,698	94%	98%	85%	89%
Antioquia	-11%	5,974,788	95%	106%	88%	90%
Quindio	-27%	509,640	81%	108%	89%	94%
Bogota	-5%	7,181,469	103%	108%	83%	84%
San Andres	3%	48,299	112%	109%	91%	86%
Atlantico	12%	2,342,265	122%	110%	91%	88%
Huila	12%	1,009,548	123%	111%	81%	85%
Risaralda	-31%	839,597	85%	116%	93%	94%
Valle del Cauca	-38%	3,789,874	80%	118%	91%	94%

Completeness
- <60%
- 61-80%
- 81-100%
- 100%+

Table 3: Death registration completeness estimates and population counts, department level

We can see that the departments with the lowest estimated rates of completeness tend to be the ones with a relatively high proportion of indigenous people (Guaviare, La Guajira) and/or afrodescendants (Choco, Cauca, Valle del Cauca). We can visualize the geographic distribution of the coverage estimates with the following maps:

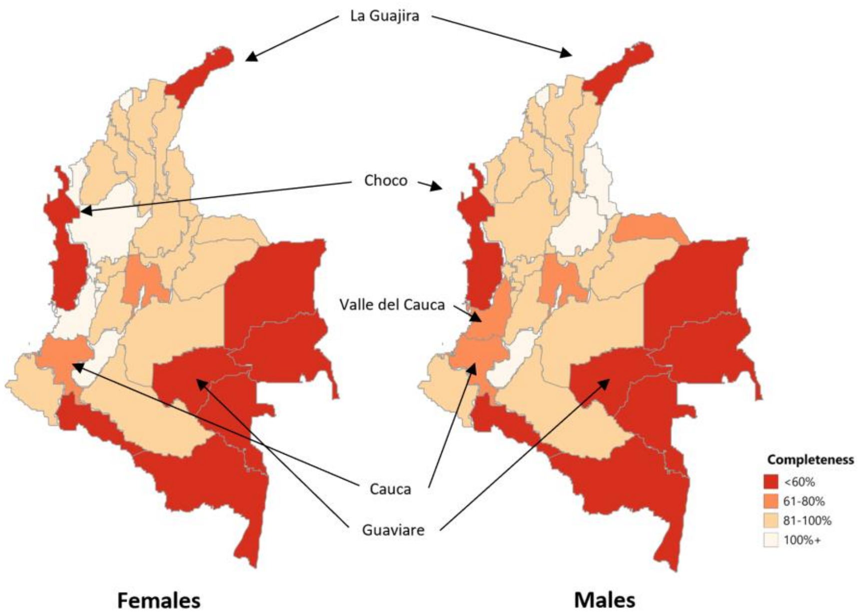

Fig 10. Maps of death registration completeness, by gender and department

IV. Limitations

Two main concerns emerge in regard to the limitations of this analysis. First, the theoretical assumptions upon which DDMs rely are often challenged in a real-life context. For instance the decades long conflict that has been unfolding since the mid-1960s involving the Colombian government, FARC rebels and paramilitary groups has led to civil populations fleeing violence, rendering census coverage homoscedasticity virtually impossible to achieve. Similarly, population displacement (whether forced or as a result of voluntary migration) challenges the closed population assumption. We would therefore encourage expanding on this analysis by measuring and subsequently adjusting for migration flows between departments.

The second concern is related to the age range parameters of the DDM functions embedded in the R packages. Adjusting the tail ends of the age groups to be considered can alter the coverage estimates, sometimes significantly. We illustrate this parameter sensitivity below with the departments of Bogota and La Guajira (Annex 1 features sets of DDM results using different age ranges, and Annex 2 features sensitivity plots for all departments using the best fit as "point estimate"). Taken together, **these limitations highlight the fact that completeness estimates should not necessarily be taken at face value; rather, they should be contextualized and interpreted as part of a broader analysis.**

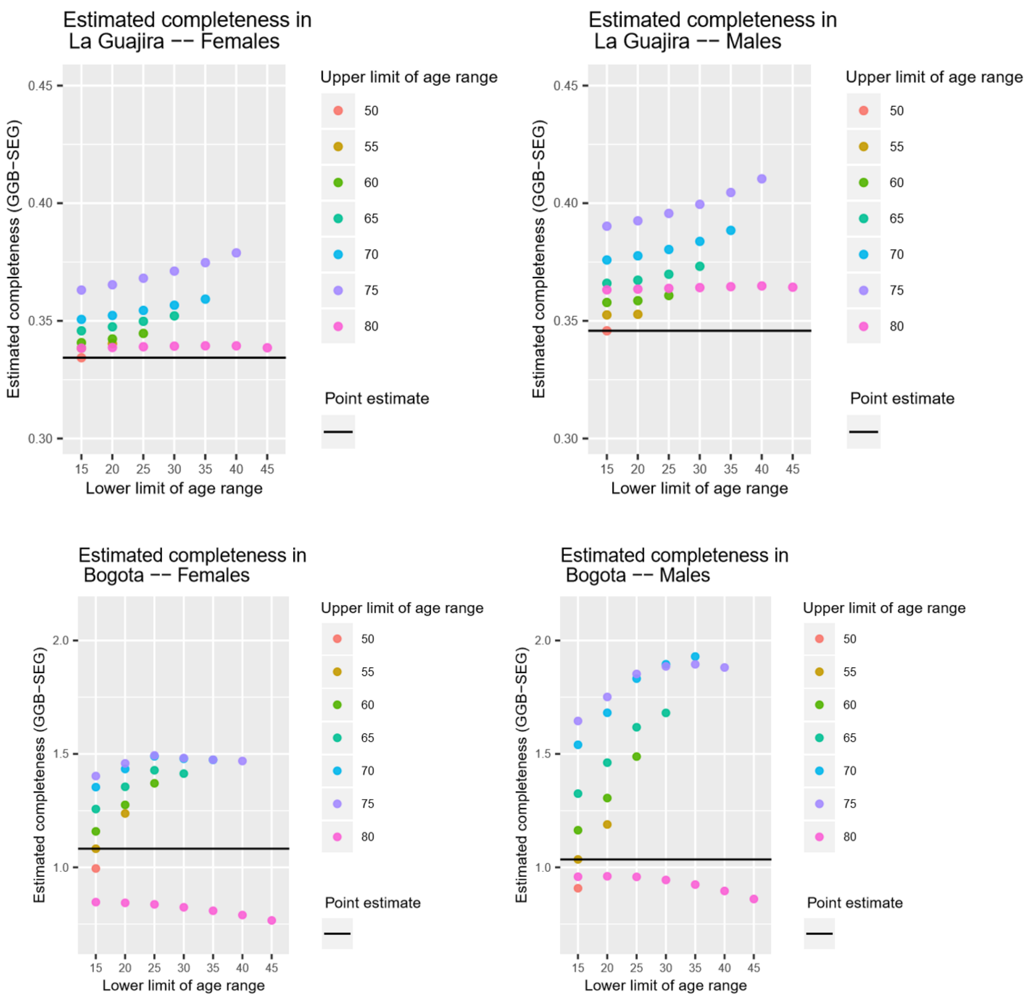

Fig 11. Age range sensitivity and its impact on results

V. Conclusion

Despite the abovementioned limitations we believe that there is significant value to be gained by applying DDM methods, and particularly the GGB-SEG technique, to CRVS data disaggregated at the subnational level. Though individual values may lack in interpretability power (e.g. results above 100%), the fact that the relative ordering of departments is consistent not only with sociological expectations, but also with peer-reviewed academic literature, points to the validity of the statistical reasoning and implementation. We believe this kind of analysis could be further deployed to CRVS data disaggregated along different socioeconomic dimensions, such as education or income level.

VI. References

Adair T, Lopez AD (2018). Estimating the completeness of death registration: An empirical method. PLoS ONE 13(5): e0197047.

Hill, K., You, D., and Choi, Y. (2009). Death distribution methods for estimating adult mortality: Sensitivity analysis with simulated data errors. Demogr Res, 21 (2009), pp. 235-254

Population Division, Department of Economic and Social Affairs, United Nations Secretariat. (2002). Methods for Estimating Adult Mortality

VII. Supporting Documentation

Annex 1: Data Analysis Results (.xlsx)

Annex 2: Plots (.pdf)

Annex 3: Code (.rmd, .html)