

1. Journal instruction requires a city and country for affiliations; however, these are missing in affiliations 2–4. Please verify if the provided city names are correct and amend if necessary.

They are correct, i.e.

2: Bellaterra, Spain

3: Bogotá, Colombia

4: Southampton, UK

2. Please check and confirm that the layout of the Table 4 is correct, amend if necessary.

Some minor amendments have been made.

There is also one important. It concerns Table 5:

The Total Mortality column forms part of the results for Women! This, I couldn't change. So please place "women" above "Total mortality".

Another thing I couldn't change is the alignment of the names of the statistical indicators in the first column in Table 4 (from Spatial Rho to Moran's I). Now it appears that they are model variables. In the last proof, the names of the variables were aligned to the left and the statistical indicators to the right. Another option is to leave a blank line between "No health insurance" and "Spatial Rho".

The same applies to Table 5 regarding "Adjusted R-squared" and "AIC" - or align it to the right or leave a blank line between Adjusted R-squared and "no health insurance".

3. Please note that the boxed values can be changed as underlined values in Table 9. Please check and confirm.

That's fine.

Original Article

Mortality by cause of death in Colombia: a local analysis using spatial econometrics

Jeroen Spijker, ¹✉

Email jspijker@ced.uab.es

Joaquín Recaño, ^{1,2}

Sandra Martínez, ³

Alessandra Carioli, ⁴

¹ Centre d'Estudis Demogràfics, Carrer de Ca n'Altayó, Edifici E2, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

² Department of Geography, Universitat Autònoma de Barcelona, Bellaterra, Spain

Received: 15 December 2017 / Accepted: 22 July 2020

Abstract

Colombia is undergoing major changes in mortality patterns. National- and department-level cause-specific analyses have previously been carried out, but very little is known about municipal-level trends, despite their epidemiological interest. We first analyze standardized mortality rates for seven cause-of-death groups to obtain high and low mortality clusters based on the spatial autocorrelation indicators *Global Moran's I* and *Local Moran's I*. The Mann–Whitney nonparametric test is then used to ascertain statistical associations between the high and low mortality clusters and known health determinants. We subsequently apply spatial lag and Durbin (when spatial autocorrelation was present) and OLS models (when not) to explain overall spatial patterns in cause-specific mortality. Age- and sex-specific cause-of-death mortality and population data were obtained from the National Administrative Department of Statistics (DANE). Deaths were corrected for each municipality due to under-registration. Results show that spatial autocorrelation declined over time for all cause-of-death categories, except male circulatory system diseases and perinatal mortality. It is highest in external causes, especially among men, with mortality hotspots moving from the central Andean area to Orinoquia and the Amazon rainforest. Male mortality is also more spatially clustered than female mortality and especially neoplasms, and external-cause mortality is also indirectly affected by the conditions of neighboring municipalities. Municipal surface area, ethnicity and public expenditure on health and education are the most frequent contextual variables explaining territorial differences in mortality. The identification of geographical mortality clusters in Colombia will allow decision makers to prioritize those regions with higher mortality.

AQ1

Keywords

Spatial cluster analysis
Spatial Durbin model
Mortality
Causes of death
Epidemiology
Colombia

JEL Classification

C21

1. Background

Mortality in Colombia is still transiting from the second to the third phase of the epidemiological transition (ET), i.e., from a pattern dominated by communicable diseases to one where chronic diseases have become most common (Omran 1971; Otero 2013; Cristancho 2017). Likewise, Frenk et al. (1991) stated that in Latin American countries the ET is characterized by its dilated and heterogeneous pattern that blends infectious with chronic degenerative and social pathological diseases. In the case of Colombia, some of the country's interior and poorer departments still have significant rates of both emerging and re-emerging communicable diseases, which accords with the unequal regional socioeconomic development in the country. For instance, in the Amazon Region acute respiratory infections are even today among the top three mortality causes, as communities there have suffered a systematic lack of targeted public policies to improve their quality of life and also lag behind compared to the rest of the country in terms of economic and social development (Otero 2013). Conversely, the wealthiest parts of the country (including the capital district of Bogotá, Cundinamarca, Quindío, Risaralda and Valle del Cauca) are clearly situated in the third phase of the ET, as infant mortality is low, life expectancy high and cancer has (almost) displaced ischemic heart disease as the primary cause of death (*ibid.*; Cristancho 2017).

Mortality studies are fundamental to understanding the health status of a population, in which analysis of mortality by cause of death allows one to gain initial insights into possible reasons for mortality differences (Spijker 2004). By also analyzing their spatial distribution and its etiology (i.e., direct and indirect factors that cause diseases), decisions can then be made in accordance with the needs of each region, whether in terms of health or other health-related aspects (López and Arce 2008). The level of mortality could therefore be considered as an indicator of the health and living condition of a population, which are, in turn, related to demographic, economic, social, biological, cultural and political factors.

Interest in the study of mortality in Colombia has increased in recent years due to the changes in the cause-of-death pattern from communicable to chronic diseases. The most studied themes have focused on mortality rates and life expectancy for which updated information is publicly available at the departmental level. Important emphasis has also been given to external causes, particularly homicide—because of the country's historic specificity and the convergence of problems of a political, economic, legal, psychological and individual nature (e.g., Franco-Agudelo 1997; Arroyave et al. 2014), non-communicable diseases such as cancer (e.g., Piñeros et al. 2013) and certain communicable diseases typical of tropical countries (e.g., dengue; Misnaza et al. 2016). On the other hand, information on the impact of specific causes of death on life expectancy is limited (Cristancho 2017), as well as analyses based on geographical models

that allow the identification of mortality clusters based on municipal data. Indeed, geographical analysis of mortality in small areas has been a neglected topic despite its epidemiological interest, not only in Colombia but in Latin American demography in general, perhaps with the exception of Brazil. However, given the known regional diversity in cause-specific mortality in Colombia, it justifies analyzing such patterns at a small scale. The same applies to the use of spatial econometrics, a tool that permits the identification of high and low mortality clusters and potential explanations for the observed spatial variation.

Conversely, international literature on the analysis of spatial autocorrelation in issues related to health and mortality is quite abundant and varied. Such studies have generally sought to provide evidence on causal relationships related to the environment and use of health services, which can be used to facilitate public health decision making. Examples based on US county-level data include adherence to breast and colorectal cancer screening (Feng et al. 2016), colorectal cancer mortality in North Carolina (Kuo et al. 2019) and stroke hospitalization and mortality risk in Florida (Roberson et al. 2016). Recent spatial analysis studies from other countries include ones by Rodrigues et al. (2013) on maternal and child health in the Brazilian state of Pernambuco, Santos et al. (2019) on child leprosy in Northeast Brasil, Piuvezam et al. (2015) on mortality from cardiovascular diseases among Brazilian elderly, Kim et al. (2016) on cancer mortality from different sites in Korea, Lin et al. (2019) on suicide mortality in Taipei City, Taiwan and Thompson and Gartner (2014) on homicide mortality in Toronto. Notwithstanding, there are perhaps only a handful of spatial studies that analyzed more than one cause-of-death category (two exceptions include Diez Roux et al. 2007; Windenberger et al. 2012) or that examined the geographical variation in mortality rates by testing the effect of potentially explanatory features of neighboring areas after accounting for the characteristics of the specific area (one exception is Yang et al. 2015).

Although different contributions have been made to the analysis of cause-specific mortality in Colombia, spatial autocorrelation studies in general and regarding (cause-specific) mortality at the municipal level in particular are still scarce in Colombia. Acevedo Bohórquez and Velásquez Ceballos (2008) analyzed municipal information of homicides recorded in 2001 in a single department of the country. Their results showed that mortality for this cause did not have a random distribution. The Colombian National Health Institute and Health Observatory (INS/ONS 2014a) estimated the spatial autocorrelation for avoidable causes of mortality between 1998 and 2011. Later that year, a geographical analysis of homicides was included (INS/ONS 2014b), which showed the presence of positive spatial autocorrelation of the standardized mortality rates for homicides (the Global Moran equaled > 0.5) as well as the presence of high and low mortality clusters. However, despite their spatial analyses, these studies were rather descriptive as they did not aim to explain those spatial patterns.

The current study therefore builds on previous research investigating the spatial distribution of major causes of death at the municipal level in Colombia. Mortality and population data came from the National Administrative Department of Statistics (DANE),

although we correct for under-registration of cause-specific deaths, as is explained in detail in the ensuing section. Our paper therefore has five main contributions:

1. to produce better spatial data on mortality by cause of death by age and sex at the municipal level in Columbia than officially available;
2. to identify the presence of spatial clusters in mortality for major causes of death (infectious diseases, neoplasms, circulatory system disease, perinatal mortality, external causes, remaining causes and ill-defined causes);
3. to show that departments are too heterogeneous for data analysis so that analysis at municipal level is more meaningful;
4. to statistically test whether several known health determinants help to explain the observed spatial variation, i.e., the existence of mortality clusters, in Colombia during the period 2004–2006.
5. Construct spatial regression models that overcome the shortcomings of classical analytical approaches used in ecological mortality research, including ordinary least squares (OLS) that ignores spatial effects.

2. Data and methods

Currently, Colombia has 33 departments (see Fig. 1) and 1122 local administrative entities,¹ comprising 1103 municipalities, 18 “non-municipalized areas”² and San Andrés and Providencia. For the spatial analysis, San Andrés and Providencia were excluded because they are two groups of islands situated too far from the Colombian mainland (about 775 km) to be considered contiguous to other municipalities. Standardized Mortality Rates (SMR) for seven major causes of death (see Table 1) were estimated from 1998 to 2014. We obtained the required mortality and population data by single ages and sex at the municipal level for the studied period directly from the Department of National Statistics of Colombia (DANE). Mortality data were derived from the national data file and contains non-fetal deaths at municipal level, and the population data came from DANE (2014).

Fig. 1

Departments and capital district of Colombia. *Source:* [https://commons.wikimedia.org/wiki/File:Colombia,_administrative_divisions_-_es_-_colored_\(%2Bbox\).svg](https://commons.wikimedia.org/wiki/File:Colombia,_administrative_divisions_-_es_-_colored_(%2Bbox).svg)



Table 1

Descriptive statistics of sex- and cause-specific standardized mortality rates of Colombian municipalities per population of 100,000 for the period 1998–2014 before and after data correction

Cause of death (Pan American Health Organization's 6/67 CIE-10 Code)	Indicator	Data correction	Men					
			1998– 2000	2001– 2003	2004– 2006	2007– 2009	2010– 2012	2013– 2014
All-cause mortality	Average	Before	448.3	465.1	398.3	356.9	314.4	363.3
		After	637.6	613.9	566.0	508.0	466.8	539.7
	CV	Before	58.4	55.7	52.0	52.1	58.9	45.3
		After	56.0	53.2	52.1	54.5	66.2	47.8

Source: Own calculations based on the microdata obtained from Vital Statistics and the denominators supplied by DANE. CV coefficient of variation

Cause of death (Pan American Health Organization's 6/67 CIE-10 Code)	Indicator	Data correction	Men					
			1998–2000	2001–2003	2004–2006	2007–2009	2010–2012	2013–2014
Infectious diseases (A00-B99, G00-G03, J00-J22, P35.0)	Average	Before	18.8	17.9	16.3	13.4	11.5	17.1
		After	35.6	31.7	28.6	23.1	21.4	23.3
	CV	Before	105.0	113.4	117.5	132.3	169.3	105.1
		After	129.2	132.4	129.4	150.6	233.6	127.9
Neoplasms (C00-D48)	Average	Before	46.4	46.3	43.8	41.6	35.5	49.9
		After	67.3	64.1	61.5	59.5	52.7	72.9
	CV	Before	78.8	74.8	71.2	73.5	82.0	63.7
		After	78.8	75.3	71.3	76.4	81.3	64.4
Circulatory system (I00-I99)	Average	Before	113.8	110.6	112.2	102.7	94.2	113.1
		After	165.5	156.4	157.5	146.2	139.4	162.9
	CV	Before	67.5	62.8	63.2	63.5	68.0	63.6
		After	66.3	62.0	62.0	65.1	68.2	64.1
Perinatal (P00-P29, P35.1-P96)	Average	Before	7.9	6.3	5.5	4.6	3.2	5.0
		After	20.5	17.6	16.2	14.5	12.6	18.5
	CV	Before	147.2	160.7	157.8	181.7	227.1	154.9
		After	165.2	187.6	181.4	213.6	265.7	176.8
External causes (V01-Y89)	Average	Before	184.1	208.2	151.6	124.7	104.0	92.9
		After	229.4	230.4	195.8	158.2	134.7	123.9
	CV	Before	82.8	84.8	81.1	75.6	79.2	68.2
		After	77.4	77.4	77.2	70.1	76.2	71.2
Other causes (D50-F99, G04-H95, J30-M98)	Average	Before	62.8	64.3	61.1	59.8	55.1	74.2
		After	96.4	95.0	91.9	90.5	87.7	115
	CV	Before	84.6	80.0	83.8	81.8	97.0	66.9
		After	82.4	79.9	85.7	84.6	113.4	68.3
Ill-defined causes (R00-R99)	Average	Before	14.5	11.5	8.9	10.1	11.0	11.2
		After	22.9	17.7	14.6	16.0	18.3	18.2
	CV	Before	165.7	162.3	149.1	136.2	119.3	152.4
		After	165.8	169.8	169.9	149.9	135.6	174.4

Source: Own calculations based on the microdata obtained from Vital Statistics and the denominators supplied by DANE. *CV* coefficient of variation

Cause of death (Pan American Health Organization's 6/67 CIE-10 Code)	Indicator	Data correction	Men					
			1998–2000	2001–2003	2004–2006	2007–2009	2010–2012	2013–2014
Cause of death (Pan American Health Organization's 6/67 CIE-10 Code)	Indicator	Data correction	Women					
			1998–2000	2001–2003	2004–2006	2007–2009	2010–2012	2013–2014
All-cause mortality	Average	Before	247.8	243.3	223.6	200.4	177.5	226.8
		After	406.7	364.0	334.5	296.6	272.5	335.0
	CV	Before	57.3	55.8	58.4	57.8	68.7	47.6
		After	58.8	56.7	62.8	60.6	80.8	50.6
Infectious diseases (A00-B99, G00-G03, J00-J22, P35.0)	Average	Before	14.2	12.9	11.9	9.3	8.5	12.3
		After	29.0	23.8	21.2	16.1	16.0	20.5
	CV	Before	119.9	110.6	139.4	129.5	463.4	116.5
		After	138.9	126.9	165.7	145.0	563.6	148.8
Neoplasms (C00-D48)	Average	Before	41.2	40.4	38.5	33.9	29.4	40.5
		After	63.4	56.6	53.9	47.1	41.8	55.5
	CV	Before	74.4	74.5	71.0	74.3	84.9	67.7
		After	74.9	72.3	72.9	74.5	87.7	69.5
Circulatory system (I00-I99)	Average	Before	92.6	89.9	88.6	78.3	69.1	84.6
		After	146.3	131.2	127.6	110.5	100.8	118.1
	CV	Before	65.0	64.8	76.5	65.4	69.8	60.3
		After	65.4	62.8	83.6	65.0	71.1	61.6
Perinatal (P00-P29, P35.1-P96)	Average	Before	5.7	4.7	4.1	3.3	2.4	4.1
		After	14.9	12.5	11.3	9.9	8.7	14.1
	CV	Before	153.0	167.6	192.5	198.7	241.5	177.9
		After	173.5	195.5	244.6	219.2	273.7	201.2
External causes (V01-Y89)	Average	Before	27.9	32.2	23.6	20.7	17.6	16.5
		After	43.3	44.1	34.9	30.5	26.3	23.3
	CV	Before	95.2	100.0	88.0	91.3	106.4	99.3
		After	97.8	96.8	94.9	99.4	115.4	108.7

Source: Own calculations based on the microdata obtained from Vital Statistics and the denominators supplied by DANE. *CV* coefficient of variation

Cause of death (Pan American Health Organization's 6/67 CIE-10 Code)	Indicator	Data correction	Men					
			1998– 2000	2001– 2003	2004– 2006	2007– 2009	2010– 2012	2013– 2014
Other causes (D50-F99, G04- H95, J30-M98)	Average	Before	53.3	53.9	50.0	47.7	42.6	60.6
		After	87.8	81.1	74.6	70.7	65.6	90.6
	CV	Before	80.6	78.9	81.4	80.7	90.6	65.8
		After	81.6	80.4	83.8	83.4	97.9	69.6
Ill-defined causes (R00-R99)	Average	Before	12.9	9.2	7.0	7.2	7.9	8.3
		After	22.0	14.8	11.1	11.8	13.4	12.9
	CV	Before	170.8	180.0	166.0	148.3	141.8	173.4
		After	169.2	185.5	168.8	157.1	165.4	175.4

Source: Own calculations based on the microdata obtained from Vital Statistics and the denominators supplied by DANE. *CV* coefficient of variation

It should be mentioned that DANE and other researchers (Flórez and Méndez 1997; Rodríguez-García 2000, 2007; Vargas and Schmalbach 2013; Piñeros and Murillo 2004; Piñeros et al. 2006) identified important shortcomings in the vital statistics in recent years related to alterations to the mortality register and the turbulent political history and violence Colombia suffered during the last decades and, accordingly, employed similar methods to ours to correct for this at the departmental level. Rodríguez-García (2000) estimated that the nationwide coverage of death records is 74%, although in departments such as Amazonas, Chocó, Vaupés, Vichada and Nariño this does not exceed 50%, including for specific groups of causes. In view of this situation, DANE decided to construct mortality tables by sex and department that incorporated significant corrections in infant mortality and other ages based on different internal studies on the coverage of the registry that showed important regional discrepancies. However, these corrections were not passed on at any time to the municipal scale or to the causes (DANE 2007). Rodríguez-García et al. (2017) underscored that for infectious diseases, the departments that exhibit the highest level of mortality are, in general, the most rural ones in the country, such as Guainía, Amazonas, Vaupés, Cauca, Vichada, Putumayo, Nariño, Chocó, Córdoba, La Guajira, Guaviare and Caquetá, departments which are also among those with the lowest mortality coverage. For this reason, the data we use in this study come from corrections applied to the obtained aggregate number of deaths in Colombia according to sex, age group, municipality of residence and cause of death. The procedure for estimating the vital statistics we do as followed:

- (a) For each year (t) we estimate the corrected total number of deaths (\bar{D}) by sex (s), age group (a) and department of residence (g) by applying the mortality rates μ by sex, age group and department of residence as published by the official DANE

mortality tables (DANE 2007). DANE prepared these mortality tables for the period 1985–2020 for each sex separately based on the information from the last three censuses, the birth records according to the residence of the mother and the deaths according to the residence of the deceased, information on the degree of omission, as well as special estimates for infants and children from 1 to 4 years. The population data (P) by sex, age and department of residence for the same period was also obtained from DANE. Life table deaths are thus estimated in the following way:

$$\bar{D}_{t,s,a,g} = \mu_{t,s,a,g} \times P_{t,s,a,g} \quad 1$$

- (b) The obtained deaths ($\bar{D}_{t,s,a,g}$) are then pro-rata distributed to the cause-specific (C) distribution by age and sex at the departmental level that were yielded from the Colombian vital statistics microdata. Accordingly, we obtain age- and sex-specific deaths for the seven large groups of causes at the departmental level, i.e.:

$$\bar{D}_{t,s,a,g,C} = \bar{D}_{t,s,a,g} \times \frac{D_{t,s,a,g,C}}{D_{t,s,a,g}} \quad 2$$

In other words, we assume that there is no bias in the registration of causes of death, suggesting that the essential problem in the Colombian death registries is overall coverage. Digging deeper into this issue, however, is beyond the scope of this paper.

- (c) Lastly, we redistribute the corrected death data at the departmental level to the municipal level (m). This was done by applying them pro-rata to the share of deaths in each municipality:

$$\bar{D}_{t,s,a,m,C} = \bar{D}_{t,s,a,g,C} \times \frac{D_{t,s,a,m,C}}{D_{t,s,a,g,C}} \quad 3$$

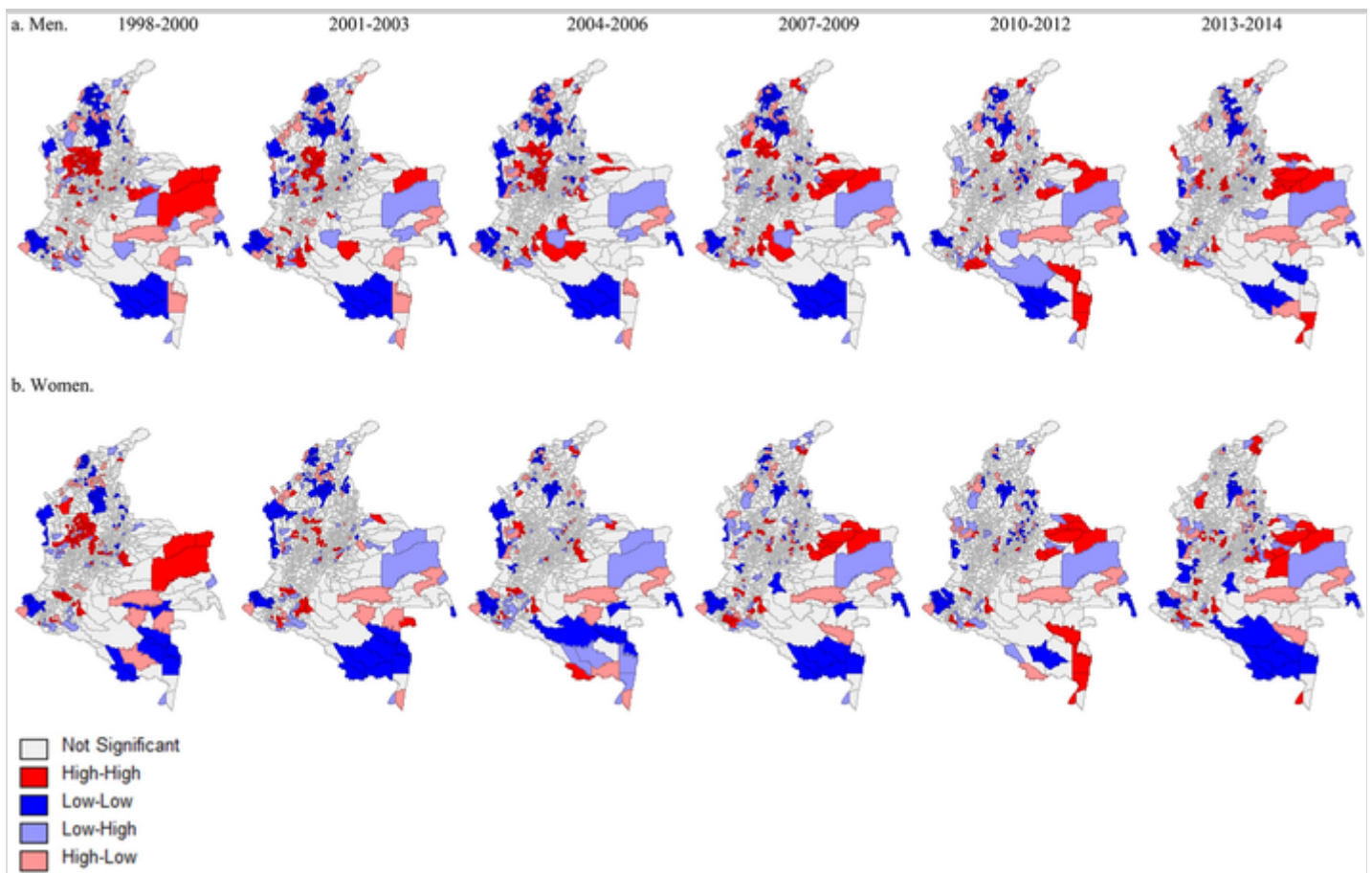
Once having obtained our small-area dataset of deaths according to sex, age group and cause-of-death categories across a 17-year period, our next step consists in calculating the age-, sex- and cause-specific death rates for the major cause-of-death categories for each municipality. In order to minimize yearly random fluctuations, we aggregate the data for the following three-year periods: 1998–2000, 2001–2003, 2004–2006, 2007–2009, 2010–2012 and 2013–2014.³ To allow for comparison over time and across municipalities, we use the direct standardization method to calculate Standardized Mortality Rates (SMR). This is done by first multiplying age- and sex-specific mortality rates of municipalities $\mu_{as}^{\text{municipality}}$ by a standard population P_a^{COL} , taken here as the national population age structure of both sexes combined according to the 2005 census. The population is distributed as per the following age structure: 0, 1–4, 5–9, ..., 80+. The open group is chosen to coincide with the lifetables and populations published by DANE. The sum of the products is then divided by the total population of Colombia p^{COL} to obtain the cause-specific SMR for each municipality:

$$SMR^{\text{municipality}} = \frac{\sum_{a=0}^{\omega} \mu_{as}^{\text{municipality}} \times P_a^{\text{COL}}}{\sum_{a=0}^{\omega} p^{\text{COL}}}$$

Before embarking on the spatial analysis, however, we first analyze whether the effect of the adjustments that are made to account for under-registration of mortality on the SMR changed over time and differed between spatial units. Given that the adjustments are only sex- and age-specific and under-registration is worst among the youngest ages, those causes of death with a young age profile (perinatal and infectious disease mortality) are most affected by the corrections. However, both the absolute and relative differences between the two rates do not change much over time, neither does the geographical pattern (as an example, see Fig. 2 for total mortality). We therefore decided to correct the death statistics by taking into account the different regional context. The level of coverage of death registries is related to a region's infrastructure and available (financial) resources. These include the presence of hospitals and the availability of qualified personnel to timely and properly carry out the registration, which is not always the case in rural areas and small towns (Piñeros and Murillo 2004). Additionally, particularly remote municipalities encounter difficulties in accessing the internet that limit completing mortality records per se and in time. Furthermore, regions with a large indigenous population do not record all deaths due to cultural factors, while there are also instances deaths are registered but not entered into the DANE database. One study estimated that 90% of maternal and perinatal deaths in Colombia in 2002 were registered, with Chocó (64%), Sucre and La Guajira (around 75%) being the departments with the highest under-registration of such deaths (DANE 2006).

Fig. 2

Ratio of corrected* to uncorrected SMR of total mortality by sex. Colombian municipalities (1998–2014). *Source:* Own calculations based on the microdata obtained from Vital Statistics and the denominators supplied by DANE. *See main text for how the municipal deaths were corrected



Conversely, the more developed departments of the Andean mountain range, the coffee-growing region and the large urban centers of Bogotá, Cali, Medellín and Barranquilla do not require major data corrections. The more accurate registration of deaths in these regions is linked to a greater presence of Central Government. In Table 1, we provide descriptive statistics of the cause-specific SMR at the national level before and after the data correction we implemented and in Appendix 1 the results from the linear regressions between the two. Briefly, results show that particularly the municipalities located in the Departments of Chocó (on the Pacific coast), La Guajira (on the Atlantic Coast bordering with Venezuela) and Amazon (in the southeast) were most affected by the corrections. As a consequence, a number of high mortality clusters emerged that the epidemiological literature on Colombia had previously identified using morbidity registers (Ochoa and Osorio 2006), but which had gone unnoticed in the analysis of vital records. Rodríguez-García (2007) identified similar levels of under-reporting at the departmental level for the year 2000 to that we obtained.

In order to explain the observed sub-national mortality patterns, information was obtained for 1087 municipalities⁴ from the 2005 census and administrative records on the following known health determinants (see also Appendixes 2 and 3 for more detail on their measurement, data sources and descriptive statistics for the high and low mortality clusters): population density and degree of urbanization (see Cyril et al. 2013 for references), temperature (Patz and Olson 2006), immigration and ethnicity⁵ (Nazroo 1998), poverty (Fiscella and Franks 1997; Rodríguez-Acosta 2016), level of education (Holmes and Zajacova 2014), public investment in health (Jaba et al. 2014), education (Holmes and Zajacova 2014), water treatment (Pérez-Vidal et al. 2016) and health

insurance coverage (Lloyd-Sherlock et al. 2012). Lastly, we also include municipal surface area as a proxy for geographical isolation (as most large municipalities are located in the Amazon area), which, in turn, is associated with fewer structural resources like social services and health care. These are known determinants of intimate partner violence (Lanier and Maume 2009) and infectious diseases (Walker et al. 2016).

2.1. Spatial Cluster Analysis

To study the possible existence of spatial dependence in the sex- and cause-specific SMR, we use the spatial autocorrelation indicators *Global Moran's I* and *Local Moran's I*.⁶ Spatial statistics is an analytical tool that treats the data of the municipalities as parts of a whole, a territorial structure where neighborhood relations are established and where it is possible to analyze to what extent statistical association exists between the values of a variable that is distributed in the territory. That is why it is necessary, before calculating the indicators, to establish a criterion based on spatial distance that clearly determines the municipalities that are neighbors. Based on this criterion, a matrix of weights is constructed that relates each municipality to all others, which then serves to calculate the value of the spatial indicator. We experimented with Rook's, Queen's and nearest neighbor (NN) case contiguity, as well as with first-order and second-order (with and without the inferior order) neighbors.

The *Global Moran's I* indicator summarizes the overall spatial associations within a territory (here Colombia) according to an analyzed variable (Anselin 1995). With its calculation, a global autocorrelation test can be performed, with as null hypothesis that the variable shows spatial independence (i.e., the values of a variable do not depend on those of its neighbors). There are several alternatives to estimate the probability that the distribution of the data is random, but here we use an approximation to the value of Global Moran's I from a random permutation (specifically 999 permutations, a methodology that has an associated probability of 0.001). However, one weakness of the Global Moran's I is that it averages out local variations in the level of spatial association. The Local Moran's I statistic was therefore conceived to be able to identify local patterns of association (hot spots) (*ibid.*), whereby the obtained value provides evidence of either:

- Local positive spatial autocorrelation, i.e., regions with high values of a variable surrounded similar neighbors (HH clusters, known as hotspots) and regions with low values surrounded by regions who also have low values (LL clusters known as coldspots) and
- Local negative spatial autocorrelation, i.e., regions with high values of a variable surrounded by neighbors with low values (HL clusters) and vice versa (LH clusters).

Thus, positive spatial autocorrelation indicates the presence of clusters of similar values in the territory and is undoubtedly information that will help us to locate the mortality patterns due to similar characteristics throughout the Colombian territory.

2.2. The Mann–Whitney nonparametric test for analyzing the exogenous characteristics of the mortality hotspots and coldspots

To ascertain an understanding of the nature of the spatial clusters with positive high and low spatial autocorrelation (HH and LL), we apply several additional statistical procedures. For the analyses, we only consider the 2004–2006 period as it approximates a period when a significant number of contextual factors are available locally as they can be derived from the 2005 census. As our initial objective is to assess whether significant differences exist in the distribution of independent, potentially explanatory, variables between both spatial groupings, these variables are first subjected to a normality test using the Kolmogorov–Smirnov test with the Lilliefors correction. Since they all present a level of significance of 0.000, the hypothesis of a normal distribution is rejected and the typical ANOVA t test not used, but rather the Mann–Whitney U test.⁷ This is a nonparametric test comparing two independent samples that has the advantage of not needing a specific a priori distribution but uses the ordinal distribution of the dependent variable (in this case of the selected regressors). It is used to compare two groups of median values and determines whether or not the differences between the two are due to chance. Specifically, we aim to show whether the distribution of the explanatory variables is significantly different between the positively correlated clusters of high and low mortality for total mortality and the first five cause-of-death categories (i.e., excluding ill-defined and remaining natural causes due to a lack of a clear etiology) for each sex. From the U tests we retain the level of significance of the test and the median of each variable. Subsequently, we apply the Spearman’s rank correlation coefficient that is not subject to the restriction of normality to obtain the strength and direction of the association at the municipal level between the independent variable and the SMRs in each of the HH and LL clusters. The results of this analysis are described in Sect. 3.3.

2.3. Using Spatial Lag and Spatial Durbin models to explain spatial patterns in mortality

One disadvantage of analyzing HH and LL mortality clusters is that only a small proportion of the municipalities are analyzed (see Appendix 3). Moreover, no consideration is made regarding any possible spatial correlation of the covariates. One way to get around this is by first applying a spatial lag model and subsequently a spatial Durbin model.⁸ A spatial lag model can be expressed as:

$$y = \alpha I_n + \rho W y + X\beta + \varepsilon, \quad \text{where } \varepsilon \sim N(0, \sigma^2 I_n) \quad 5$$

y is a $n \times 1$ vector and denotes the dependent variable of $n = 1087$ municipalities, in this case cause-specific mortality; αI_n is the $n \times 1$ identity vector associated with the intercept; ρ is the spatial autoregressive coefficient associated with the lagged dependent variable $W y$, with W being the $n \times n$ spatial weight matrix.

In turn, the spatial Durbin model is applied to consider both spatial autocorrelation and spatial dependence in a set of explanatory variables are added as spatially lagged (but

usually only when significant):

$$y = \alpha I_n + \rho W y + X\beta + W X \theta + \varepsilon, \quad \text{where } \varepsilon \sim N(0, \sigma^2 I_n) \quad 6$$

X is a $n \times k$ matrix of k explanatory variables related to $\beta k \times 1$ parameters in the set of explanatory variables; θ is a $k \times 1$ vector of effects of the spatially lagged covariates, $W X$. The error term is normally distributed with mean 0 and variance $\sigma^2 I_n$, with I_n an $n \times n$ identity matrix, following LeSage and Pace (2009).

The main difference between the spatial lag and Durbin models is that the latter includes not only endogenous interaction relationships ($W y$), but also the exogenous interaction relationships that enters as spatially lagged explanatory variables ($W X$). This way it is possible to assess if and how the specific mortality cause measured in a specific unit i is related to the features of its neighbors.

Initial model covariate selection has been conducted through a best subset selection approach (Heinze et al. 2018), where covariates are progressively deleted by evaluating each model performance (both spatial lag and spatial Durbin models) through Akaike's information criterion (AIC) (Akaike 1973) estimators and the final set of variables is checked for collinearity through a variance inflation factor (VIF), which has to score below 4.0 (Fox and Weisberg, 2018). The appropriateness of the spatial lag and Durbin models against spatial error alternatives has been assessed using Anselin (1988) Lagrange multiplier tests. We only apply these models to those causes of death where spatial autocorrelation is present in the dependent variable for the period 2004–2006 (see Table 2). When this is not the case, an OLS regression is conducted.

Table 2

Global Moran's I according to period and cause of death. Colombian municipalities (1998–2014)

	1998– 2000	2001– 2003	2004–2006	2007– 2009	2010– 2012	2013– 2014
<i>Men</i>						
All causes	0.1916*	0.2091*	<i>0.1473*</i>	0.1155*	0.0751*	0.1160*
Infectious diseases	0.1333*	0.1101*	<i>0.0433</i>	0.0650*	0.2058*	0.0948*
Neoplasms (cancer)	0.1490*	0.1748*	<i>0.1350*</i>	0.1084*	0.0902*	0.1288*
Circulatory system diseases	0.1130*	0.1881*	<i>0.1708*</i>	0.1266*	0.0967*	0.1212*
Perinatal mortality	0.0200	0.0285	<i>0.0313</i>	0.0602*	0.0606*	0.0343
External causes	0.4077*	0.4306*	<i>0.4443*</i>	0.4044*	0.3568*	0.3191*

Source: Own calculations based on the microdata obtained from Vital Statistics and the denominators supplied by DANE

*Significant at $p < 0.01$. In italics: time period and causes of death studied in explanatory analyses

	1998– 2000	2001– 2003	2004–2006	2007– 2009	2010– 2012	2013– 2014
Remaining causes	0.1081*	0.0586*	0.0113	0.0101	– 0.0009	0.0619*
Ill-defined causes	0.1315*	0.1260*	0.2247*	0.1168*	0.1797*	0.1041*
<i>Women</i>						
All causes	0.1367*	0.0585*	<i>0.0140</i>	0.0494*	0.0471*	0.0945*
Infectious diseases	0.1251*	0.0595*	<i>0.0165</i>	0.0348	0.0837*	0.0814*
Neoplasms (cancer)	0.1708*	0.1584*	<i>0.1220*</i>	0.1354*	0.1202*	0.1341*
Circulatory system diseases	0.1473*	0.1327*	<i>0.0132</i>	0.1019*	0.0428*	0.0866*
Perinatal mortality	0.0208	0.0197	<i>0.0162</i>	0.0583*	0.0340	0.0696*
External causes	0.0226*	0.2633*	<i>0.1983*</i>	0.1505*	0.1263*	0.0875*
Remaining causes	0.1748*	0.0198	0.0095	0.0056	0.0195	0.0535*
Ill-defined causes	0.2505*	0.0876*	0.0886*	0.1882*	0.1643*	0.0574*
<i>Source:</i> Own calculations based on the microdata obtained from Vital Statistics and the denominators supplied by DANE						
*Significant at $p < 0.01$. In italics: time period and causes of death studied in explanatory analyses						

Evaluation metrics are included for each model and comprise the Lagrange multiplier test (LM test) for residual autocorrelation, which determines whether the model residuals are spatially autocorrelated; the AIC for the selected model and the OLS counterpart.

2.3.1. The interpretation of the spatial Durbin model

The idea behind the spatial Durbin approach is that the effects of the explanatory variables are not only felt in the area where they are measured (as it is the case for the spatial lag model), but that there is a feedback process that makes it possible that its effects expand to/from neighboring areas. A simple example would be the presence of a large health care facility, whose intake capacity exceeds that of the area where it is located, thus benefiting all its neighbors. Thus, the spatial Durbin model allows separating the direct effect of an explanatory variable on the dependent variable from the indirect effect (LeSage and Pace, 2009). Isolating the dependent variable, Eq. (6) can be rewritten as:

$$(I_n - \rho W) y = \alpha I_n + X\beta + XW\theta + \varepsilon \quad 7$$

where the partial derivative with respect to the explanatory variables set can be expressed as:

$$\frac{\partial y}{\partial X_i} = \frac{(I_n \beta_i + W \theta_i)}{(I_n - \rho W)} \quad 8$$

with $\frac{\partial y}{\partial X_i}$ a $n \times n$ matrix. The direct effects are the average of the diagonal elements of $\frac{\partial y}{\partial X_i}$, whereas the average of the off-diagonal elements of $\frac{\partial y}{\partial X_i}$ represents the indirect effects. The direct effect represents the expected average change across all observations for the dependent variable of a particular area due to an increase of one unit for a specific explanatory variable in this area (feedback effect), and the indirect effects represent how changes in the independent variable arising from a particular area influence the dependent variable in all other neighboring areas (spillover effect). It should be noted that some independent variables may not have an indirect effect as we only separate the indirect effect from an independent variable's total effect when the additional lagged explanatory variable improves the model fit (i.e., the AIC value is the same or lower). In these instances, the direct effect equals the total effect.

3. Results

Colombia is currently characterized by a decline in mortality levels at regional and municipal levels and shows certain stability in the territorial pattern in this reduction (see the coefficient of variation (CV) in Table 1). We do not rule out that the latter result is due to improvements that have been made in the mortality register since 2010 as variance increased during the latter two periods for all causes of death. This is an aspect worthy of a more in-depth study. However, strong territorial disparities still exist in terms of life expectancy at birth between the different departments (a maximum of 10 years in the case of males and 7 in the case of females; see Appendix 5). The spatial hierarchy of life expectancy at the departmental level remained at constant levels since the late 1980s (Cristancho 2017). This territorial stability makes it possible to identify a series of departments in the northern Andean region, Bogotá, the Atlantic coast and more urbanized departments as those that experienced higher life expectancies over the last decades. In contrast, all the departments to the east of the Andean mountain range located in the Amazonian area, those bordering Ecuador and the Pacific coast are the areas that are lagging behind in life expectancy.

The department-specific results are filled with nuances when we descend to the municipal scale and address the causes of death. The correction of deaths implies an increase in mortality disparities between municipalities in the categories infectious diseases, perinatal mortality, other causes and ill-defined diseases. This is verified through the time trend of the coefficient of variation (CV; Table 1). Conversely, neoplasms, diseases of the circulatory system and external causes experience a marked stability and even a slight reduction for the same indicator. Between 1998 and 2014, a generalized reduction in SMR can be observed in both sexes, affecting most of the causes with the exception of neoplasms among men and the remaining causes in both sexes.

3.1. Spatial autocorrelation in Colombian mortality

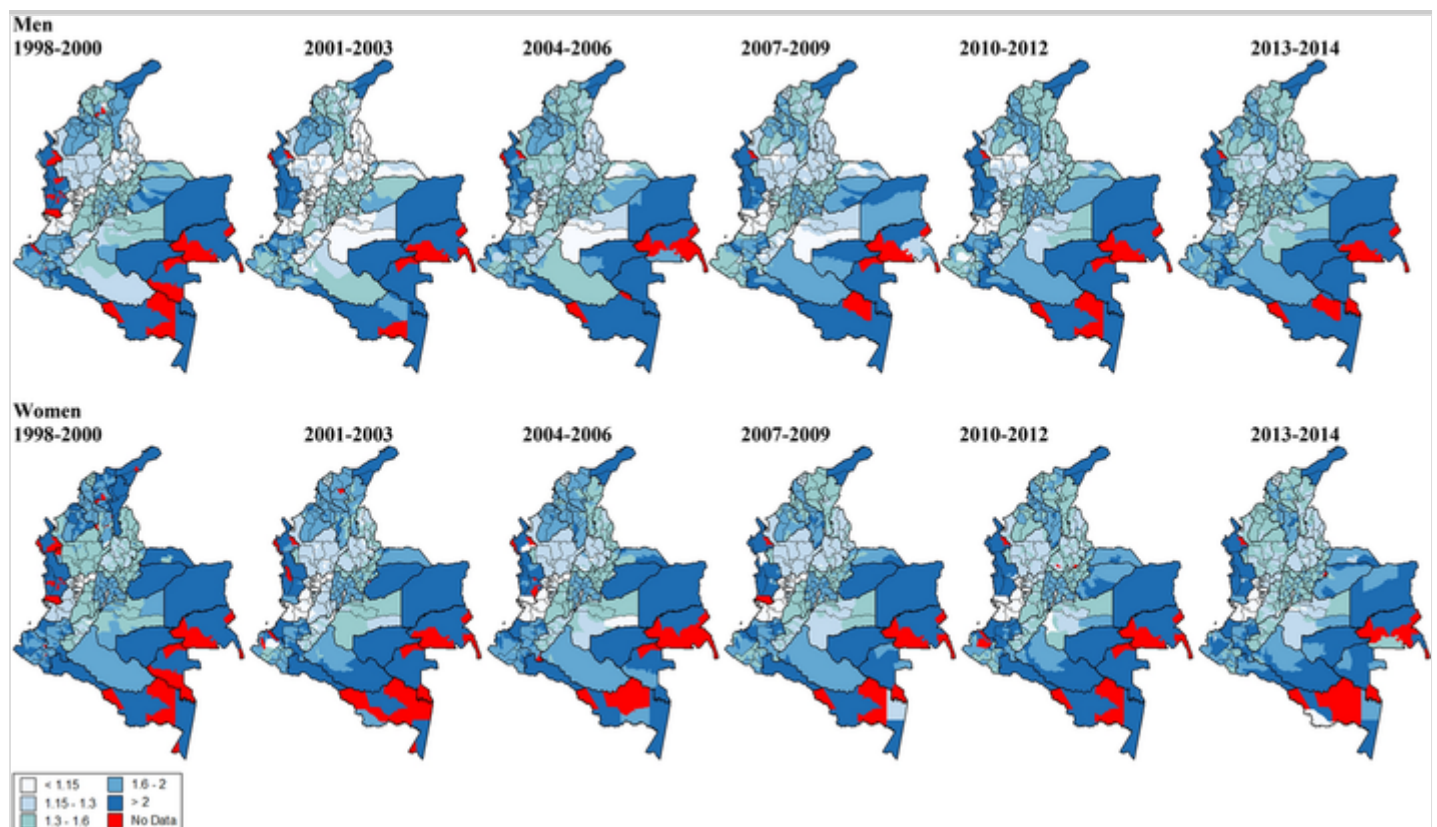
Based on the Global Moran's I indicator, results provide a clear indication of spatial autocorrelation in Colombia during the period 1998–2014 for both all-cause and cause-specific mortality (Table 2). In neoplasms, external causes and ill-defined causes there is

statistically significant spatial autocorrelation for both men and women and during all of the periods studied. In addition, regarding men this also applies to infectious diseases, diseases of the circulatory system and total mortality. This is despite the overall reduction in mortality and some territorial convergence in the decrease in the CV in total mortality. In other words, there is a clear continuity over time in the spatial patterns observed. External causes observed the highest level of spatial association.

Mapping the municipal results allows the identification of mortality hotspots as well as contiguous areas with low levels of mortality (see Fig. 3 for all-cause mortality and the Supplementary Figures S1-S7 in the OSF Preprints repository (<https://osf.io/n5pk8/>) for the specific causes of death). In the first place, the general trend in total mortality among both sexes is a decrease in areas with autocorrelation. The slight break in the trend since 2010 may be related to the spatial variations in the Colombian Vital Statistics registry but also for the effects associated with the increase in infectious diseases in specific local areas of the country. The fact that the Global Moran's I is superior among men can be easily explained by the leading role that men have regarding external causes. As the general level of mortality is the product of different cause-specific trends, the overall pattern conceals many interesting panoramas, including territorial contrasts in the specific causes that we analyzed. These are succinctly described below.

Fig. 3

Municipal clusters. All-cause mortality. Colombia (1998–2014). *Source:* Own calculations based on the microdata obtained from Vital Statistics and the denominators supplied by DANE



Infectious diseases show different clusters in the Colombian Pacific zone, south of the Cauca Valley, north of Nariño, Guajira, the area of Orinoquia (in the east of the country) and Amazonia present high levels of clusterisation. Some of these areas, as is well known by Colombian epidemiologists, are characterized by being jungle, in which there is transmission of diseases such as malaria and dengue where access to drinking water is scarce and health intervention not always optimal (Padilla et al. 2017; Quintero et al. 2014). During the period 2010–2012, there was a rise in malaria and dengue mortality in the Eastern part of the Amazon region, which is reflected in a very clear increase in the Global Moran's I value for infectious diseases (for men it goes from 0.0650 in 2007–2009 to 0.2058 in 2010–2012). The change is also observed in women, but less strong. Some authors have highlighted a link between the increase in illegal mining and the appearance of malaria outbreaks in regions with widespread illegal exploitation of minerals spurred on by the historical highs of the price of gold between 2010 and 2012 (Salas-Zapata et al. 2014). These mining communities, basically composed of men, are mainly located in difficult to access jungle regions.

As to cancer, the Andean zones located in the center of the country have the most aged population. As cancer is a degenerative disease related to having survived to older ages, particularly this area observes the highest levels of clustering of cancer mortality. Moreover, the spatial structure is maintained throughout the study period. The municipalities in blue correspond to areas with low life expectancy where cancer mortality is much lower, in part also due to a lack of diagnosis and almost absence of smoking (Murillo et al. 2004).

The geographical pattern of circulatory system diseases is very similar between time periods, although the amount of clusterisation is less in the more recent periods. The spatial distribution of these causes is associated with areas that still lag behind in the ET. Men have a higher Global Moran's I than women.

Regarding perinatal mortality, there is clustering in Orinoquia, Amazonas and certain areas in Pacific and Nariño, where health services are not accessible to all communities (Cristancho 2017). The municipalities in this part of the country are extensive in territory, what makes access to health care difficult, especially for infants resident in isolated communities in the Amazon rainforest and along the Pacific coast. The spatial autocorrelation among males is higher. A more detailed analysis of the spatial structure showed some recurrence in incomplete coverage in some areas, even after the administered corrections. These problems appear to be related to the health care that municipalities with hospitals provide for outlying (often rural) areas that inadvertently lead to an inflation of death registration of certain urban areas. This assertion should, however, be verified empirically in the future.

Perhaps the most interesting maps are those for external causes as they maintain the highest degree of clustering and continuity in the areas involved. Initially located in Antioquia and Córdoba where armed conflict at different levels (guerrillas, paramilitaries, drug trafficking and illegal mining) overlaps throughout the studied period, these areas are

maintained during the period and extended to the departments east of the mountain range, to the plain (Meta) and to the southern guerrilla area (Caquetá and Putumayo) and later consolidated in Orinoquia and Amazonas. Neither negligible is the mortality linked to urban delinquency, which is of predominantly male character. Also worthy to highlight is the slightly different geography of female mortality from external causes: the Antioquia and Cordoba areas are no longer significant, while municipalities in Meta, Caquetá and Putumayo are also characterized by high levels of mortality from non-natural causes (although no longer during the last period studied 2013–2014, except for Meta). The peculiarity of the female trend is explained by the masculinization of the areas where the violence is concentrated. In any case, the gradual reduction of Global Moran's I associated with external causes is good news for Colombia.

Concerning the remaining natural causes and ill-defined causes, results indicate a clustering of under-registration (in red color) located in departments such as Chocó, Cauca, La Guajira and the municipalities in Orinoquia and Amazonia. Nevertheless, the decrease in the Global Moran's I indicator is a sign of progress as this is in line with the improvement that has been made in the quality of vital statistics in Colombia in recent years (Rodríguez-García 2007).

Finally, to evaluate the use of uncorrected vital records in the calculation of SMR, we repeated the analyses for the first and last period using the Rook criterion of order 1 for the uncorrected data and compared the results. Results show that uncorrected data provides a biased and inaccurate view of the spatial structure of mortality in Colombia. Indeed, a high degree of under-registration in municipalities may yield contiguous areas with similar mortality rates and thus produce an inflated (and misleading) Global Moran's I (see also Appendix 6).

3.2. Explanations of the mortality hotspots and coldspots in Colombia in 2004–2006

Colombia has a heterogeneous geography that hosts very diverse ecosystems, a climate marked by extreme differences in altitude in populated spaces that extend from sea level to the tropical moorland at altitudes above 3000 m. At the sociocultural level, Colombia also has a very polarized social structure with strong inequalities that are spatially accentuated due to territorial-specific complex ethnic structures and the history of armed conflict that has plagued the country for more than sixty years, joined later by drug trafficking and illegal mining. This has generated a context of latent violence that, despite the recent peace process, has implied a disproportionate weight of deaths linked to external causes that are unequally distributed throughout the country. This heterogeneity implies the existence of multiple factors that can explain the links that emerge between the existence of spatial autocorrelation of different types of causes of death in certain territorial spaces. To better understand this complex geographical scheme, we use the Mann–Whitney *U* test to see whether the medians of the explanatory variables are significantly different between the high and low mortality clusters for total mortality and the first five cause-of-death categories in the 2004–2006 period (Appendix 3). Subsequently, for the significant

variables ($p \leq 0.05$) we obtain the Spearman's correlation coefficients to analyze the strength and direction of the association between the independent variables and the cause-specific SMRs among the municipalities in the high and low mortality clusters (Table 3).

Table 3

Spearman's rank correlation coefficient between the causes of death and the explanatory variables among

Variable	Cause of death ^a						
	Specific	C0	C1	C2	C3	C4	C5
Category	Variable	Men					
Geography	Temperature	-.531**		-.489**	-.399**		-.359 [‡]
	Area	.112	.447**	-.580**	-.589**	.586**	.368**
Settlement	Main town			.461**			
	Rurality	.107		-.322**			
	Density	-.099		.681**	.481**	-.311**	-.382 [‡]
Population	Fertility proxy		.264*			.339**	
	Older population	.259**	-.348**	.535**	.694**	-.444**	-.085
	Internal immigrants	.527**		.399**	.421**		.537**
Socioeconomic status	Low educational level		-.260*	-.594**		-.252*	.146*
	Poverty	-.554**	.169	-.746**	-.687**	.505**	-.321 [‡]
Ethnicity	Afrodescendant	-.228**	.305*	-.218*	-.459**	.438**	
	Indigenous		.676**	-.267**	-.479**	.763**	.244**
	Mixed/white	.290**	-.629**	.400**	.619**	-.688**	
Public investment	Health		-.403**	-.461**	-.223**		.154**
	Education	-.354**	.377**	-.501**	-.507**	.402**	
	Water supply		-.436**	-.544**			
Health insurance	Health insurance	.407**		.711**	.444**		.204**
	<i>N</i>	182	70	124	141	68	294

^aC0 = total mortality; C1 = infectious diseases; C2 = neoplasms; C3 = circulatory system diseases; C4 = respiratory system diseases; C5 = other causes. Only associations are shown for the significant variables in Appendix 3

Results are very similar for both sexes. Starting with infectious diseases, there is a clear geographical pattern with surface area (used as a proxy of isolation) being positively

associated as the HH cluster municipalities are 3–4 times smaller on average than LL cluster ones. The HH cluster municipalities also tend to be more populated and least advanced in the context of the Demographic Transition (i.e., higher fertility and proportionally fewer elderly). The association with low education is in the expected direction, but relatively low (-0.260 for men and -0.472 for women) and only just statistically significant and does not yield clear conclusions as the cluster difference in the median equals only 4 percentage points. The ethnic variable is also significantly associated with mortality from infectious diseases. Results show that mortality hotspots are linked to a greater presence of Afrodescendant in the case of men (+ 1 percentage point) and the indigenous population regarding both sexes (median of men 15% and women 11% vs. $< 0.1\%$ in LL cluster municipalities). Not surprisingly, the Spearman's correlation coefficient is high (0.676 for men and 0.643 for women). Results also show that greater public investment per capita in health and water supply are done in areas with high levels of infectious diseases, while HH mortality cluster municipalities also have on average higher investment in education.

Between the HH and LL cancer mortality clusters, significant differences are observed in all variables. Although their contributions are different in intensity and sign, they follow the general scheme of demographic and socioeconomic development adapted to the geographical and socioeconomic peculiarities of Colombia, where, due to competing causes of death, cancer is linked to survival to older ages and therefore low mortality. As the U test shows, HH clusters are made up of mainly Andean municipalities, with a lower average temperature and surface area, a more urban context (higher proportion residing in the main municipal town), more elderly and internal migrants, fewer low-educated people but more with a mestizos/white background and access to health care through social security system contributions. Conversely, municipalities in LL clusters have a higher degree of poverty, greater per capita public investment in health, education and water supply and a younger population. Finally, there are no important gender differences in the results.

The contextual factors associated with circulatory system disease mortality follows a similar pattern to that of cancer, although with several exceptions (the two settlement variables and public investment in health and water supply show no statistically significant association). In other words, again variables related to greater urban, demographic and socioeconomic development are positively associated with this cause. Of all variables, the proportion of people over 50 has the highest Spearman's correlation coefficient with the SMR of circulatory system diseases (0.694 for men and 0.759).

Factors affecting perinatal mortality repeat to some extent what was observed for infectious diseases. Regarding the geography variables, the municipal surface area is positively associated and density negatively associated with SMR in both sexes. This can be explained by noting the indirect association with (isolated) areas, which have less government presence and more public health issues. According to the U test, the median surface area is approximately three times greater in HH clusters than in LL clusters. Other variables associated with high perinatal mortality rates are higher fertility, fewer older

people, ethnicity (similar pattern as observed for infectious diseases), poverty and public investment in education. The association in the latter is positive because per capita investment is higher in less economically developed areas, particularly indigenous reservations, where perinatal mortality is high. Results for women are similar, except that the proportion elderly is the only significant population variable. The highest Spearman's coefficient is between perinatal mortality and the proportion of indigenous people in the municipality (males 0.763; females 0.721).

Mortality from external causes presents very interesting associations, especially with respect to the geography and settlement variables. The U test shows that male mortality hotspots are linked to larger-sized municipalities located in mid-mountain heights with low densities and a rural character. This corresponds to the geographical space of the Andean foothills, places that are characterized by a predominance of jungle with poor government control and the presence of armed groups. Municipalities with a greater presence of immigrants are positively associated with external mortality in both sexes, observing in fact the highest Spearman's correlation coefficients (0.537 and 0.631 for men and women, respectively). A greater indigenous presence is also positively associated with external-cause mortality, although its explanatory power is low.

Lastly, we describe the results for total mortality. Its pattern is not very different to that of neoplasms and circulatory system as the two causes dominate the overall mortality structure. For instance, total mortality is negatively associated with temperature, which may be considered a proxy for altitude (see Appendix 2 note h): Particularly in the case of men, most municipalities in HH clusters are located in urban areas of the Andean mountain range and the LL cluster municipalities in the lower lying coastal areas (Fig. 3). Internal migration, that also has a strong urban component, is also positively associated with total mortality. The same applies to being of mixed/white race and health insurance. On the other hand, household poverty (measured as unsatisfied basic needs, particularly in housing) is negatively associated with mortality, but this is also more common in (healthier) urban areas. Statistically significant is also public investment in education (negatively associated) and having an older population (positive). The latter association is stronger for women than for men, which is consistent with the fact that they have a higher life expectancy.

3.3. Explanations for the overall geographical mortality patterns in Colombia in 2004–2006

To better capture spatial dependence, causes of death were first tested for different neighborhood definitions. First-order Rook criterion of contiguity, i.e., a stricter definition of neighbors where neighboring units are shared by more than a single point, was considered the best to use for establishing the neighborhood relations when Moran's I was high and comparable with a more complex weight matrix (men: all causes, cancer and circulatory system diseases; women: external causes). NN $k = 4$ was chosen instead as the spatial weight matrix for male external cause-of-death mortality and NN $k = 3$ for female cancer mortality (Appendix 4). We also found little evidence of a Modifiable Area Unit

Problem (see, e.g., Meliker and Sloan 2011), given the small differences in the results between the different spatial weight matrices.

Table 4a, b presents the results for the spatial lag and Durbin models and their AIC measurements and Table 5 those of the OLS approach when little to no spatial autocorrelation was present in the cause of death (only women: total mortality and circulatory system diseases; both sexes: infectious diseases and perinatal mortality). The fact that there are fewer spatial regression models for women is a clear indication of the lower importance of geography as an explanatory factor for small-area differences in mortality. According to the AIC, using spatial Durbin is the best approach wherever spatial autocorrelation appears to be positive and significant (Moran's $I > 0$ and p -value < 0.05).

Table 4

(a) Spatial regression results for spatial lag model and spatial Durbin model: Men. (b) Spatial regression spatial Durbin models: Women

(a)						
Variable	Total mortality				Neoplasm	
	Lag	Durbin			Lag	Durbin
	Estimate	Direct	Indirect	Total	Estimate	Direct
<i>Intercept</i>	- 49.724			- 137.832	66.460***	
Temperature						
ln(Area in km ²)	51.933***	55.455***	-	55.455***	- 1.306	- 1.697
Rurality	- 0.300	- 3.289***	6.119***	2.830***	- 0.203***	- 0.512***
Older population						
Internal immigrants	- 0.581	- 0.556	3.620***	3.064***	- 0.351***	- 0.259
Afrodescendant					- 0.239***	- 0.233
Mixed/white	1.568***	1.692***	- 0.913	0.779 [†]		
Low educational level						
Investment in water supply						
No health insurance	4.950***	3.324***		3.324***	0.972***	0.877***
Spatial rho please align to the right	0.230	0.280			0.141	0.238

[†] $p < 0.1$; * $p < .05$; ** $p < 0.01$; *** $p < 0.001$. The unit of analysis is Colombian municipalities. The age-standardized mortality rate per population of 100,000. The explanatory variable values are properly specified in the table. Public investment is in ~~thousands+1000-~~ of Colombian pesos per capita. See a definitions. In the spatial Durbin models, the indirect effect of an explanatory variable is only separ when the model fit improves (i.e., the AIC value decreases) as a result. If not, the variable only has equal to the total effect

(a)

Variable	Total mortality				Neoplasm	
	Lag	Durbin			Lag	Durbin
	Estimate	Direct	Indirect	Total	Estimate	Direct
LR test <small>please align to the right</small>	33.05***	51.1***			11.135***	28.938***
AIC <small>please align to the right</small>	15,345	15,225			11,107	11,044
AIC for OLS <small>please align to the right</small>	15,376	15,275			11,116	11,071
LM test for residual autocorrelation <small>please align to the right</small>	24.2***	0.9			30.149***	1.27
Spatial weight matrix <small>please align to the right</small>	First-order Rook				First-order Rook	
Moran's I <small>please align to the right</small>	0.150***				0.134***	
Variable	Circulatory system diseases				External causes	
	Lag	Durbin			Lag	Durbin
	Estimate	Direct	Indirect	Total	Estimate	Direct
Intercept	77.980***			35.883*	- 18.013	
Temperature (°C)					0.595	10.387***
ln(Area in km ²)						
Rurality	- 0.225	- 0.993***	1.664***	0.689***		
Older population					- 2.406**	- 4.386**
Internal immigrants	- 0.756***	- 0.963**	1.591***	0.628 [†]	1.179***	1.217**
Afrodescendant					- 0.242	0.000
Mixed/white	0.613***	0.576***	-	0.576***		
Low educational level					0.178	1.403**
Investment in water supply	- 0.0001*	- 0.0001**	-	- 0.0001**		

[†] $p < 0.1$; * $p < .05$; ** $p < 0.01$; *** $p < 0.001$. The unit of analysis is Colombian municipalities. The age-standardized mortality rate per population of 100,000. The explanatory variable values are properly specified in the table. Public investment is in ~~thousands~~ ^{1000-s} of Colombian pesos per capita. See a definitions. In the spatial Durbin models, the indirect effect of an explanatory variable is only separ when the model fit improves (i.e., the AIC value decreases) as a result. If not, the variable only has equal to the total effect

(a)

Variable	Total mortality				Neoplasm	
	Lag	Durbin			Lag	Durbin
	Estimate	Direct	Indirect	Total	Estimate	Direct
No health insurance	1.265*	0.978***	-0.555	0.423*		
<i>Spatial rho</i> same as above (allian this and the names of the indicators below to the right)	0.284	0.326			0.611	0.599
LR test	44.2***	60.0***			364.42***	342.3***
AIC	12,940	12,868			13,661	13,619
AIC for OLS	12,983	12,926			14,023	13,960
LM test for residual autocorrelation	12.7***	0.3			4.425*	3.684
Spatial weight matrix	First-order rook				Nearest neighbor $k = 4$	
Moran's I	0.170***				0.485***	

(b)

Variable	Neoplasm				External	
	Lag	Durbin			Lag	Durbin
	Estimate	Direct	Indirect	Total	Estimate	Direct
<i>Intercept</i>	16.300			3.811	-7.979	
Temperature (°C)					-0.063	0.803
ln(Area in km ²)	5.600***	5.765***	-	5.765***	5.643***	4.148***
Rurality	-0.038	-0.208**	0.290***	0.082	0.112	0.117
ln(Pop. density)	4.238**	4.537***	-	4.537***		
Mixed/white	0.190***	0.185***	-	0.185***		
Low educational level	-0.518*	-0.415*	-	-0.415*	0.088	-0.061
Poverty					-0.288***	-0.331***
No health insurance	0.383***			0.426***		

† $p < 0.1$; * $p < .05$; ** $p < 0.01$; *** $p < 0.001$. The unit of analysis is Colombian municipalities. The age-standardized mortality rate per population of 100,000. The explanatory variable values are properly specified in the table. Public investment is in ~~thousands~~ ~~1000-s~~ of Colombian pesos per capita. See a definitions. In the spatial Durbin models, the indirect effect of an explanatory variable is only separ when the model fit improves (i.e., the AIC value decreases) as a result. If not, the variable only has equal to the total effect

(a)

Variable	Total mortality				Neoplasm	
	Lag	Durbin			Lag	Durbin
	Estimate	Direct	Indirect	Total	Estimate	Direct
Spatial rho same as above (align this and the names of the indicators below to the right)	0.170	0.249			0.347	0.318
LR test	16.7***	35.3***			67.0***	53.8***
AIC	10,881	10,823			10,652	10,637
AIC for OLS	10,896	10,856			10,717	10,689
LM test for residual autocorrelation	29.8***	0.4			11.1***	2.0
Spatial weight matrix	Nearest neighbor $k = 3$				First-order Rook	
Moran's I	0.199***				0.196***	

$\dagger p < 0.1$; $*p < .05$; $**p < 0.01$; $***p < 0.001$. The unit of analysis is Colombian municipalities. The dependent variable is the age-standardized mortality rate per population of 100,000. The explanatory variable values are proportions, unless otherwise specified in the table. Public investment is in thousands of Colombian pesos per capita. See Appendix 2 for exact definitions. In the spatial Durbin models, the indirect effect of an explanatory variable is only separated when the model fit improves (i.e., the AIC value decreases) as a result. If not, the variable only has an effect equal to the total effect

Table 5

OLS models for the causes of death that showed no or very little spatial autocorrelation

Variable	Men			Women		
	Infectious diseases	Perinatal mortality	Total mortality Please move the Total Mortality column to the results for Women	Infectious mortality	Circulatory system	Perinatal mortality
Intercept	-37.527***	-5.908	-14.031***	-7.757***	25.695*	-9.168*
Temperature (°C)		-0.857***	-8.724***			-0.840**
ln(Area in km ²)	5.211***	5.370***	42.623***	3.675***		3.810**

$*p < .05$; $**p < 0.01$; $***p < 0.001$. The unit of analysis is Colombian municipalities. The dependent variable is the age-standardized mortality rate per population of 100,000. The independent variable values are proportions, unless otherwise specified in the table, except for public investment (in thousands of Colombian pesos per capita). See Appendix 2 for exact definitions.

Variable	Men			Women		
	Infectious diseases	Perinatal mortality	Total mortality <small>Please move the Total Mortality column to the results for Women</small>	Infectious mortality	Circulatory system	Perinatal mortality
Main town	0.372***	0.145***	2.782***	0.203***	0.544**	0.202**
ln(Pop. density in km ²)					- 7.271***	
Older population	0.900			9.772***		5.028**
Internal immigrants				- 0.228***	- 0.744***	
Afrodescendant		0.137**				
Indigenous	0.474***	0.132*	0.814*	0.338***		
Mixed/white					0.291**	
Poverty						0.158**
Investment in health	- 0.048***	- 0.034***	- 0.177***	- 0.024***	- 0.042*	- 0.031**
Investment in education	0.225***	0.192***	0.876***	0.108***	0.185**	0.145**
Investment in water supply			- 0.223**		- 0.099**	
No health insurance				0.194***	1.014***	
Adjusted R-squared <small>align to the right</small>	0.21	0.17	0.26	0.20	0.17	0.15
AIC <small>align to the right</small>	10,769	10,306	14,427	10,183	12,691	9946

* $p < .05$; ** $p < 0.01$; *** $p < 0.001$. The unit of analysis is Colombian municipalities. The dependent variable is the age-standardized mortality rate per population of 100,000. The independent variable values are proportions, unless otherwise specified in the table, except for public investment (in thousands of 1000s Colombian pesos per capita). See Appendix 2 for exact definitions.

AQ2

Male total mortality shows a clear positive spatial association with the common characteristics of Colombia's urban population (mestizo/white background and health insured) and a negative association with rural areas. The municipalities with the largest surface area are characterized by a positive association. The direct negative effect of rural population on general mortality is more than nullified by a high indirect effect of rural municipalities that contribute a high positive association with this variable. This points to

an effect not measured in our model, namely the location of health care facilities in urban areas. The negative association as a direct effect of the coefficients linked to the migratory contribution (lower mortality in centers with little immigration) and positive as an indirect effect, which suggests a higher mortality in the high-migration municipalities that surround the reference municipality.

The importance of characteristics of neighboring municipalities is also observed for cancer, and for both sexes. The spatial Durbin model shows a greater concentration in urban areas (direct negative effect with rurality and positive with the availability of health insurance). On the contrary, the indirect effects associated with rurality are positive, which seems to indicate, as with general mortality, enclaves with a greater concentration of the population in urban centers that are surrounded by areas of greater rural population tend to have a greater incidence of this cause of mortality and may contribute to some of the cancer mortality in the surrounding urban areas where hospitals are located. In the case of women, it is also associated with population density and mixed/white race. These variables are associated with advanced stages of the epidemiological transition in the more developed areas of Colombia. On the other hand, in the case of men the spatial lag model for this cause is characterized by a negative association with the Afro-Colombian population—which has a much younger structure than the country as a whole—and with high-migration areas (these municipalities tend the most advanced in terms of economic and social development). On the contrary, the effect of the superficial extension of the municipality plays against it, unlike the other causes, although its significance is much lower.

Patterns of spatial dependence in male circulatory system disease are similar to those of neoplasms. The most significant direct effects are found in two characteristics associated with greater urban development, a greater proportion of the population being of mixed race/white race and with health insurance. The direct effect of rural population and immigration is in the same line as for cancer, but its indirect (neighborhood) effects are larger. Likewise, among the significant indirect effects, urban areas surrounded by more rural areas are characterized by higher circulatory system disease mortality.

~~Finally, among~~ With respect to the deaths linked to the external causes, there is a direct and indirect effect of temperature, which is highly negatively associated with altitude. There is of course no causal effect here, but results suggest that mortality is much higher in municipalities in the foothills (direct effect) surrounded by other municipalities at a higher altitude (indirect effect with a high negative association). This scheme corresponds to the classic model of mortality due to violent guerrilla warfare, which throughout the conflict in Colombia has been characterized by attacks on more populated municipalities in lower altitudes while maintaining their basecamps higher up in the Andean mountains. The effect of the low educational levels of some populations that have historically constituted a local guerrilla and paramilitary recruitment sector, very close to the military actions carried out by these violent groups, goes along the same lines. Another important component of external-cause mortality is traffic accidents, and municipalities characterized by a low average socioeconomic status (of which education is a marker) tend

to have higher rates of lethal motorcycle accidents (Martínez-Cabezas 2020). Effects are also stronger for men than for women. The exception is large municipalities in terms of area surrounded by likewise large municipalities (both direct and indirect effect, suggesting the extreme vulnerability of women in isolated areas).

Regarding the OLS results of the causes of death without statistically significant spatial autocorrelation, the explanatory variables explained between 15% (perinatal mortality, women) and 26% (total mortality, women) of the municipal variation in mortality (Table 5), although the AIC is relatively high in the latter model. This may suggest some overfitting. Turning to the specific results, there is some overlap with the earlier Mann–Whitney U tests that were performed on the HH and LL clusters. For instance, considering infectious diseases (for both sexes), the greater the surface area of the municipality, the higher the mortality. This can be explained by the fact infectious diseases remains an important health issue in the large and isolated municipalities of the Amazon area. These areas also have large concentrations of indigenous people but at the same time fewer internal migrants (women only) who are more likely to live in healthier, more developed and economically dynamic areas. Once controlling for all other significant factors, mortality levels are also higher in more aged municipalities (men only) and those with less investment in health, but also where investment is higher in education. The latter association can be explained by the fact that per capita investment is higher in less economically developed areas where perinatal mortality is high. The association with health and education investment is found for perinatal mortality and the same geographical and settlement variables are significant, with the addition of the temperature (although not in the expected positive association, but this is possibly because respiratory diseases are located in the most populated areas of the Andean mountain range with lower temperatures). In the case of males, mortality rates are also positively associated with the proportion of Afro-Americans and indigenous people.

Finally, turning to female total and circulatory system diseases, mortality is positively associated with living in a more urban context (higher proportion residing in the main municipal town), more elderly, but also an indigenous background (total mortality) and a mestizos/white background (circulatory system diseases). The latter is likely because among mestizos/white women a large proportion of circulatory system disease mortality is due to stroke, which is related to having survived to older ages. Finally, total and circulatory system mortality is also lower in municipalities where more is invested in health care and water treatment.

4. Discussion

The objective of this study was to analyze the possible existence of geographical clusters of mortality in Colombia, their evolution over time and local factors that may explain the geography of mortality from causes in Colombia between 1998 and 2014. Our results—based on spatial indicators and cartography—show that the geographical distribution of causes of death in Colombia has a significant degree of spatial autocorrelation in which male mortality from external causes stands out. Results also show that the applied data

corrections considerably improved the quality of the municipal data, leading to the identification of clusters that had already been previously identified in epidemiological studies which used morbidity registers. In this sense, the correction we applied is a valuable instrument of health policies.

Colombia experienced a general decrease in all-cause mortality over the 16-year period that was studied. This occurs in a context of some stability in the geographical patterns of mortality. From 2004 to 2006 among men, but already earlier among women as they are further ahead in the ET, cancer mortality began to contribute more significantly to the overall mortality pattern. At the same time, spatial autocorrelation between municipalities declined over time for all cause-of-death categories except male circulatory system diseases and perinatal mortality and was highest in external causes. The latter was especially the case among men, with hotspots moving from the central Andean area to Orinoco and the Amazon rainforest. The identification of geographical clusters of mortality in Colombia and their evolution over time will allow decision makers to prioritize those regions with higher mortality. On the other hand, particularly in LL cluster municipalities non-communicable diseases are now clearly dominating, denoting that they are clearly in the third stage of the ET.

We tested the two most important ones, cancer and circulatory system diseases, and according to our explanatory analysis both appeared to be significantly associated with the mestizo/white population who have the highest life expectancy. However, we believe that this link tells us more about the competition between causes than about a real statistical association. Municipalities with a younger population structure correspond to those with a significant presence of indigenous and Afrodescendants, where lower life expectancies are associated with mortality from other causes, particularly infectious diseases and perinatal mortality. On the other hand, HH external mortality clusters are more associated with geographical factors linked to isolation and the presence of poorly accessible jungle areas, where the low presence of the Colombian public administration instigated violence by different actors of armed conflicts in Colombia. Likewise, the negative association between total mortality, cancer, circulatory system disease and external causes with poverty reinforces the idea that urban Andean populations are on the one hand the most developed and richest, but on the other also the geographical spaces in Colombia where extreme inequality is manifested.

While the Mann–Whitney nonparametric test provides interesting and useful insights into mortality extremes, it does so for a small subset of municipalities and does not consider spatial correlation of the covariates. We therefore also performed spatial lag and spatial Durbin models. In particular, the interpretations of the latter results are richer than those of other conventional analytical approaches. To summarize the key findings, the spatial Durbin model is superior to the spatial lag model for all cause-of-death categories explored. Moreover, it allows the estimation of direct and indirect effects of covariates, which improves the overall spatial model. Indeed, results showed the existence of significant spillover effects, thus providing strong evidence to support our argument that the characteristics of surrounding municipalities are important determinants of mortality

for different types of causes. That said, this effect is less discerning for women than for men as well as certain causes of death (infectious diseases and perinatal mortality). On the other hand, the prominent role of different variables of a geographical nature, isolation, temperature/altitude, urbanization/rural population and their spatial lag as determining variables in mortality levels confirm the strong impact of the geographical context in explaining mortality patterns in Colombia.

In addition, the negative association between public investment in health and most causes of death except external causes in both the cluster analysis and the OLS models clearly shows the importance of efficiently providing widespread coverage and equitable access to health care. As summarized by Dávila-Cervantes and Agudelo-Botero (2018), this is to ensure that the most vulnerable populations (such as children, women, indigenous communities and persons of African descent) will have their basic needs covered. This would help to reduce inequalities that are mediated by variables such as gender, age, social status, education level and geographical area of residence, which have a differential and unjust effect on the health status of communities. But besides coverage, quality should also be the goal of the public services, with distinct goals for urban and rural areas. Although the decentralized transfers' system is complex in terms of structure and criteria of resource allocation, this would be facilitated by strengthening of the management capacity of the departments (Bonet et al. 2014).

A suggestion for future research is to examine more specific causes of death such as types of cancer, diabetes, homicides and traffic accidents. For instance, it is possible that a higher degree of spatial autocorrelation will be observed for some of these causes due to similar territorial patterns related to their etiology, such as nutritional factors (e.g., low fruit and vegetable and high salt intake are considered risk factors of stomach cancer) and smoking (the most important risk factor of lung cancer). Moreover, as the new political conditions after the 2012–2016 Colombian peace process also led to a considerable reduction in homicides, this might have spatial repercussions that we have not been able to observe. Given the obvious importance of disease risk factors (both proximate and indirect), spatial econometric models of an explanatory nature can be applied to test the effect on these specific causes of death with additional local indicators to those used here (e.g., related to homicide, the sex ratio, alcohol consumption, unemployment, housing conditions and income inequality, or including information on diet, hygiene, alcohol, smoking and access to health care when analyzing different types of cancer). Regarding the explanatory analysis, we do need to acknowledge its ecological nature, although the interpretation of results of sub-national studies (in our case, municipalities) is less subject to ecological bias because they are more homogeneous than countries. According to Higgs et al. (1998), the identification of statistical associations between mortality and exogenous factors at the small area level may even be interpreted in conjunction with and suggestive of epidemiological investigations at the microlevel.

The high heterogeneity of the 1118 Colombian **local administrative entities** ~~municipalities~~ **analyzed** suggests another possible route of investigation, namely ~~the~~ **grouping of municipalities** ~~them~~ into “sub-regions” that are situated between local authorities and

departments. This new territorial division would allow the absorption of at least part of the registration errors linked to the regional hospital infrastructure and also raise the statistical and spatial significance of the calculated indicators. Along the same lines, the elimination of the large and sparsely populated municipalities of the Amazon region may imply significant changes in subsequent analyses due to the “noise” they contribute in the calculation of the Global and Local Moran’s I. The choice of using one or another criterion of contiguity for establishing the neighborhood relations is something that we did evaluate and have tried different methods, the Rook criterion of order 1 was the one that offered the most robust and consistent results (Appendix 3). The fact that all high-value loadings were positive also implies that the potential issue of “competing causes” (i.e., the decrease in one cause of death leads to the increase in another) does not play a role in our spatial analysis. This is, however, no surprise as most causes of death that we analyzed have a specific age- and sex-structure. For instance, perinatal mortality only occurs around the time of birth and most external causes during late adolescence and early adulthood (ages 15–34). Even the possible competing causes circulatory system diseases and neoplasms did not result in a significant negative association. On the contrary, the association was clearly positive, but also this has a plausible explanation as regions with low levels of mortality tend to have low levels in both circulatory system diseases and neoplasms as their etiologies are similar (e.g., dietary and behavioral factors and medical technology).

Despite our interesting findings, some limitations of our study should be highlighted. Perhaps most importantly is the under-registration of the Colombian mortality data. Specific adjustments were thus applied. As a result, our adjusted mortality rates are quite consistent with those published by the INS/ONS (2013) and the Ministry of Health and Social Protection (2016) reported a decline in the adjusted death rate from 524 in 2005 to 444 per 100.000 in 2014. As a comparison, our adjusted rates for the period 2004–2006 were, respectively, 566 and 334 for men and women and for the period 2013–2014, respectively, 540 and 335.

Finally, the application of spatial regression models allows the introduction of a new conceptual element in the analysis of epidemiological transitions, namely geographical space. If we can operationalize this concept, we would have an indicator of “system convergence” that represents changes in cause-of-death patterns. To illustrate, using the results we obtained for Colombia: although there is a gradual reduction of mortality, this occurs without territorial convergence. That is to say, we could have epidemiological transition models with or without territorial convergence, which would add a geographical dimension to the analysis and description of mortality patterns, but we leave this for future research.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements

Financial support for this research came from the Spanish Ministry of Economy and Competitiveness under the projects “Geographical mobility and access to housing: Spain in international perspective” (Ref CSO2013-45358-R) and “Comportamientos demográficos y estrategias residenciales: apuntes para el desarrollo de nuevas políticas sociales” (CSO2016-79142-R) and the “Ramón y Cajal” Programme (RYC-2013-14851).

5. Appendix 1

See Table 6.

Table 6

Results of the linear regression between the original and corrected* age-standardized mortality rate in Colombian municipalities for the period 1998–2014 by sex and cause of death

Cause of death	Indicator	Men						Women	
		1998–2000	2001–2003	2004–2006	2007–2009	2010–2012	2013–2014	1998–2000	2001–2003
All-cause mortality	Slope	1.19	1.10	1.26	1.32	1.46	1.36	1.47	1.32
	Constant	102.34	101.56	63.63	37.24	6.32	46.02	41.73	43.89
	R^2	0.77	0.76	0.78	0.78	0.77	0.75	0.76	0.75
Infectious diseases	Slope	1.97	1.81	1.66	1.73	2.28	1.67	2.14	1.85
	Constant	-1.42	-0.75	1.50	-0.11	-4.98	-0.32	-1.47	-0.09
	R^2	0.71	0.77	0.74	0.78	0.80	0.69	0.72	0.76
Neoplasm	Slope	1.33	1.32	1.32	1.41	1.41	1.40	1.43	1.26
	Constant	5.66	3.24	3.74	0.79	2.61	3.09	4.40	5.62
	R^2	0.84	0.89	0.88	0.90	0.92	0.90	0.85	0.86
Circulatory system	Slope	1.30	1.29	1.30	1.39	1.41	1.38	1.45	1.31
	Constant	17.76	13.80	12.35	3.79	6.78	7.22	12.34	13.23
	R^2	0.82	0.85	0.88	0.90	0.90	0.90	0.83	0.86
Perinatal	Slope	2.51	2.80	2.88	3.15	4.15	3.74	2.57	2.67
	Constant	0.63	-0.13	0.33	0.08	-0.57	-0.04	0.19	-0.10
	R^2	0.75	0.74	0.73	0.72	0.80	0.77	0.76	0.75

Source: Own calculations based on the microdata obtained from Vital Statistics and the denominators

*See main text for how the deaths were corrected for the municipalities

Cause of death	Indicator	Men						Women	
		1998–2000	2001–2003	2004–2006	2007–2009	2010–2012	2013–2014	1998–2000	2001–2003
External causes	Slope	1.12	0.97	1.18	1.09	1.12	1.26	1.45	1.21
	Constant	24.02	29.25	16.69	22.08	18.21	7.19	2.77	4.94
	R^2	0.92	0.91	0.92	0.86	0.81	0.81	0.83	0.84
Other causes	Slope	1.36	1.36	1.41	1.45	1.71	1.42	1.55	1.39
	Constant	11.23	8.36	5.98	3.61	– 6.77	9.41	5.30	6.21
	R^2	0.82	0.83	0.84	0.86	0.85	0.81	0.86	0.82
Ill-defined causes	Slope	1.50	1.52	1.66	1.63	1.73	1.69	1.61	1.59
	Constant	1.25	0.20	– 0.24	– 0.40	– 0.66	– 0.66	1.31	0.12
	R^2	0.89	0.89	0.80	0.87	0.83	0.82	0.90	0.93
Municipalities	N	1112	1115	1115	1119	1118	1117	1112	1115

Source: Own calculations based on the microdata obtained from Vital Statistics and the denominato

*See main text for how the deaths were corrected for the municipalities

6. Appendix 2

See Table 7.

Table 7

Exogenous variables used in the explanatory analysis

Variable category	Specific variable ^h	Definition	Source
Geography	Temperature	Mean temperature (°C) in period 1981–2010	IDEAM ^a
	Area	Size of municipality in km ²	IGAC ^b
Settlement	Main town	Proportion resident in municipality's main town	2005 Census ^c

Sources: ^aInstituto de Hidrología, Meteorología y Estudios Ambiental, <https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Promedios-Precipitacion-y-Temperatura-media-Promed/nsxu-h2dh>. ^bInstituto Geográfico Agustín Codazzi, <http://igac.gov.co/>; ^cOwn calculations based on the census microdata data; ^dNational Administrative Department of Statistics, <https://www.dane.gov.co/>; ^ePrecise definition and methodology is described in <https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-vida/necesidades-basicas-insatisfechas-nbi>; ^f<https://www.dnp.gov.co/programas/inversiones-y-finanzas-publicas/Paginas/inversiones-y-finanzas-p%C3%BAblicas.aspx>; ^g<http://www.sispro.gov.co>. ^hMunicipal data on altitude was also obtained but not further analyzed due to its high correlation with temperature (> 0.95)

Variable category	Specific variable ^h	Definition	Source
Population	Rurality	Proportion resident in municipality's rural areas	2005 Census ^c
	Density	Number of inhabitants per km ²	DANE ^d , based on 2005 Census
	Fertility proxy	Rate per 1000 of children aged 0–4 to women aged 15–34	2005 Census ^c
	Older population	Proportion of persons aged 50+	2005 Census ^c
	Internal immigrants	Proportion of persons born in other municipalities	2005 Census ^c
Socioeconomic	Low educational level	Proportion of persons older than 20 with low education	2005 Census ^c
Status	Poverty	Proportion of households with unsatisfied basic needs ^e	DANE ^e , based on 2005 census
Ethnicity	Afrodescendant	Proportion of afrodescendants	2005 Census ^c
	Indigenous	Proportion of indigenous persons	2005 Census ^c
	Mixed/white	Proportion of mixed race and white people	2005 Census ^c
Public investment	Health	Investment in health, in Colombian pesos per capita (2004–2006)	DNP ^f
	Education	Investment in education, in Colombian pesos per capita (2004–2006)	DNP ^f
	Water supply	Investment in water treatment, in Colombian pesos per capita (2004–2006)	DNP ^f
Health insurance	Health insurance	Proportion of people with access to health care through social security contributions (2009–2011)	Ministerio de salud y protección social ^g

Sources: ^aInstituto de Hidrología, Meteorología y Estudios Ambiental, <https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Promedios-Precipitacion-y-Temperatura-media-Promed/nsxu-h2dh>. ^bInstituto Geográfico Agustín Codazzi, <http://igac.gov.co/>; ^cOwn calculations based on the census microdata data; ^dNational Administrative Department of Statistics, <https://www.dane.gov.co/>; ^ePrecise definition and methodology is described in <https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-vida/necesidades-basicas-insatisfechas-nbi>; ^f<https://www.dnp.gov.co/programas/inversiones-y-finanzas-publicas/Paginas/inversiones-y-finanzas-p%C3%BAblicas.aspx>; ^g<http://www.sispro.gov.co>. ^hMunicipal data on altitude was also obtained but not further analyzed due to its high correlation with temperature (> 0.95)

7. Appendix 3

See Table 8.

Table 8

Variable medians in the high (H–H) and low (L–L) mortality clusters and level of significance according

Variable	Total mortality			Infectious diseases			Neoplasm		
	H-H	L-L	Sig.	H-H	L-L	Sig.	H-H	L-L	Sig.
<i>Men</i>									
Geography									
Temperature in °C	22.0	28.6	0.000	24.0	19.3	0.073	18.4	28.0	0.000
Area in Km2	455.5	358.5	0.028	626.0	168.0	0.000	195.0	829.0	0.000
Settlement									
Main town	43.8	40.0	0.268	35.7	26.5	0.259	49.2	33.3	0.000
Rurality	45.8	36.6	0.022	60.2	66.7	0.244	43.0	55.9	0.020
Density	28.2	36.7	0.026	21.9	27.5	0.269	87.9	16.8	0.000
Population									
Fertility proxy		N/A		858.1	728.4	0.009		N/A	
Older population	16.9	15.5	0.024	14.5	18.9	0.000	18.6	14.2	0.000
Internal immigrants	34.0	14.4	0.000	16.8	22.3	0.161	35.4	16.8	0.000
Socioeconomic status									
Low educational level	83.0	81.7	0.270	82.5	86.6	0.044	77.2	85.8	0.000
Poverty	37.4	63.2	0.000	62.6	47.0	0.012	26.6	68.0	0.000
Ethnicity									
Afrodescendant	1.68	2.92	0.036	1.49	0.20	0.003	0.98	2.86	0.020
Indigenous	0.14	0.05	0.265	15.14	0.01	0.000	0.05	0.26	0.010
Mixed/white	95.84	88.52	0.004	54.73	98.20	0.000	96.73	85.06	0.000
Public investment									
Health	92,266	86,781	0.489	81,101	109,407	0.021	64,655	103,595	0.000
Education	15,406	20,753	0.000	20,510	15,339	0.000	13,304	22,755	0.000
Water supply	35,397	38,709	0.432	37,233	66,641	0.005	25,954	49,742	0.000
Health insurance									
Health insurance	22.4	14.0	0.000	17.8	11.5	0.212	34.9	10.8	0.000

^a I noticed that the values that were in red in the last proof are now black. In any case, as the paper version will be in black and white, it is best to j
~~the 0.05 level.~~ N/A: Not applicable (only the association between the fertility proxy and infectious c

Variable	Total mortality			Infectious diseases			Neoplasm		
	H–H	L–L	Sig.	H–H	L–L	Sig.	H–H	L–L	Sig.
N° of municipalities	80	102	182	31	39	70	61	63	124
% of all municipalities	7.1%	9.1%	16.3%	2.8%	3.5%	6.3%	5.4%	5.6%	11.1%
<i>Women</i>									
Geography									
Temperature in °C	18.4	28.5	0.000	25.9	19.3	0.063	19.8	28.3	0.000
Area in km ²	423.0	543.0	0.182	764.5	241.0	0.000	261.5	829.0	0.000
Settlement									
Main town	36.6	38.8	0.799	38.6	24.9	0.106	49.8	32.2	0.000
Rurality	55.3	44.9	0.050	47.2	64.1	0.119	43.8	55.9	0.010
Density	39.3	26.1	0.590	20.9	33.7	0.092	66.1	14.7	0.000
Population									
Older population	13,029	14,007	0.535	18,099	8586	0.000	18,315	9708	0.000
Fertility proxy		N/A		826.3	704.0	0.016		N/A	
Older population	17.6	14.9	0.001	14.6	18.3	0.000	17.9	13.2	0.000
Internal immigrants	19.8	18.1	0.036	25.9	25.0	0.773	35.9	18.8	0.000
Socioeconomic status									
Low educational level	83.7	81.9	0.312	82.5	87.0	0.006	79.9	85.6	0.000
Poverty	40.2	64.1	0.000	53.8	46.9	0.382	30.3	71.3	0.000
Ethnicity									
Afrodescendant	0.43	7.06	0.000	1.14	0.33	0.027	1.90	10.02	0.000
Indigenous	0.15	0.21	0.595	11.90	0.04	0.000	0.13	0.29	0.310
Mixed/white	94.84	74.55	0.005	81.35	98.86	0.000	95.75	39.78	0.000
Public investment									
Health	100,634	90,183	0.753	79,530	98,141	0.119	74,783	102,714	0.000
Education	15,696	21,239	0.000	18,886	16,168	0.056	14,819	22,750	0.000

^a I noticed that the values that were in red in the last proof are now black. In any case, as the paper version will be in black and white, it is best to j
~~the 0.05 level.~~ N/A: Not applicable (only the association between the fertility proxy and infectious c

Variable	Total mortality			Infectious diseases			Neoplasm		
	H–H	L–L	Sig.	H–H	L–L	Sig.	H–H	L–L	Sig.
Water supply	35,358	43,813	0.832	32,505	52,208	0.002	28,255	53,012	0.00
Health insurance									
Health insurance	18.9	13.1	0.020	16.6	11.3	0.041	31.3	10.9	0.00
N° of municipalities	33	56	89	38	24	62	54	61	115
% of all municipalities	2.9%	5.0%	7.9%	3.4%	2.1%	5.5%	4.8%	5.4%	10.3%

^a I noticed that the values that were in red in the last proof are now black. In any case, as the paper version will be in black and white, it is best to j
~~the 0.05 level.~~ N/A: Not applicable (only the association between the fertility proxy and infectious c

8. Appendix 4

See Table 9.

Table 9

Moran's I of SMR in Colombia municipalities (2004–2006) according to different contiguity criteria

	Rook 1	Queen 1	Nearest neighbor			
			K = 3	K = 4	K = 5	K = 6
<i>Men</i>						
All causes	<u>0.150***</u>	0.150***	0.189***	0.191***	0.190***	0.175***
Infectious diseases	0.042*	0.041 [†]	0.060**	0.060**	0.050**	0.056***
Neoplasms (cancer)	<u>0.134***</u>	0.134***	0.158***	0.144***	0.145***	0.138***
Circulatory system diseases	<u>0.170***</u>	0.170***	0.180***	0.181***	0.185***	0.182***
Perinatal mortality	0.032 [†] +	0.033 [†]	0.039	0.039*	0.040*	0.041**

Source: Own calculations based on the microdata obtained from Vital Statistics and the denominators supplied by DANE

[†] I've inserted 4 more †. ... $p < 0.1$; * $p < .05$; ** $p < 0.01$; *** $p < 0.001$. Highest values in **bold**. Underlined value signals the weight matrix chosen for the Spatial lag and Spatial Durbin models. Tests were also performed for second-order Rook and Queen (with and without including the lower order), but results were either similar or worse, so are not shown here (but can be obtained from the authors upon request). Although nearest neighbor may have a higher Moran's I value than Rook 1, Rook 1 is more reliable to use these when spatial polygons differ greatly in terms of size, which is the case in the Colombian context (e.g., compare the very vast Amazonian region with the many small-sized municipalities of the Andean region). Rook 1 was therefore chosen when Moran's I was high and reasonably comparable with a more complex spatial weight matrix. On two occasions a different weight matrix was chosen (external causes-males and cancer-women). Spatial autocorrelation was not considered to be present in the remaining causes of death

	Rook 1	Queen 1	Nearest neighbor			
			K = 3	K = 4	K = 5	K = 6
External causes	0.450***	0.440***	0.478***	<u>0.485***</u>	0.454***	0.436***
<i>Women</i>						
All causes	0.016	0.016	0.055	0.055*	0.054*	0.046*
Infectious diseases	0.015	0.014	0.022	0.019	0.018	0.024
Neoplasms (cancer)	0.124***	0.123***	<u>0.199***</u>	0.178***	0.167***	0.152***
Circulatory system diseases	0.016	0.015	0.060 [†]	0.045 ^{†‡}	0.045 ^{†‡}	0.041 ^{†‡}
Perinatal mortality	0.015	0.016	0.021	0.014	0.006	0.006
External causes	<u>0.196***</u>	0.196***	0.228***	0.213***	0.197***	0.191***

Source: Own calculations based on the microdata obtained from Vital Statistics and the denominators supplied by DANE

[†] I've inserted 4 more †. $p < 0.1$; * $p < .05$; ** $p < 0.01$; *** $p < 0.001$. Highest values in **bold**. Underlined value signals the weight matrix chosen for the Spatial lag and Spatial Durbin models. Tests were also performed for second-order Rook and Queen (with and without including the lower order), but results were either similar or worse, so are not shown here (but can be obtained from the authors upon request). Although nearest neighbor may have a higher Moran's I value than Rook 1, Rook 1 is more reliable to use these when spatial polygons differ greatly in terms of size, which is the case in the Colombian context (e.g., compare the very vast Amazonian region with the many small-sized municipalities of the Andean region). Rook 1 was therefore chosen when Moran's I was high and reasonably comparable with a more complex spatial weight matrix. On two occasions a different weight matrix was chosen (external causes-males and cancer-women). Spatial autocorrelation was not considered to be present in the remaining causes of death

AQ3

9. Appendix 5

See Table 10.

Table 10

Life expectancy at birth in the Colombian departments. 2005

Department	Men	Women	Difference
Bogotá	73.81	79.30	5.49
Atlántico	71.84	77.58	5.74
Sucre	71.68	76.87	5.19
San Andrés	71.34	77.03	5.69

Source: DANE (2007)

The Amazon group refers to the departments of the Amazon, Guainía, Guaviare, Vichada. A more detailed analysis of sex differences in life expectancy at the departmental level can be found in Cristancho (2017; pp. 215–272)

Department	Men	Women	Difference
Bolívar	71.28	76.37	5.09
Boyacá	71.01	76.72	5.71
Magdalena	70.86	75.71	4.85
Santander	70.77	77.17	6.40
Cundinamarca	70.70	77.00	6.30
La Guajira	70.70	77.31	6.61
Córdoba	70.37	75.88	5.51
Nariño	70.34	75.98	5.64
Colombia	70.20	77.11	6.91
Huila	69.51	75.16	5.65
César	69.36	75.48	6.12
Quindío	69.32	75.87	6.55
Antioquia	69.26	76.83	7.57
Valle del Cauca	68.49	77.66	9.17
Tolima	68.33	75.40	7.07
Caldas	67.70	76.79	9.09
Norte de Santander	67.70	75.51	7.81
Cauca	67.36	73.92	6.56
Putumayo	66.85	74.82	7.97
Risaralda	66.80	76.47	9.67
Meta	66.51	74.34	7.83
Casanare	65.23	72.20	6.97
Arauca	64.80	72.96	8.16
Grupo Amazonía	64.07	74.46	10.39
Caquetá	63.58	72.45	8.87
Chocó	63.07	71.93	8.86

Source: DANE (2007)

The Amazon group refers to the departments of the Amazon, Guainía, Guaviare, Vichada. A more detailed analysis of sex differences in life expectancy at the departmental level can be found in Cristancho (2017; pp. 215–272)

10. Appendix 6

See Table 11.

Table 11

Global Moran's I by Rook First-Order Weight Contiguity of uncorrected and corrected SMR. Colombian municipalities. 1998–2000 and 2013–2014

	1998–2000		2013–2014	
	Uncorrected	Corrected	Uncorrected	Corrected
<i>Men</i>				
All causes	0.3945*	0.1916*	0.2121*	0.1160*
Infectious diseases	0.0488	0.1333*	0.0709*	0.0948*
Neoplasms (cancer)	0.2689*	0.1490*	0.1901*	0.1288*
Circulatory system diseases	0.2521*	0.1130*	0.1925*	0.1212*
Perinatal mortality	0.0017	0.0200	0.0029	0.0343
External causes	0.4763*	0.4077*	0.3916*	0.3191*
Remaining causes	0.2306*	0.1081*	0.1120*	0.0619*
Ill-defined causes	0.1489*	0.1315*	0.0797*	0.1041*
<i>Women</i>				
All causes	0.2765*	0.1367*	0.1375*	0.0945*
Infectious diseases	0.0612*	0.1251*	0.0404	0.0814*
Neoplasms (cancer)	0.2962*	0.1708*	0.1799*	0.1341*
Circulatory system diseases	0.2790*	0.1473*	0.1480	0.0866*
Perinatal mortality	– 0.0123	0.0208	0.0305	0.0696*
External causes	0.2498*	0.2267*	0.0835*	0.0875*
Remaining causes	0.2633*	0.1748*	0.0557	0.0535*
Ill-defined causes	0.2313*	0.2055*	0.0557	0.0574*
<i>Source:</i> Own calculations based on the microdata obtained from Vital Statistics and the denominators supplied by DANE				
* < 0.001				

References

Acevedo Bohórquez I, Velásquez Ceballos E (2008) Algunos conceptos de la econometría espacial y el análisis exploratorio de datos espaciales. *Ecos de Economía* 279–234

Akaike H (1973) Formation theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F (eds) *Second international symposium on information theory*. Akadémiai Kiadó, Budapest, pp 267–281

Anselin L (1988) Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical analysis* 20(1):1–17

Anselin L (1995) Local indicators of spatial association—LISA. *Geogr Anal* 27(2):93–115

Arroyave I, Burdorf A, Cardona D, Avendano M (2014) Socioeconomic inequalities in premature mortality in Colombia, 1998–2007: the double burden of non-communicable diseases and injuries. *Prev Med* 64:41–47

Bivand R, Piras G (2015) Comparing implementations of estimation methods for spatial econometrics. *J Stat Softw* 63(18):1–36. <https://doi.org/10.18637/jss.v063.i18>

Bonet J, Pérez G, Ayala J (2014) Contexto histórico y evolución del SGP en Colombia. Centro de Estudios Económicos Regionales (CEER), Cartagena.

https://www.banrep.gov.co/sites/default/files/publicaciones/archivos/dtser_205.pdf

Cristancho C (2017) Niveles, tendencias y determinantes de la mortalidad reciente en Colombia. PhD thesis, Universitat Autònoma de Barcelona.

https://ddd.uab.cat/pub/tesis/2017/hdl_10803_406099/ccf1de1.pdf

Cyril S, Oldroyd JC, Renzaho A (2013) Urbanisation, urbanicity, and health: a systematic review of the reliability and validity of urbanicity scales. *BMC Public Health* 13(1):513. <https://doi.org/10.1186/1471-2458-13-513>

DANE (Departamento Administrativo Nacional de Estadística) (2006) La mortalidad materna y perinatal en Colombia en los albores del siglo XXI.

<http://www.paho.org/relacsis/index.php/biblioteca-usuarios/publicaciones-relacsis/at11-cobertura-de-nacimientos-y-defunciones/323-col-unfpa-dane-estudiocabalidad-mm-2003/file>. Accessed 14 Nov 2018

DANE (Departamento Administrativo Nacional de Estadística) (2007) Proyecciones de población 2005-2020. Colombia. Tablas abreviadas de mortalidad nacionales y departamentales 1985–2020.

http://www.dane.gov.co/files/investigaciones/poblacion/proyepobla06_20/8Tablasvida1985_2

DANE (Departamento Administrativo Nacional de Estadística) (2014) Proyecciones de Población municipal. <https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion>

Dávila-Cervantes C, Agudelo-Botero M (2018) Changes in life expectancy due to avoidable and non-avoidable deaths in Argentina, Chile, Colombia and Mexico, 2000–2011. *Cadernos de Saúde Pública* 34(6):e00093417

Diez Roux AV, Green Franklin T, Alazraqui M, Spinelli H (2007) Intraurban variations in adult mortality in a large Latin American city. *J Urban Health* 84(3):319–333. <https://doi.org/10.1007/s11524-007-9159-5>

Feng X, Tan X, Alenzi EO, Rai P, Chang J (2016) Spatial and temporal variations of screening for breast and colorectal cancer in the United States, 2008 to 2012. *Medicine* 95(51):1–7

Fiscella K, Franks P (1997) Poverty or income inequality as predictor of mortality: longitudinal cohort study. *BMJ* 314(7096):1724

Flórez C, Méndez R (1997) Estimaciones de cobertura del registro de defunciones. Colombia 1993. Informe al Ministerio de Salud

Fox J, Weisberg S (2018) *An R companion to applied regression*, 3rd edn. Sage publications, Thousand Oaks

Franco-Agudelo S (1997) Violence and health in Colombia. *Pan Am J Public Health* 2(3):170–180. <https://doi.org/10.1590/S1020-49891997000900002>

Frenk J, Bobadilla JL, Stern C, Frejka T, Lozano R (1991) Elements for a theory of the health transition. *Health Trans Rev* 1(1):21–38

Heinze G, Wallisch C, Dunkler D (2018) Variable selection—A review and recommendations for the practicing statistician. *Biometrical J* 60(3):431–449. <https://doi.org/10.1002/bimj.201700067>

Higgs G, Senior ML, Williams HC (1998) Spatial and temporal variation of mortality and deprivation 1: widening health inequalities. *Environ Plan A* 30(9):1661–1682

Holmes CJ, Zajacova A (2014) Education as “the great equalizer”: health benefits for black and white adults. *Soc Sci Q* 95(4):1064–1085

INS/ONS (Instituto Nacional de Salud/Observatorio Nacional de Salud) (2014a) Cuarto Informe: Violencia Homicida en Colombia. Instituto Nacional de Salud, Bogotá, D.C., Colombia.

<https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/IA/INS/INFORME%20I>

INS/ONS (Instituto Nacional de Salud/Observatorio Nacional de Salud) (2014b) Tercer Informe. Mortalidad Evitable en Colombia. Instituto Nacional de Salud, Bogotá, D.C., Colombia.

<https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/IA/INS/Informe3-ME-ONS-Definitivo.pdf>

INS/ONS (Instituto Nacional de Salud/Observatorio Nacional de Salud), Gaviria Uribe A, Muñoz Muñoz N, Ruiz Gómez F, Ospina Martínez M, De la Hoz Restrepo F, Restrepo Trujillo C (2013) Primer Informe ONS, aspectos relacionados con la frecuencia de uso de los servicios de salud, mortalidad y discapacidad en Colombia, 2011. Instituto Nacional de Salud, Bogotá, D.C., Colombia.
<https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/IA/INS/informefinal.pdf>

Jaba E, Balan CB, Robu I-B (2014) The relationship between life expectancy at birth and health expenditures estimated by a cross-country and time-series analysis. *Procedia Econ Finance* 15:108–114

Kim J, Lee M, Jung I (2016) A comparison of spatial pattern detection methods for major cancer mortality in Korea. *Asia Pac J Public Health* 28(6):539–553.
<https://doi.org/10.1177/1010539516657871>

Kuo T-M, Meyer AM, Baggett CD, Olshan AF (2019) Examining determinants of geographic variation in colorectal cancer mortality in North Carolina: a spatial analysis approach. *Cancer Epidemiol.* <https://doi.org/10.1016/j.canep.2019.01.002>

Lanier C, Maume MO (2009) Intimate partner violence and social isolation across the rural/urban divide. *Viol Against Women* 15(11):1311–1330

LeSage J, Pace R (2009) Introduction to spatial econometrics. Taylor and Francis, Boca Raton. <https://doi.org/10.1201/9781420064254>

Lin C-Y, Hsu C-Y, Gunnell D, Chen Y-Y, Chang S-S (2019) Spatial patterning, correlates, and inequality in suicide across 432 neighborhoods in Taipei City, Taiwan. *Soc Sci Med.* <https://doi.org/10.1016/j.socscimed.2018.12.011>

Lloyd-Sherlock P, Minicuci N, Beard J, Chatterji S (2012) Social protection and preventing illness in developing countries: establishing the health effects of pensions and health insurance. *Int Soc Secur Rev* 65(4):51–68. <https://doi.org/10.1111/j.1468-246X.2012.01447.x>

López E, Arce P (2008) Efectos de las causas de mortalidad adulta en la esperanza de vida, entre 1985 y 1999, según regiones colombianas. *Biomédica* 28(3):414–422

Martínez-Cabezas SP (2020) Cuatro ensayos sobre la mortalidad por causas externas en Colombia. PhD thesis, Universitat Autònoma de Barcelona

Meliker JR, Sloan CD (2011) Spatio-temporal epidemiology: principles and opportunities. *Spatial Spatio-temporal Epidemiol* 2(1):1–9

Ministry of Health and Social Protection (2016) Análisis de Situación de Salud. Colombia, 2016. Ministerio de Salud y Protección Social, Bogotá.

<https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/ED/PSP/asis-colombia-2016.pdf>. Accessed 06 Nov 2018

Misnaza SP, Roncancio CP, Peña IC, Prieto FE (2016) Áreas de alta mortalidad perinatal debida a anomalías congénitas: análisis de estadísticas vitales, Colombia, 1999–2008. *Biomédica* 36(3):359–367. <https://doi.org/10.7705/biomedica.v36i3.2867>

Murillo R, Piñeros M, Hernández G (2004) Atlas de mortalidad por cáncer en Colombia. Instituto Nacional de Cancerología, Instituto Geográfico Agustín Codazzi, Bogotá

Nazroo JY (1998) Genetic, cultural or socio-economic vulnerability? Explaining ethnic inequalities in health. *Sociol Health Illn* 20(5):710–730. <https://doi.org/10.1111/1467-9566.00126>

Ochoa J, Osorio L (2006) Epidemiology of urban malaria in Quibdó, Chocó. *Biomedica* 26(2):278–285

Omran AR (1971) The epidemiologic transition: a theory of the epidemiology of population change. *Milbank Memor Fund Q* 49(4):509–538

Otero A (2013) Diferencias departamentales en las causas de mortalidad en Colombia. Documentos de Trabajo sobre Economía Regional del Banco de la República de Colombia 187. Centro de Estudios Económicos Regionales (CEER), Cartagena. https://www.banrep.gov.co/sites/default/files/publicaciones/archivos/dtser_187_0.pdf

Padilla JC, Lizarazo FE, Murillo OL, Mendigaña FA, Pachón E, Vera MJ (2017) Epidemiología de las principales enfermedades transmitidas por vectores en Colombia, 1990–2016. *Biomédica*. <https://doi.org/10.7705/biomedica.v37i0.3769>

Patz JA, Olson SH (2006) Malaria risk and temperature: influences from global climate change and local land use practices. *Proc Natl Acad Sci* 103(15):5635–5636

Pérez-Vidal A, Torres-Lozada P, Escobar-Rivera J (2016) Hazard identification in watersheds based on water safety plan approach: case study of Cali-Colombia. *Environ Eng Manag J* 15(4):861–872

Piñeros M, Murillo R (2004) Incidencia de cáncer en Colombia: importancia de las fuentes de información en la obtención de cifras estimativas. *Rev Colomb Cancerol* 8(1):5–14

Piñeros M, Ferlay J, Murillo R (2006) Cancer incidence estimates at the national and district levels in Colombia. *Salud Pública de México* 48(6):455–465

Piñeros M, Gamboa O, Hernández-Suárez G, Pardo C, Bray F (2013) Patterns and trends in cancer mortality in Colombia 1984–2008. *Cancer Epidemiol* 37(3):233–239

Piuevam G, Medeiros WR, Costa AV, Emerenciano FF, Santos RC, Seabra DS (2015) Mortality from cardiovascular diseases in the elderly: comparative analysis of two five-year periods. *Arq Bras Cardiol* 105(4):371–380. <https://doi.org/10.5935/abc.20150096>

Quintero J et al (2014) Ecological, biological and social dimensions of dengue vector breeding in five urban settings of Latin America: a multi-country study. *BMC Infect Dis* 14(1):38

Roberson S, Dutton M, Macdonald M (2016) Does place of residence or time of year affect the risk of stroke hospitalization and death? A descriptive spatial and temporal epidemiologic study. *PLoS ONE* 11(1):1–13. <https://doi.org/10.1371/journal.pone.0145224>

Rodrigues M, Bonfim C, Portugal JL, Gurgel IGD, Medeiros Z (2013) Using spatial analysis to identify areas vulnerable to infant mortality. *Revista Panamericana de Salud Pública* 34(1):36–40

Rodríguez-Acosta S (2016) La desigualdad de ingreso y salud en Colombia [Income inequality and health in Colombia]. *Perfiles Latinoamericanos* 24(48):265–296

Rodríguez-García J (2000) Descripción de la mortalidad por departamentos. Colombia año 2000. Cendex, Pontificia Universidad Javeriana. <https://www.javeriana.edu.co/cendex/pdf/DT%20016-05.pdf>. Accessed 14 Nov 2018

Rodríguez-García J (2007) Desigualdades socioeconómicas entre departamentos y su asociación con indicadores de mortalidad en Colombia en 2000. *Revista Panamericana de Salud Pública* 21(2/3):111–124

Rodríguez-García J, Peñaloza-Quintero RE, Amaya-Lara JL (2017) Estimación de la carga global de enfermedad en Colombia 2012: nuevos aspectos metodológicos. *Revista de Salud Pública* 19:235–240

Salas-Zapata WA, Ríos-Osorio LA, López-Hernandez LF, Gómez-Arias RD (2014) Sustainability analysis of the malaria control policy in the municipality of El Bagre, Colombia. *Revista Gerencia y Políticas de Salud* 13(27):128–146

Santos MB et al (2019) Clinical and epidemiological indicators and spatial analysis of leprosy cases in patients under 15 years old in an endemic area of Northeast Brazil: an ecological and time series study. *BMJ Open* 9(7):e023420

Spijker J (2004) Socioeconomic determinants of regional mortality differences in Europe. PhD Thesis, University of Groningen

Thompson SK, Gartner R (2014) The spatial distribution and social context of homicide in Toronto's neighborhoods. *J Res Crime Delinq* 51(1):88–118

Vargas YGS, Schmalbach JHE (2013) Inequidades en mortalidad materna por departamentos en Colombia para los años 2000–2001, 2005–2006 y 2008–2009. *J Public Health* 15(4):577–588

Walker RS, Kesler DC, Hill KR (2016) Are isolated indigenous populations headed toward extinction? *PLoS ONE* 11(3):e0150987

Windenberger F, Rican S, Jouglu E, Rey G (2012) Spatiotemporal association between deprivation and mortality: trends in France during the nineties. *Eur J Public Health* 22(3):347–353. <https://doi.org/10.1093/eurpub/ckr029>

Yang TC, Noah AJ, Shoff C (2015) Exploring geographic variation in US mortality rates using a spatial Durbin approach. *Population Space Place* 21(1):18–37

¹ Since 1998, this number has fluctuated between 1112 and 1122. See also Appendix 1.

² El Encanto, La Chorrera, La Pedrera, La Victoria, Miriti-Paraná, Puerto Alegria, Puerto Arica, Puerto Santander, Tarapaca, Cacahual, La Guadalupe, Morichal, Pana, Puerto Colombia, San Felipe, Pacoa, Papunaua and Yavarate.

³ The latter period only covers 2 years as no data could be obtained for 2015 at the time of conducting the analysis.

⁴ Although we corrected the mortality data for 1115 local administrative entities in Colombia for the 2004–2006 period (see Appendix 1), data on the selected explanatory variables were not available for the 18 non-municipalized areas and 3 municipalities. In addition, another 7 municipalities were excluded because they did not have any neighboring entities.

⁵ For the purpose of the study, the variable is broken down into Afrodescendants (mainly located on the Atlantic coast and Pacific coasts), the indigenous population (distributed throughout the country but particularly in jungle areas in enclaves *-resguardos-* and generally populated by members of the same ethnic group) and mestizos/whites (the country's economically and politically dominant group and overrepresented in the country's urban areas).

⁶ Calculations were performed using the software GeoDa (version 1.8.14) that was originally designed by Luc Anselin.

⁷ The null hypothesis (H_0) is stated as follows: There are no differences between the median values of the independent variables between the two groups of positive spatial autocorrelation. If p value ≤ 0.05 the H_0 is rejected.

⁸ For this analysis we used the spatial regression software *spatialreg* in R, see [www://cran.r-project.org](http://www.cran.r-project.org) (Bivand and Piras 2015).