RELAP - Revista Latinoamericana de Población
Vol. 15 - Número 29: 211-233
http://doi.org/10.31406/relap2021.v15.i2.n29.8
ISSN 2393-6401

Evaluation of the death records quality in Brazil: Sociodemographic determinants of incomplete education information

Evaluación de la calidad del registro de defunciones en Brasil: determinantes sociodemográficos de información incompleta sobre la educación

Walter Pedro Silva Júnior
ORCID: 0000-0002-6908-6995
wp_jr@hotmail.com
Statistician at the Universidade Federal do Rio Grande do Norte with expertise in data science, Brazil

Flávio Henrique Miranda de Araujo Freire
ORCID: 0000-0002-7416-9947
flavio.freire@ufrn.br
Titular professor at the Demography and Actuarial Sciences Department and coordinator at the Laboratory of Estimates and Population Projections at the Universidade Federal do Rio Grande do Norte, Brazil

Damião Nóbrega da Silva
ORCID: 0000-0003-3031-0870
damiao.silva@ufrn.br
Associate professor at the Department of Statistics of the Universidade Federal do Rio Grande do Norte, Brazil

Marcos Roberto Gonzaga
ORCID: 0000-0002-6088-3453
marcos.gonzaga@ufrn.br
Associate professor at the Demography and Actuarial Sciences Department and coordinator at the Demography Graduate Programme (Programa de Pós-Graduação em Demografia) at the Universidade Federal do Rio Grande do Norte, Brazil

Abstract

This paper aimed to evaluate the relationship between the lack of filling in education level information of adult deaths in the Brazilian Mortality Information System (SIM, by its Portuguese acronym) with sociodemographic characteristics of the deceased. Using sociodemographic information from SIM, a descriptive temporal analysis of the percentages of missing data in the education variable was performed. Next, the relationship between the lack of completion for education information was evaluated with sociodemographic variables. It was observed that men, at older ages, white, widowed, who died on public roads and due to external causes have a greater chance for failure in filling out the education information. Furthermore, it was found that the chance of presenting missing data is lower in the northern region compared to the other regions. It is generally concluded that the distribution of failure to complete the education information for deaths registered in the SIM is not uniform across spatial, socioeconomic and demographic characteristics.

Keywords

Mortality registries
Death certificates
Educational status
Data quality

Resumen

Este trabajo tuvo como objetivo evaluar la relación entre la falta de llenado de la información del nivel de educación de las muertes de adultos en el Sistema de Información de Mortalidad Brasileño (SIM, por su sigla en portugués) sobre las características sociodemográficas de los fallecidos. A partir de la información sociodemográfica del SIM, se realizó un análisis descriptivo temporal de los porcentajes de datos faltantes en la variable educación. A continuación, se evaluó la relación entre la falta de información sobre la educación y las variables sociodemográficas. Se observó que los hombres de mayor edad, raza blanca, viudos, fallecidos en la vía pública y por causas externas tienen una mayor probabilidad de tener incompleta su información sobre educación. Además, se comprobó que la probabilidad de presentar datos faltantes es menor en la región norte en comparación con las demás regiones. En general, se concluye que la distribución de la falta de llenado de la información sobre educación en las defunciones registradas en el SIM no es uniforme según las características espaciales, socioeconómicas y demográficas.

Palabras Clave

Registros de mortalidad
Actas de defunción
Escolaridad
Calidad de datos

Recibido: 09/03/2021
Aceptado: 08/06/2021

Introduction

Schooling has long been identified as one of the most important dimensions of socioeconomic status to study differentials in mortality (Kitagawa & Hauser, 1973; Muller, 2002; Preston & Taubman, 1994). If schooling is a dimension of and demographic and socioeconomic status, mortality and epidemiological transitions occur unevenly across social strata, then less educated people are expected to be more "lagging behind" in these change processes compared to more educated ones. For example, the less educated tend to die earlier from external causes (homicide) than the more educated (Filho et al., 2007).

Travassos et al. (2006) assessed these inequalities in access to health services in large regions, according to income and education in the period 1998 and 2003, using data from Pesquisa Nacional por Amostra de Domicílios (Brazilian National Household Surveys or PNADs, by its Portuguese acronym). The results indicated that access is strongly influenced by these variables. Individuals who lived in regions with a higher degree of socioeconomic development (Southeast and South) had greater access to health services compared to those who lived in less developed regions, in addition to a worsening in regional inequality across time. On the other hand, social inequalities decreased in the period. Although income had a greater influence than schooling, the chance that an individual with 9 years or more of study had used health services in Brazil was 20.9% higher than one from 0 to 4 years of schooling (Travassos et al., 2006).

Education is a variable that captures the differentials of the population's socioeconomic status well. Analyzing mortality based on educational levels makes it possible to make important inferences for the Brazilian population. However, the study of mortality differentials by education level requires good available information according to socioeconomic variables.

The quality of health information systems has been assessed in a multidimensional way (Lima et al., 2009). Two of these dimensions have been widely discussed in the case of the Sistema de Informação de Mortalidade (Mortality Information System or SIM, by its Portuguese acronym) of the Ministério da Saúde (Ministry of Health or MS, by its Portuguese acronym) of Brazil, namely the coverage and completeness of the information, and mainly in studies on mortality. Regarding the coverage of mortality records, due to the development and improvement of demographic and statistical techniques, or the combination of these, which measure

the underreporting of deaths, the literature has advanced considerably regarding the assessment on the coverage degree of deaths (Paes, 2005; Paes & Albuquerque, 1999; Queiroz, 2012; Queiroz et al., 2017; Schmertmann & Gonzaga, 2018; Szwarcwald, 2008).

Regarding the efficient and complete filling of death certificates, the analysis of mortality according to variables such as education remain a challenge, making it more difficult to understand the socioeconomic differentials of mortality in Brazil. In 2010, 31.96 % of deaths registered in the SIM did not have information about the deceased's education. As an alternative, information on deaths at home in the last twelve months has been used, which is available in the 2010 Brazilian Census (Silva, Freire & Pereira, 2010). Despite of 2010 Demographic Census can be used as an alternative source to measure mortality by education in Brazil, an important limitation in this data source is the fact that the information refers to the person responsible for the household surveyed and not to the person who died. Even though there may be a positive correlation between the education level of the household head and other residents, it is an important limitation. In addition, there is evidence of inaccuracy in schooling data from the 1991 and 2000 Censuses. When analyzing the inter-census survival ratios, gradients in schooling mortality were inconsistent with the literature, which suggests a positive relationship between survival and level of education of the individual (Nepomuceno & Turra, 2020).

The main advantage of analyzing the mortality differentials with the SIM data lies in the fact that the system is fed with annual records, and the education information is from the person who died (Ribeiro, Turra & Pinto, 2016). In this context, and given the limitations and importance of education as a socioeconomic variable for mortality studies, an analysis on the completeness of this information in the Mortality Information System is necessary, investigating whether the non-declaration has any selectivity according to sociodemographic variables. This information is relevant for applying imputation techniques, as it enables evaluating the pattern of the missing data, the mechanisms which generate them, and consequently to select methods which are better suited (Enders, 2010).

There are several studies on evaluating the quality of filling in SIM data according to some variable of interest. In a systematic review of the literature, Lima et al. (2009) identified 15 studies which assessed the completeness involving some SIM variable addressing a specific target audience or cause of death. Moreover, Romero & Cunha (2006) analyzed the completeness

of socioeconomic and demographic aspects of the death certificates of children under one year old from 1996 to 2001 according to regions and federation units, concluding that maternal education was classified as very poor (50 % or more of missing data) in most Brazilian regions. There is evidence of a certain heterogeneity in the failure percentage to fill in the information on education due to causes of death (Felix et al., 2012; Macente & Zandonade, 2010). Nevertheless, there is also a downward trend in the failure percentage to complete the education variable in the southeast region and in Brazil as a whole (Felix et al., 2012).

However, these studies only descriptively and/or temporally measure the ignored data of the SIM variables, but do not seek to relate this incompleteness with other variables. Thus, the objective of this study was to evaluate the relationship between the ignored record of education of adult death in the Mortality Information System and some sociodemographic characteristics of the deceased in Brazil in 2010.

Data and Methods

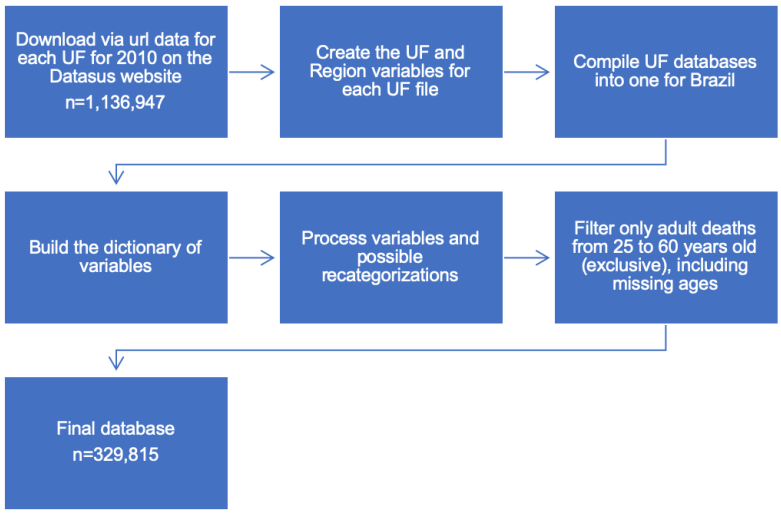
We used the death microdata from the Ministry of Health's Mortality Information System (SIM) in this study to assess the completeness of information on the education of adults, aged 15 to 59, who died in Brazil in 2010. As the main study variable is education, we chose to use only the deceased aged between 25 and 60 years old (exclusive), assuming it is reasonable to accept that from the age of 25 onwards there is a consolidation of the education level of most people, including those with school lag. When evaluating the non-declaration of education in deaths by age, we noticed that it is higher in the population aged 60 years or older than among adults aged 25 to 59 (Silva Júnior & Freire, 2018). However, we opted for the exclusion of older adults because there is evidence of problems in mortality estimates at advanced ages due to inaccurate data in both the numerator (deaths) and in the denominator (population) (Gomes & Turra, 2009; Queiroz et al., 2017; Queiroz & Sawyer, 2012).

The use of the 2010 Census' year is justified by the fact that it was the limit year in which the Ministry of Health began to recommend the preferential use of the new death certificate which introduced changes in the way of measuring education, thus minimizing possible errors when matching the data to be published (Ministério da Saúde [MS], 2011; Ribeiro, 2016). In addition, mortality estimates already have sources of uncertainty, such as underreporting of deaths and, in the case of the analysis of mortality by education, a

high percentage of missing in the records of that information. In this sense, when estimating mortality rates by education level, it is necessary to use the population from census's year in order do not introduce other sources of uncertainty in the estimates of mortality rates by education level.

The data were analyzed using the R software program (R Core Team, 2019). Every software routine used to generate the database is shown in Figure 1 and can be obtained from the repository available at GitHub (see Silva Júnior et al., 2021).

Figure 1. Operational flow for obtaining and consolidating the database of deaths registered by the Ministry of Health/Datasus/SIM in 2010



Source: Own elaboration based on data from the Mortality Information System/Ministry of Health.

Regarding the available variables, it is important to note that the SIM makes 61 variables available in its microdata. However, many are only filled out for fetal deaths or for deaths under 1 year or for women. Others refer to the date, time and codes of the death certificate. There are also those which are not even in the dictionary and that all cases are ignored. These variables were not considered in the study. As the interest lies in the sociodemographic characteristics of the deaths, the variables used in this assessment are education, large regions, gender, age, race/skin color, marital status, place of death occurrence, and cause of death. Table 1 shows the categorization used for each study variable.

Table 1. Sociodemographic variables of deaths registered in the SIM in 2010 and categorizations used

Variable	Description	Microdata categories	Final categories
Region	Residence region of the deceased	NA*	1: North
			2: Northeast
			3: Southeast
			4: South
			5: Center West
Gender	Gender of the deceased	1: Male	1: Male
		2: Female	2: Female
		9: Ignored	9: Ignored
Age (five-year groups)	Age of the deceased	Discrete variable (in years)	6: 25 to 29 years
			7: 30 to 34 years
			8: 35 to 39 years
			9: 40 to 44 years
			10: 45 to 49 years
			11: 50 to 54 years
			12: 55 to 59 years
		000, 999 and blank: Ignored	999: Ignored (000, 999, blank)
Skin color/race	Race or skin color of the deceased	1:White	1: White
		2:Black	2: Black
		3:Yellow	4: Brown
		4: Brown	6: Yellow or indigenous
		5: Indigenous	9: Ignored (Blank)
		Blank	
Civil status	Civil status of the deceased	1: Single	1: Single
		2: Married	2: Married
		3: Widower	3: Widower
		4: Judicially separated	6: Divorced/Others (4 and 5)
		5:	
		9: Ignored	9: Ignored
Place of occurrence	Place of death occurrence	1: Hospital	1: Hospital
		2: Other health establishment	2: Other health establishment
		3: Home	3: Home
		4: Public road	4: Public road
		5: Others	5: Others
		9: Ignored	9: Ignored

(continues)

Table 1 (continuation)

Variable	Description	Microdata categories	Final categories
Basic cause of death (in chapters)	Basic cause of death according to ICD10	ICD-10 codes	2: II. Neoplasms (tumors) 9: IX. Circulatory system diseases 10: X. Respiratory system diseases 20: XX. External causes 99: Other
Education	Education of the deceased (in years of study completed)	1: None 2: 1 to 3 years 3: 4 to 7 years 4: 8 to 11 years 5: 12 years and more 9: Ignored 0: Blank	1: None 2: 1 to 3 years 3: 4 to 7 years 4: 8 to 11 years 5: 12 years and more 9: Ignored (9, 0 and in blank)
Education (3 categories)	Education with recategorization into low, medium and high education using the variable "Education"	NA	1, 2, 3 => 1: low 4 => 2: medium 5 => 3: high 9 => 9: Ignored
Education missing	Indicates whether education was missing or not. Coded from the variable "Education"	NA	1, 2, 3, 4, 5 => 0 : No 9 => 1: Yes

**NA = not applicable*
Source: Mortality Information System/Ministry of Health.

SIM microdata is made available by the Unidade Federativa, or States in Brazil (UF, by its Portuguese acronym). With this, a variable called "UF" was created in the database so that it was possible to aggregate the respective UFs in a variable called "Region". The age variable is available with 3 digits, which have been transformed into a discrete variable (in years), as suggested by the dictionary and subsequently in five-year groups. Regarding the skin color/race variable, the original categorization was maintained, except for the "Yellow" and "Indigenous" categories. They represent only 0.3 % and 0.2 % of adult deaths, respectively, and have been grouped. For the variable

marital status, the levels "judicially separated" (Divorced) were added with the category "Others", and the others were maintained. There was no recategorization for the variables gender and place of death occurrence.

Information on cause of death was categorized in order to include the causes of death chapters where the failure to declare the education of the deceases was more expressive. A cut-off criterion of 10% was adopted, meaning that in relation to the total number of non-declarations of education. The causes of death chapters with ignored data above this value were treated individually. The categories below the cut-off were grouped in the category "other chapters". The value of 10% was based on the classification of completeness presented by Romero & Cunha (2006), where variables with up to 10% of ignored records are classified as excellent or good and above 10% as fair, bad or very bad. Another alternative for categorizing the cause of death would be: preventable, non-preventable and ill-defined causes (Malta et al., 2011). However, after performing the categorization in this way, the death totals did not correspond to those observed in the tables on the Datasus website (see Ministério de Saúde, n.d.) and were therefore not used.

An indicator variable called "Missing education" was created based on the education variable, in which the value 1 was assigned if the education data is ignored or blank, and 0 otherwise. This variable was used as the response variable in this study. The other two forms of the education variable, one in the original form of the SIM base and the other with 3 categories, are used for descriptive analysis. It is worth mentioning that deaths whose education was classified as higher education were grouped into a single category, either for those who were taking a higher education course or for those who had already completed this stage.

It is also important to note that in addition to incompleteness and coverage, the SIM microdata of deaths lacks standardization and greater control of the tabulation of some variables. There is a discrepancy for some of these between the categories proposed in the dictionary, those observed in the microdata and those which Datasus makes available in the tables consulted on the website. The dictionary establishes the code 000 for ignored data in the age variable, but three codes appear for the ignored records in the microdata (000, 999, and "blank").

The skin color/race variable is also non-compliant. The dictionary does not establish a code for ignored data, but these were left blank in microdata. Another problem was found in the variable marital status. The encoding

"5" appears in the microdata in this attribute, but does not appear in the dictionary. This code was considered as the "other" category when consulting the tables on the Datasus website.

A similar situation is found in the main study variable: education. The code "0" appears in the microdata, but it is not included in the dictionary. There were also records of blank cells for the education of the deceased. The official tables of the website considered the code "9", "0" and "blank" as missing data. The other variables used here did not present problems of this nature.

These inconsistencies represent a possible source of error, mainly for the education variable. The code "0" was considered to be ignored, but we have no way of knowing if these zeros were not to be counted in the "no" education category as being 0 (zero) years of study. This can underestimate the number of deaths in the less educated and consequently underestimate the mortality estimates of the less educated.

It is also important to warn of possible problems with the information provided about the education of the deceased in the death certificates (DC). Despite only being mandatory since 2011, the new death certificate model may be used simultaneously in different UFs since 2009. This can be a source of error, since the education level has changed in the new DC. The Ministry of Health makes both forms compatible before making microdata available (MS, 2011; Ribeiro, 2016). Another source of error is the fact that DC information is provided by third parties who often may not have sufficient knowledge about the victim (Guedes et al., 2011; Pérez, 2010; Ribeiro, 2016). Hummer & Hernandez (2013) show that third parties have a preference for reporting higher education for those who died on death certificates, which may underestimate the category of less education. However, this problem is more evident in older adults (Sorlie & Johnson, 1996), which does not affect the approach adopted in this study very much.

With regard to data analysis, univariate and bivariate descriptive analysis was used by means of absolute (n) and relative (%) frequencies. The missing data in the education variable of adult death was univariately analyzed in the first moment how it behaved by region throughout the availability period of SIM data (1979 to 2018). Focusing on the year 2010, a univariate descriptive analysis of the explanatory variables, in addition to education, is performed. The explanatory variables are analyzed in a bivariate manner according to the missing education response variable.

The odds ratio (OR) was used as an association measure to identify the relationship between the Ignored education (Yes/No) variable and the explanatory variables (sociodemographic). Here it is important to note that since SIM death data are from the universe and not sampled, it is not necessary to use inferential analysis. Some studies used inferential techniques when analyzing the completeness of SIM variables, in which they used regression models to assess the trend in the completeness of sociodemographic variables (Felix et al., 2012; Macente & Zandonade, 2010). However, despite controversies based on the overpopulation argument, we understand that statistical techniques for inferential data such as linear, logistic or Poisson regression models are not applicable in this type of study, since we are not dealing with sample data.

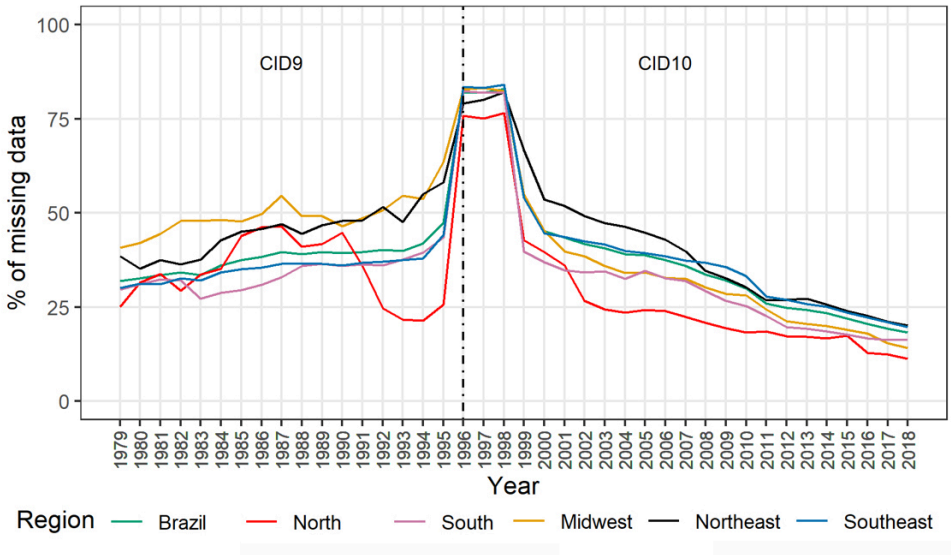
Results

A preliminary analysis of the data quality history from the completeness perspective of the item "Education" of adult death (including cases with missing age) registered in the SIM is shown in Figure 2. It was found that the percentages of missing data in this variable from 1979 to 2018 (except 1996 to 1998) for Brazil ranged from 21% to 54%. There was negligence in filling out death certificates in the years 1996 to 1998, reaching an average missing data percentage of 82% in the period. It is also evident that this neglect occurred in all Brazilian regions in the same period. A hypothesis as to why this occurred may be related to the changes which occurred from the nineteenth to the tenth version of the International Classification of Diseases (ICD-9 to ICD-10), which occurred between 1995 and 1996, except for mere coincidence.

With regard to the evolution of the completeness of education information in the death record, there is a trend after 1998 towards improving the quality of this item in all Brazilian regions, going from levels of 70 to 80% of missing data in 1998 to 10% to 20% in 2016 (Figure 2). A positive highlight is the North region, where it has had the lowest percentage of records with ignored education since 1991, with a slight exception between 1999 and 2001 and with an important highlight for the period from 1992 to 1995. A total of 329,815 adult deaths of victims between 25 and 59 years old, including deaths with missing ages were recorded in Brazil in 2010, which is the focus of this study. Of this total, the percentage of missing records of the adult deceased's education was 29.96% (98,820 cases). The Midwest, South and North regions showed missing record percentages below the level of Brazil for 2010, with 27.11%, 24.79% and 17.42%, respectively, while the Northeast (30.19 %) and Southeast (33.29%) were higher (Figure 2). These regional differences can

also be observed at the state level, where the completeness of the education information in the 2010's death record ranges from 9.5% in Paraná to 54.8% in Espírito Santo (see Table 4 in the Appendix section of this paper).

Figure 2. Percentages of missing data in the education variable of deceased persons from 25 to 59 years including cases with missing age, Brazil, large regions, 1979 to 2018



Source: Mortality Information System/Ministry of Health, 1979 to 2018.

Table 2 shows the distribution of deaths according to the variables of interest. It appears that there were no ignored cases regarding the distribution of deaths by region, since the microdata were extracted by federal unit, which enabled creating the variable region without registration loss.

The largest share of deaths is in the Southeast (46.3%), followed by the Northeast (24.9%). The basic cause of death also did not present ignored data. The most frequent causes of death were those in Chapter XX, which deals with external causes, with 23.9% of cases. Chapters IX, of the circulatory system diseases, and II of Neoplasms (tumors) were also quite representative, with percentages of 21.3% and 17.3% , respectively.

All other variables showed missing data: the gender variable had 0.1% of ignored data; age had 1.2% of records without declaration; place of death had 0.4% of records without this information; skin color/race had 6.3% with non-information; and marital status had 8.4% of cases without this information. The education variable presented a percentage of 30.0% of death certificates without information.

Table 2. Distribution of the number of adult deaths in Brazil in 2010 according to demographic variables and education

Variables	Categories	n	%
Region	North	21,619	6.6
	Northeast	82,048	24.9
	Southeast	152,661	46.3
	South	49,647	15.1
	Center West	23,840	7.2
Gender	Male	221,753	67.2
	Female	107,858	32.7
	Ignored	204	0.1
Age	25 to 29 years	28,769	8.7
	30 to 34 years	29,916	9.1
	35 to 39 years	32,089	9.7
	40 to 44 years	40,549	12.3
	45 to 49 years	52,849	16.0
	50 to 54 years	65,460	19.8
	55 to 59 years	76,216	23.1
	Ignored	3,967	1.2
Skin color/race	White	144,998	44.0
	Black	31,000	9.4
	Brown	131,370	39.8
	Yellow or indigenous	1,590	0.5
	Ignored	20,857	6.3
Civil status	Single	143,070	43.4
	Married	120,659	36.6
	Widower	12,856	3.9
	Divorced/Others	25,422	7.7
	Ignored	27,808	8.4
Place of death occurrence	Hospital	209,194	63.4
	Other health establishment	11,118	3.4
	Home	55,177	16.7
	Public road	35,534	10.8
	Others	17,412	5.3
	Ignored	1,380	0.4
Basic cause of death (in chapters)	II. Neoplasms (tumors)	56,900	17.3
	IX. Circulatory system diseases	70,246	21.3
	X. Respiratory system diseases	18,898	5.7
	XX. External causes of morbidity and mortality	78,935	23.9
	Others	104,836	31.8

(continues)

Table 2 (continuation)

Variables	Categories	n	%
Education	None	24,912	7.6
	1 to 3 years	56,381	17.1
	4 to 7 years	82,220	24.9
	8 to 11 years	46,294	14.0
	12 years or more	21,188	6.4
	Ignored	98,820	30.0
Education (with 3 categories)	Low	163,513	49.6
	Medium	46,294	14.0
	High	21,188	6.4
	Ignored	98,820	30.0
Education missing	No	230,995	70.0
	Yes	98,820	30.0
Total (N)		329,815	100.0

Source: Mortality Information System/Ministry of Health, 2010.

Table 2 generally reveals that deaths are more frequent in males (67.2%), increase as age increases, affect more whites (44.0%) and browns (39.8%), is more frequent among singles (43.4%) and married (36.6%), and the majority (63.4%) occur in hospitals. Regarding the distribution of deaths according to education, there is a large concentration of deaths in the lower ranges of study years, where almost half of the deceased have up to 7 years of studying. Only 6.4% of those who died in 2010 had 12 years or more of studying.

Given this general overview of the distribution of deaths by each variable considered in the study, Table 3 provides the absolute (n), relative (%) and odds ratios (OR) of the missing education variable for each considered variable.

When analyzing by region, we found that the highest percentage of deaths with missing education was the southeastern region (33.3%), while the lowest was the northern region (18.3%). There is almost no difference in the percentages regarding the gender of the deceased, where male and female deaths presented 30.7% and 28.3% of missing education data. Of the cases where gender is not informed, 94.6% also did not have their education registered. Regarding age, there is a uniformity of missing data for education around 29% in each age group. For the deaths of undeclared age, 97.5% had not completed their education.

Still in Table 3, the percentage of missing education when analyzing color or race among blacks was 28.7%, while the lowest percentage was among yellow or indigenous people, with 25.0%. Information about education was also not recorded in 66.2% of the cases in which the color or race variable was ignored. Regarding marital status, singles had the lowest percentage of ignored education (24.2%), while widowers had the most ignored (26.5%). Among the cases with undeclared marital status, 83.7% were without an informed education level.

Regarding the place where the death occurred, 35.5% of the deaths which occurred on public roads were missing data for education; this percentage was 29.0% for those who died in the hospital, while 21.7% for those who died in another health facility did not have their education recorded. Still regarding the place of death occurrence, 61.3% among the death certificates on which this variable was ignored were left without education records.

When disaggregating for basic cause of death, it appears that the percentages of missing education were higher in Chapter XX - External causes, with 33.3%.

Still in relation to Table 3 and analyzing the odds ratios, we found that the chance of finding missing data on education of deceased in the Southeast region is 2.223 times the chance verified in the North region.

The chance of missing education in male deaths is 1.126 times compared to female. All age groups had only a slightly greater chance of presenting ignored education data than the youngest deceased aged 25 to 29, with odds ratios ranging from 1.029 to 1.062 times.

Regarding the race or skin color of the deceased (Table 3), except for the black skin color or race (there was practically no difference between the odds ratios), all categories were less likely to have missing data for education compared to whites, with the largest being between yellow or indigenous and whites with a 1.192 times lower chance (OR = 1/0.839) in deaths declared as yellow or indigenous. From the point of view of the marital status of the deceased, it is observed that the chances of this outcome occurring in the married, widowed and divorced/others are 1.085, 1.130 and 1.093 times higher, respectively, compared to the deceased who were single.

Table 3. Distribution and Odds Ratios (OR) of the ignored education of adult deaths variable according to sociodemographic variables, Brazil, 2010

Variables	Categories	Missing education				OR*
		No		Yes		
		n	%	n	%	
Region (North)	North	17,655	81.7	3,964	18.3	1.000
	Northeast	57,275	69.8	24,773	30.2	1.926
	Southeast	101,833	66.7	50,828	33.3	2.223
	South	37,084	74.7	12,563	25.3	1.509
	Center West	17,148	71.9	6,692	28.1	1.738
Gender (Female)	Male	153,605	69.3	68,148	30.7	1.126
	Female	77,379	71.7	30,479	28.3	1.000
	Ignored	11	5.4	193	94.6	44.544
Age (25 to 29 years)	25 to 29 years	20,608	71.6	8,161	28.4	1.000
	30 to 34 years	21,062	70.4	8,854	29.6	1.062
	35 to 39 years	22,707	70.8	9,382	29.2	1.043
	40 to 44 years	28,569	70.5	11,980	29.5	1.059
	45 to 49 years	37,411	70.8	15,438	29.2	1.042
	50 to 54 years	46,392	70.9	19,068	29.1	1.038
	55 to 59 years	54,147	71.0	22,069	29.0	1.029
	Ignored	99	2.5	3,868	97.5	98.661
Skin color or Race (White)	White	103,705	71.5	41,293	28.5	1.000
	Black	22,116	71.3	8,884	28.7	1.009
	Brown	96,942	73.8	34,428	26.2	0.892
	Yellow or indigenous	1,192	75.0	398	25.0	0.839
	Ignored	7,040	33.8	13,817	66.2	4.929
Civil status (Single)	Single	108,501	75.8	34,569	24.2	1.000
	Married	89,665	74.3	30,994	25.7	1.085
	Widower	9,452	73.5	3,404	26.5	1.130
	Divorced/Others	18,854	74.2	6,568	25.8	1.093
	Ignored	4,523	16.3	23,285	83.7	16.158
Place of death occurrence (Home)	Hospital	148,441	71.0	60,753	29.0	1.050
	Other health establishment	8,108	72.9	3,010	27.1	0.952
	Home	39,698	71.9	15,479	28.1	1.000
	Public road	22,903	64.5	12,631	35.5	1.414
	Others	11,311	65.0	6,101	35.0	1.383
	Ignored	534	38.7	846	61.3	4.063

(continues)

Table 3 (continuation)

Variables	Categories	Missing education				OR*
		No		Yes		
		n	%	n	%	
Basic cause of death (II. Neoplasms (tumors))	II. Neoplasms (tumors)	41,954	73.7	14,946	26.3	1.000
	IX. Circulatory system diseases	50,320	71.6	19,926	28.4	1.112
	X. Respiratory system diseases	13,139	69.5	5,759	30.5	1.230
	XX. External causes of morbidity and mortality	52,681	66.7	26,254	33.3	1.399
	Others	72,901	69.5	31,935	30.5	1.230

**The response variable is Ignored education (Yes/No) where the outcome is Yes for the OR calculation.*
Source: Mortality Information System/Ministry of Health, 2010.

When considering the place of death, those who died on public roads have a 1.414 times greater chance of having their education ignored on the SIM death record compared to those deaths which occurred in the deceased's home. This chance is practically the same when death occurs in the hospital, only 1.050 times greater.

When the basic causes of death are taken into account, deaths due to external causes are those which present a greater chance (1.399 times) of the occurrence of lacking education information compared to deaths due to neoplasms (tumors).

It is essential to call attention to the behavior of the chances of observing the missing education when the explanatory variables are also not observed. They are extremely larger compared to the reference categories for each item. This value was 44.564 times for the gender variable, while the following values were found for: age (98.661), color or race (4.929), marital status (16.158) and place of death occurrence (4.063).

Discussion and conclusion

Through the performed analyzes, it was noticed that the chances of education of the deceased being ignored when filling out the death certificate differ according to sociodemographic characteristics. Regarding the region variable, it was found that the chance of presenting missing data is lower in the northern region compared to the other regions. This result is

surprising since studies which evaluated the quality of the data under the aspect of the degree of death coverage by region highlight that the SIM death data in the northern region has less coverage. Queiroz et al. (2017) found that all states in the South and Southeast regions had coverage of 100% of SIM death records in 2010, while under-registration was detected in some of the North region. Thus, it appears that better coverage does not necessarily mean better data quality under the completeness dimension.

On the other hand, it is possible that the higher incompleteness observed may be influenced by the different degrees of coverage in these variables, with the argument that UFs with a lower degree of coverage (as is the case in the North) tend to possibly have higher proportions of death certificates filled out by doctors and administrative agents in better condition (Romero & Cunha, 2006).

The results showed that the distribution of non-declaration of education for deaths registered in the Brazilian Mortality Information System is not uniform across socioeconomic and demographic characteristics. If this were true, it would even be possible to eliminate cases of death with no information on education for studies whose objective was to measure the differential in the risk of death according to education levels, meaning that we could know the mortality pattern by education. However, it would still not be possible to carry out studies to measure the mortality level by education.

Nevertheless, the findings of this study indicate that men of older age, white, widowed, who died on public roads and from external causes are more likely to have their education record ignored on the death certificate compared to the reference category for each one of the variables (gender, color or race, marital status, place of death and cause of death), as already shown in Table 3. In fact, one would expect that missing information about education level in deaths certificates should be higher for those who died on public roads or from other external causes, especially for men. Probably, no schooling information is reported for most people who died in public roads, for example.

The chances were also extremely higher in all cases when the explanatory variables were ignored. This concomitance of ignored data in more than one study variable may represent a limitation for the use of methods which aim to input the educational variable based on the distribution of auxiliary variables. Some of them require that these variables be complete, meaning without missing data.

Therefore, it is expected that the results found in this study are useful for applying imputation methods which depend on an occurrence pattern of the missing data due to the sociodemographic characteristics of the registered deaths. In turn, the results of imputing information on the educational level of people who died are extremely relevant for analyzing socioeconomic gradients of mortality (Ribeiro et al., 2016; Turra, Ribeiro & Pinto, 2018).

References

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford press.

Felix, J. D., Zandonade, E., Amorim, M. H. C. & Castro, D. S. (2012). Avaliação da completude das variáveis epidemiológicas do Sistema de Informação sobre Mortalidade em mulheres com óbitos por câncer de mama na Região Sudeste: Brasil (1998 a 2007). *Ciência & Saúde Coletiva, 17*(4), 945-953. https://doi.org/10.1590/S1413-81232012000400016

Filho, A. M. S., Souza, M. de F. M. de, Gazal-Carvalho, C., Malta, D. C., Alencar, A. P., Silva, M. M. A. da & Neto, O. L. de M. (2007). Análise da mortalidade por homicídios no Brasil. *Epidemiologia e Serviços de Saúde, 16*(1), 7-18. http://dx.doi.org/10.5123/S1679-49742007000100002

Gomes, M. M. F. & Turra, C. M. (2009). The number of centenarians in Brazil: Indirect estimates based on death certificates. *Demographic Research, 20*(20), 495-502. https://dx.doi.org/10.4054/DemRes.2009.20.20

Guedes, G. R., Siviero, P. C. L., Queiroz, B. L. & Machado, C. J. (2011). Approximating the educational differences in mortality: demographic indirect techniques. *Cadernos de Saúde Coletiva, 19*(2), 240-243. https://gilvanguedes.com/wp-content/uploads/2016/09/guedes_etal_2011b_csc.pdf

Hummer, R. A. & Hernandez, E. M. (2013). The Effect of Educational Attainment on Adult Mortality in the United States. *Population Bulletin, 68*(1), 1-16. https://www.prb.org/wp-content/uploads/2013/07/Population-bulletin-2013-68-1-us-education-mortality.pdf

Kitagawa, E. M. & Hauser, P. M. (1973). *Differential mortality in the United States: a study in socioeconomic epidemiology*. Cambridge, MA: Harvard University Press. https://doi.org/10.4159/harvard.9780674188471

Lima, C. R. A., Schramm, J. M. A., Coeli, C. M. & Silva, M. E. M. (2009). Revisão das dimensões de qualidade dos dados e métodos aplicados na avaliação dos sistemas de informação em saúde. *Cadernos de Saúde Pública, 25*(10), 95-109. https://doi.org/10.1590/S0102-311X2009001000002

Macente, L. B. & Zandonade, E. (2010). Avaliação da completude do sistema de informação sobre mortalidade por suicídio na região Sudeste, Brasil, no período de 1996 a 2007. *Jornal Brasileiro de Psiquiatria, 59*(3), 173-181. https://doi.org/10.1590/S0047-20852010000300002

Malta, D. C., França, E., Abreu, D. X., Oliveira, H., Monteiro, R. A., Sardinha, L. M. V., Duarte, E. C. & Azevedo e Silva, G. (2011). Atualização da lista de causas de mortes evitáveis (5 a 74 anos de idade) por intervenções do Sistema Único de Saúde do Brasil. *Epidemiologia e Serviços de Saúde*, *20*, 409-412. http://dx.doi.org/10.5123/S1679-49742011000300016

Ministério da Saúde. (n.d.). *Datasus*. https://datasus.saude.gov.br/

Ministério da Saúde. (2011). *Sistema de Informações sobre Mortalidade - SIM: Consolidação da base de dados de 2011*. http://tabnet.datasus.gov.br/cgi/sim/Consolida_Sim_2011.pdf

Muller, A. (2002). Education, income inequality, and mortality: a multiple regression analysis. *BMJ, 324*, 23. https://doi.org/10.1136/bmj.324.7328.23

Nepomuceno, M. R. & Turra, C. M. (2020). Assessing the quality of education reporting in Brazilian censuses. *Demographic Research, 42*(15), 441-460. https://dx.doi.org/10.4054/DemRes.2020.42.15

Paes, N. A. (2005). Avaliação da cobertura dos registros de óbitos dos estados brasileiros em 2000. *Revista de Saúde Pública, 39*, 882-890. https://doi.org/10.1590/S0034-89102005000600003

Paes, N. A. & Albuquerque, M. E. E. (1999). Avaliação da qualidade dos dados populacionais e cobertura dos registros de óbitos para as regiões brasileiras. *Revista de Saúde Pública, 33*(1), 33-43. https://doi.org/10.1590/S0034-89101999000100006

Pérez, E. R. (2010). Estimativas de mortalidade adulta feminina por nível de escolaridade no Brasil (Doctoral thesis). Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais. http://hdl.handle.net/1843/AMSA-8ELQ3Y

Preston, S. H. & Taubman, P. (1994). Socioeconomic Differences in Adult Mortality and Health Status. In L. G. Martin & S. H. Preston (Eds.), *Demography of Aging* (pp. 279-318). Washington, D.C.: National Academy Press. https://doi.org/10.17226/4553

Queiroz, B. L. (2012). Estimativas do Grau de Cobertura e da Esperança de Vida para as Unidades da Federação no Brasil entre 2000 e 2010. In *XVIII Encontro Nacional de Estudos Populacionais*. Associação Brasileira de Estudos Populacionais, Águas de Lindóia, São Paulo, Brazil.

Queiroz, B. L., Freire, F. H. M. de A., Gonzaga, M. R. & Lima, E. E. C. de. (2017). Estimativas do grau de cobertura e da mortalidade adulta (45q15) para as unidades da federação no Brasil entre 1980 e 2010. *Revista Brasileira de Epidemiologia, 20*(Suppl. 1), 21-33. https://doi.org/10.1590/1980-5497201700050003

Queiroz, B. L. & Sawyer, D. O. T. (2012). O que os dados de mortalidade do Censo de 2010 podem nos dizer. *Revista Brasileira Estudos de População, 29*(2), 225-238. https://www.rebep.org.br/revista/article/view/31

R Core Team. (2019). R: A Language and Environment for Statistical Computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing.

Ribeiro, M. M. (2016). *Mortalidade adulta por níveis de escolaridade no estado e município de São Paulo: uma proposta de estimação a partir do Censo Demográfico de 2010* (Doctoral thesis). Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais. http://hdl.handle.net/1843/FACE-AKWJ6C

Ribeiro, M. M., Turra, C. M. & Pinto, C. C. X. (2016). Estimativas de mortalidade adulta por níveis de escolaridade no estado São Paulo em 2010. In *XX Encontro Nacional de Estudos Populacionais*. Associação Brasileira de Estudos Populacionais, Foz do Iguaçu, Paraná, Brazil.

Romero, D. E. & Cunha, C. B. da. (2006). Avaliação da qualidade das variáveis sócio-econômicas e demográficas dos óbitos de crianças menores de um ano registrados no Sistema de Informações sobre Mortalidade do Brasil (1996/2001). *Cadernos de Saúde Pública, 22*(3), 673-681. https://doi.org/10.1590/S0102-311X2006000300022

Schmertmann, C. P. & Gonzaga, M. R. (2018). Bayesian estimation of age-specific mortality and life expectancy for small areas with defective vital records. *Demography, 55*(4), 1363-1388. https://doi.org/10.1007/s13524-018-0695-2

Silva, L. E. da, Freire, F. H. M. de A. & Pereira, R. H. M. (2010). Diferenciais de mortalidade por escolaridade da população adulta brasileira, em 2010. *Cadernos de Saúde Pública, 32*(4), 2-12. https://doi.org/10.1590/0102-311X00019815

Silva Júnior, W. P. & Freire, F. H. M. de A. (2018). Determinantes socio-demográficos da incompletitude dos dados de escolaridade do óbito adulto no Brasil em 2010. In *XXI Encontro Nacional de Estudos Populacionais*. Associação Brasileira de Estudos Populacionais, Poços de Caldas, Minas Gerais, Brazil.

Silva Júnior, W. P., Freire, F. H. M. de A., Silva, D. N. da, & Gonzaga, M. R. (2021). *Artigo - Avaliação da qualidade do registro de morte no Brasil: determinantes sociodemográficos de informações incompletas* de educação [Database and R language scripts]. https://github.com/walterpedro/artigo-completitude-escolaridade-obito

Sorlie, P. D. & Johnson, N. J. (1996). Validity of Education Information on the Death Certificate. *Epidemiology, 7*(4), 437-439. https://journals.lww.com/epidem/Abstract/1996/07000/Validity_of_Education_Information_on_the_Death.19.aspx

Szwarcwald, C. L. (2008). Strategies for improving the monitoring of vital events in Brazil. *International Journal of Epidemiology, 37*(4), 738-744. https://doi.org/10.1093/ije/dyn130

Travassos, C., Oliveira, E. X. G. de & Viacava, F. (2006). Desigualdades geográficas e sociais no acesso aos serviços de saúde no Brasil: 1998 e 2003. *Ciência & Saúde Coletiva, 11*(4), 975-986. http://dx.doi.org/10.1590/S1413-81232006000400019

Turra, C. M., Ribeiro, M. M. & Pinto, C. C. X. (2018). Padrões de mortalidade por escolaridade no Brasil: evidências a partir do Sistema de Informação sobre Mortalidade. In *XXI Encontro Nacional de Estudos Populacionais*. Associação Brasileira de Estudos Populacionais, Poços de Caldas, Minas Gerais, Brazil.

Appendix

Table 4. Distribution and Odds Ratios (OR) of the ignored education of adult deaths variable according to sociodemographic variables and states, Brazil, 2010

Region	Unidade Federal	Missing education				OR*
		No		Yes		
		n	%	n	%	
North	Amazonas	3,867	90.3	414	9.7	1.000
	Acre	466	48.6	492	51.4	9.862
	Amapá	493	70.7	204	29.3	3.865
	Rondônia	1,433	56.5	1,103	43.5	7.190
	Pará	9,230	88.7	1,173	11.3	1.187
	Roraima	471	72.6	178	27.4	3.530
	Tocantins	1,695	80.9	400	19.1	2.204
Northeast	Maranhão	7,043	88.2	942	11.8	1.249
	Piauí	3,492	84.2	653	15.8	1.747
	Ceará	8,382	70.2	3,564	29.8	3.972
	Rio Grande do Norte	2,425	58.8	1,698	41.2	6.540
	Paraíba	3,038	49.2	3,134	50.8	9.636
	Pernambuco	11,333	72.9	4,218	27.1	3.476
	Alagoas	3,208	56.3	2,490	43.7	7.250
	Sergipe	2,973	88.3	395	11.7	1.241
	Bahia	15,381	66.7	7,679	33.3	4.663
Southeast	Minas Gerais	21,137	60.0	14,101	40.0	6.231
	Rio de Janeiro	29,611	81.7	6,621	18.3	2.089
	São Paulo	48,051	64.5	26,433	35.5	5.138
	Espírito Santo	3,034	45.2	3,673	54.8	11.308
South	Paraná	17,478	90.5	1,825	9.5	0.975
	Rio Grande do Sul	11,862	58.7	8,342	41.3	6.569
	Santa Catarina	7,744	76.4	2,396	23.6	2.890
Center West	Mato Grosso do Sul	3,299	75.5	1,070	24.5	3.030
	Mato Grosso	4,603	87.1	680	12.9	1.380
	Goiais	6,055	57.4	4,492	42.6	6.929
	Distrito federal	3,191	87.6	450	12.4	1.317

**The response variable is Ignored education (Yes/No) where the outcome is Yes for the OR calculation.*
Source: SIM - Mortality Information System/Ministry of Health, 2010.