DOI: 10.1590/1413-81232021264.15922019

Underreporting of AIDS deaths in Brazil: linkage of hospital records with death certificate data

Rondinelle Alves do Carmo (https://orcid.org/0000-0002-5328-8170) [1]
Gabriela Moreira Policena (https://orcid.org/0000-0002-3191-1435) [2]
Gizelton Pereira Alencar (https://orcid.org/0000-0002-2354-9050) [3]
Elisabeth Barboza França (https://orcid.org/0000-0001-6984-0233) [4]
Ana Luiza Bierrenbach (https://orcid.org/ 0000-0002-6837-0636) [1]

Abstract *This study investigated the underreporting of deaths due to Acquired Immunodeficiency Syndrome (AIDS) in Brazil, using the product of the linkage between the Mortality Information System (SIM) and the Hospital Information System (SIH) for the years 2008 to 2012. The hypothesis was that there were deaths recorded in the SIM that should have AIDS as the underlying cause of death, but that had been poorly classified. Many of the decedents had previous hospitalizations due to the illness. Underreported AIDS deaths were defined taking into account the coding rules of the International Classification of Diseases 10th Revision, in order to discriminate AIDS deaths from those due to other causes in people living with HIV / AIDS. In this period, 60,362 deceased had AIDS as the underlying cause, and we found another 2,671 (4.2%) as underreported causes. Underreported deaths increased the average AIDS mortality rate from 6.3/100,000 to 6.6/100,000. In conclusion, this study demonstrated that, through linkage in between SIH and SIM, it is possible to find underreporting of AIDS deaths in Brazil. These results serve as a warning for the need for actions that would promote a better certification of the causes of death among AIDS patients.*

Key words *AIDS, Underreporting, Death, Information system*

[1] Instituto de Ensino e Pesquisa, Hospital Sírio-Libanês. R. Prof. Daher Cutait 69, Bela Vista. 01308-060 São Paulo SP Brasil. rondiron@hotmail.com
[2] Instituto de Patologia Tropical e Saúde Pública, Universidade Federal de Goiás. Goiânia GO Brasil.
[3] Departamento de Epidemiologia, Escola de Saúde Pública, Universidade de São Paulo. São Paulo SP Brasil.
[4] Escola de Medicina, Universidade Federal de Minas Gerais. Belo Horizonte MG Brasil.

Introduction

Brazil has historically made enormous investments in the diagnosis, monitoring and treatment of patients living with AIDS, being a reference for many countries[1-3]. The government's goal is to rapidly reduce new HIV infections, control the disease and reduce AIDS-related deaths so as to meet the international goals. From the beginning of the epidemic in 1980 to 2016, 316,088 deaths were recorded in the Mortality Information System (SIM) with AIDS as the underlying cause, with an annual average of 12,189 deaths in a ten-year period from 2006 to 2016[4]. Unreliable or incomplete data can lead to errors in planning the distribution of public resources, or actions that may be ineffective in controlling, reducing, or eliminating the diseases[5]. Thus, it is necessary to obtain reliable epidemiological information to allow for the good planning of the public health actions.

Brazilian studies using national secondary data such as the Hospitalization Information System (SIH) and the Mortality Information Systems (SIM) have been expanding in the field of collective health, being used to strengthen epidemiological surveillance and to evaluate hospital morbidity and mortality in SUS[6,7]. In this context, it is valid to try to complete, correct or simply verify the records of a health information database with data from another base, both of which are routinely collected[8,9]. Some studies have shown that there is underreporting of AIDS disease and that the inventory of cases carried out with the application of the record linkage method between the notification (SINAN), mortality (SIM), laboratory (SISCEL) and drug logistics (SISCLOM) information systems contributes to decrease this underreporting[10-12]. But not many studies have been dedicated to the possibility of identifying underreporting of AIDS deaths in Brazil and elsewhere[13-16].

Linkage is the pairing of data from multiple registers or database that allows for the qualification of the information contained in each one, filling in of missing data and finding cases and deaths that should be contained in one specific registry, but that are only available in another, in order to discover underreporting and obtain more reliable answers regarding the morbidity and mortality situation[17].

Death certificates (DC) in Brazil use the international form recommended by the World Health Organization. The underlying cause of death is the illness or injury that initiated the succession of events that led to death. Among the multiple causes of death stated on the DC, the underlying cause of death (UCD) is identified based on the International Classification of Diseases 10th Revision (ICD-10) coding rules[18,19]. The underlying cause listed by the physician in DC is usually chosen to be the underlying cause that is recorded in the SIM, although in some cases coding rules need to be applied. Registering HIV / AIDS deaths as UCD is a challenge for public health authorities around the world. Underreporting may compromise public actions needed to address the problem and the implementation of individual or collective health care services.

In this study, we intend to investigate this underreporting of deaths through record linkage between the SIM and the SIH national databases. We hypothesise that there were deaths recorded in the SIM that should have AIDS as the UCD, but instead were poorly classified, even if the patient had previous hospitalizations due to such illness. The present study could only be carried out because of the existence of a consolidated database, product of the linkage between SIM and SIH for the years 2008 and 2012. This database was created in the context of the study "Global Burden of Disease in Brazil - GDB Brasil", for which the results of the present study contributed[20].

Material and methods

Data sources and linkage procedures

This is a descriptive population-based study. The consolidated SIM-SIH database for the years 2008-2012 was the product of a two-step record-linkage process. First, an in-house linkage algorithm was developed in which the probability that two records belong to the same patient was based on the name and date of birth. Municipality of residence was used as a blocking variable. For each record a Bloom Filter was constructed following the methods developed by Schnell et al.[21]. Then, in order to increase the specificity of the matches found using the first step, a deterministic linkage was performed using an algorithm similar to the ones validated by Pacheco et al.[22] and Oliveira et al.[23]. This algorithm was developed using a set of rules based on the combination of several key variables (name, date of birth, address, zip code, hospital code, telephone, etc.) and their fragments, including those obtained with Soundex transformations.

After the linkage procedures, data on personal identification were deleted. The SIM data comprised all causes of death listed in the death certificates, the date of death and the municipality of residence. The SIH data contain records of all hospitalizations funded by the Unified Health System (SUS), which represents about 2/3 of the total hospitalizations in the country[24]. Since the SIH database is primarily used for financial reimbursement purposes, the probability of hospitalizations being underreported is small[25]. The main variable used in the analysis was the one containing the ICD-10 codes for the primary diagnosis, i.e. the condition that caused the hospitalization. Other variables that were used were admission and discharge dates.

The SIM was devised in 1975 due to the recognition of the importance of mortality data for epidemiological knowledge. The system is currently considered as universal, that is, every death that occurs in the country must be registered in it. The system has been continually improved to try to increase the consistency and quality of the information. Coverage has also improved over the past two decades[26].

Coding rules for identifying AIDS deaths

ICD-10 codes B20-B24 were the ones used for AIDS identification in both databases. We do not intend to fully explain the coding rules used to select the UCD due to their complexity. In short, the rules use a logic that always seeks to choose the cause that initiated the process of events that led to death. This can easily be done by applying the "General Principle" when the certifying physician declares the causes following what is recommended in the standard form, that is, starting with the "immediate cause" and following back by the logic sequence of "intermediate causes" up to the "UCD" listed on the lowest used line on part 1 of the DC. When the physician does not state the causes correctly, or when he/she lists causes without a logical sequence, the trained coders select one of the listed causes to be the underlying one using the several ICD-10 rules and an electronic algorithm, the SCB program. Unlisted causes cannot be selected[18,19].

Chart 1 presents a series of causes that are considered predictive or presumptive of AIDS according to the coding rules manual in use in Brazil during the years of the study[27,28]. These rules, as explained below, were the basis to find underreported AIDS deaths. If one of these causes is wrongly listed on the last line of part 1 of the DCs but there is an AIDS code mentioned in any other line, by rule AIDS will be selected as the UCD.

Case definitions

SIM records were separated according to their linkage status with the SIH:

"Unpaired deaths" were those for which no corresponding record of the same individual was found in the SIH.

"Paired deaths" were those for which one or more records corresponding to the same individual were found in the SIH.

"Paired deaths linked with a difference of up to one day between hospital discharge and death" were paired deaths according to the definition above, but had the peculiarity that the date of departure from the SIH and the date of death were equal or had one day of difference. This definition, which is a subset of the previous definition, intends to point out that the death occurred during the hospitalization whose cause was AIDS, according to the SIH.

SIM records were also separated according to the reported UCD:

"All-cause deaths" were those with any code listed in the UCD, including AIDS.

"Reported AIDS deaths" were those with UCD = AIDS (B20 to B24, O97,8).

Aids deaths were separated into the following categories according to their reporting status:

"Reported AIDS deaths" were those with UCD = AIDS (B20 to B24, O97,8).

"Underreported AIDS deaths" are those who should have AIDS in the underlying cause, but who were underreported.

Four types of underreporting of AIDS deaths were defined.

Type A deaths were those that had an AIDS predictive cause registered as the UCD and a record of current or previous hospitalization due to AIDS in the SIH à Death with an AIDS predictive disease listed as the UCD and a hospitalization for HIV / AIDS.

Type B deaths were those that had an AIDS presumptive cause registered as the UCD and a record of current or previous hospitalization due to AIDS in the SIH à Death with an AIDS presumptive disease listed as the UCD and a hospitalization for HIV / AIDS.

Type C deaths had an ill-defined cause listed as the UCD, that is, with a code from Chapter XVIII - symptoms, signs, and abnormal findings from clinical and laboratory exams not elsewhere

Chart 1. ICD-10 codes for predictive and presumptive causes of AIDS.

Groups of disease	Codes	Category
Infectious diseases	(A01.0 - A05.0) (A05.2 - A19.9) (A24.0 - A32.9) (A40.0 - A69.9) (A71.0 - A74.9) (A81.0 - A81.9) (A88.0 - A89) (A93.0 - A94) (A96.8 - A96.9) (A99)	Predictive causes of AIDS
Infectious diseases	(B00.0 - B02.9) (B07 - B15.9) (B18.0 - B19) (B25.0 - B25.9) (B27.0 - B49) (B58.0 - B89) (B94.8 - B94.9) (B99)	Predictive causes of AIDS
Kaposi's sarcoma	(C46.0 - C46.9)	Predictive causes of AIDS
Burkitt's lymphoma	(C83.7)	Predictive causes of AIDS
Neoplasms of lymphatic, hematopoietic and related tissues	(C81.0 - C96.9)	Predictive causes of AIDS
Secondary or unspecified anaemia	(D50.0 - D53.9)	Presumptive cause of AIDS
Acute post-haemorrhagic anaemia	(D62)	Presumptive cause of AIDS
Acquired deficiency of coagulation factor, other coagulation defects, coagulation defect not specified	(D68.4) (D68.8) (D68.9)	Presumptive cause of AIDS
Secondary thrombocytopenia, unspecified thrombocytopenia	(D69.5) (D69.6)	Presumptive cause of AIDS
Specified haemorrhagic conditions, unspecified bleeding disorder	(D69.8) (D69.9)	Presumptive cause of AIDS
Agranulocytosis	(D70)	Presumptive cause of AIDS
Other specified and unspecified disorders of white blood cells	(D72.8 - D72.9)	Presumptive cause of AIDS
Other specified and unspecified disorders of white blood cells	(D73.0 - D84.9)	Presumptive cause of AIDS
Other disorders of the blood and hematopoietic organs and immunodeficiencies	(E40 - E46)	Presumptive cause of AIDS
Marasmus and Malnutrition	(J12.0 - J18.9)	Predictive causes of AIDS
Pneumonia	(J96)	Presumptive cause of AIDS
Chronic Respiratory Insufficiency	(R64)	Presumptive cause of AIDS
Cachexia		Presumptive cause of AIDS

Source: Mortality Coding Manual. São Paulo: University of São Paulo, 2012.

classified, and a record of current or previous hospitalization due to AIDS in the SIH à Death with an ill-defined UCD and a hospitalization for HIV / AIDS.

Type D deaths were those for whom the registered UCD does not fall in the previous definitions, i.e. it is not predictive or presumptive of AIDS or an ill-defined cause. In addition, the reg-

istered UCDs were also not external, that is, did not have an ICD-10 code from Chapter XX - External causes of morbidity and mortality. Differently from the previous types, these deaths needed to have a record of current hospitalization due to AIDS in the SIH, i.e. with a discharge date on the same date of death, therefore indicating that the death occurred during the current hospitalization à Death with current hospitalization due to AIDS, but not type A, B or C.

Examples of each type can be seen in Chart 2.

Analysis

All analyzes were performed using Stata-13 (Stata Corporation, College Station, USA). Using the consolidated SIM-SIH database, algorithms were developed to identify each of the case definitions enumerated above. The percentage of underreporting, which is the main result, was calculated by year and by region of residence, considering the following equation:

$$\textit{Percentage of underreporting} = \frac{\textit{total number of underreported AIDS deaths}}{\textit{total number AIDS deaths}} * 100$$

The annual rates of AIDS mortality (per 100,000 inhabitants) were calculated for the country as a whole, before and after the inclusion of the underreported deaths.

1. Before the inclusion of underreported deaths;

$$\textit{Mortality rate} = \frac{\textit{total AIDS deaths reported in the year}}{\text{population in the year}} * 100{,}000$$

2. After inclusion of underreported deaths.

$$\textit{Mortality rate} = \frac{\textit{total AIDS deaths (report.+ underreport.) in the year}}{\text{population in the year}} * 100{,}000$$

The present study was included as an amendment to the national project called GLOBAL BURDEN OF DISEASES BRAZIL - GBD Brazil, approved by the Research Ethics Committee of the Federal University of Minas Gerais.

Results

In Figure 1, between 2008 and 2012, there were a total of 5,829,167 deaths due to all causes recorded in the SIM. Of this total, 60,362 (1%) were due to AIDS, that is, they had AIDS as the UCD. Of the total deaths due to all causes, 2,561,818 (44%) linked with one or more SIH records and, of these, 1,108,743 (43.3%) linked with a SIH record with a difference of up to one day between the dates of death and hospital discharge, so that they refer to the hospitalization in which the death occurred. The proportion of linked records was much higher for AIDS deaths than for those due to all causes. Of the total number of deaths due to AIDS, 43,049 (71.3%) linked with a SIH record, of which 21,893 (50.1%) linked with an SIH record with a difference of up to one day between the dates of death and of hospital discharge.

Table 1 shows the number and percentage of AIDS deaths reported and underreported per year and region. During the whole period, 2,671 AIDS deaths were underreported, what represents an underreporting percentage of 4.2%, which had a certain annual variation, but with no apparent trend. The first year was the one that had the lowest proportion of underreporting. Considering the distribution of reported and underreported AIDS deaths by region, it can be noted that the percentage of underreporting was similar for the regions, from 4.2 to 4.5%, except for the Center-West region where it was lower (2.5%).

Table 2 presents, per year, the total underreporting of each type, according to the study definitions. Although there is some variation in the proportional contribution of each type over the years, in general type D (death with a current hospitalization due to AIDS, but not type A, B or C) had the highest proportion, followed of type A (death with an AIDS predictive disease listed as the UCD and a hospitalization for HIV / AIDS), C (death with an ill-defined UCD, and a hospitalization for HIV / AIDS)

and B (death with an AIDS presumptive disease listed as the UCD, and a hospitalization for HIV / AIDS). A and C had a rising trend and a D had a decreasing one. In the last year of the series, type A had a similar proportion to type D.

AIDS annual mortality rates for the country as a whole before the inclusion of the underreported deaths varied from a minimum of 6.2 in 2008 to a maximum of 6.4 in 2010, with an average of 6.3 for the period (per 100,000 inhabi-

Chart 2. Examples of each type of underreporting of AIDS deaths.

SIH			SIM	
Data of entry	**Date of discharge**	**Primary diagnosis**	**Date of death**	**Underlying cause of death**
Case 1 –Type A - identification 1023571				
30 may 2009	03 june 2009	E40		
07 july 2009	17 july 2009	B20.7		
07 june 2011	28 june 2011	B20.0		
14 july 2011	22 de july 2011	B20.0	22 july 2011	A15.3
In case 01, the patient has hospitalizations due to AIDS in the SIH, more specifically, HIV disease resulting in mycobacterial infections (B20.0). Pulmonary tuberculosis (A15.3) has been selected as the underlying cause of death. Using the selection and modification rules, pulmonary tuberculosis should be considered as a direct consequence of AIDS, and the case should have been coded with code B20.0. This case was defined as type A underreporting (death with an AIDS predictive disease listed as the UCD and a hospitalization for HIV / AIDS).				
Case 2 – Type B - identification 1008738				
09 november 2009	18 november 2009	A08.5		
09 december 2009	10 december 2009	B20.4		
17 december 2009	11 january 2010	A09		
11 january 2010	12 january 2010	K80.8		
19 january 2010	01 february 2010	A49.9	01 february 2010	R64
In case 02, the patient has a hospitalization for unspecified bacterial infection (A49.9) and cachexia (R64) as the underlying cause of death in SIM. The dates of discharge and death coincide. As the patient contains, in his history of records in SIH, the diagnosis of HIV / AIDS (B20.4), using the rules of selection and modification, cachexia (R64) could be considered as a direct consequence of AIDS. This case was defined in our study as type B underreporting (death with an AIDS presumptive disease listed as the UCD, and a hospitalization for HIV / AIDS).				
Case 3 – Type C - identification 1025584				
24 may 2010	29 may 2010	B20.0	11 june 2010	R68.8
In case 3, the main diagnosis in the last registry in the SIH is the HIV disease resulting in mycobacterial infections (B20.0) and as a underlying cause of death other specified general signs and symptoms (R68.8 - cause ill defined in Chapter XVIII). Death occurred less than 2 weeks after discharge. So while the above-mentioned underlying cause code is not considered a direct consequence of HIV, we believe that the information that the patient had been hospitalized for AIDS is more valuable than a malformed code in establishing the underlying cause of death. The death in question was defined in our study as type C underreporting (death with an ill-defined UCD and a hospitalization for HIV / AIDS).				
Case 4 – Type D - identification 1072281				
10 june 2010	10 june 2010	B20.8	10 june 2010	J98.8
In case 04, the main diagnosis of the patient in the current SIH registry is the diagnosis of HIV disease resulting in other infectious diseases (B20.8) and other respiratory disorders specified as the underlying cause of death (J98.8). Although the aforementioned underlying cause is not considered within the rules of selection and modification as predictive or presumptive of AIDS, the date of death occurred on the same date of departure from the hospital, ie, the death occurred during hospitalization. Thus, we consider that AIDS should be the underlying cause, replacing the underlying cause selected in the DC. Type D underreporting (death with a current hospitalization due to AIDS, but not type A, B or C).				

Source: SIM - SIH, 2008 to 2012.

tants). After the inclusion of the underreported deaths, the rates varied from a minimum of 6.5 in 2008 and 2012 to a maximum of 6.7 in 2010, with an average of 6.6 for the period (per 100,000 inhabitants).

Discussion

Using records linked in between the mortality and the hospitalization databases, we found 2,671 AIDS deaths that were underreported

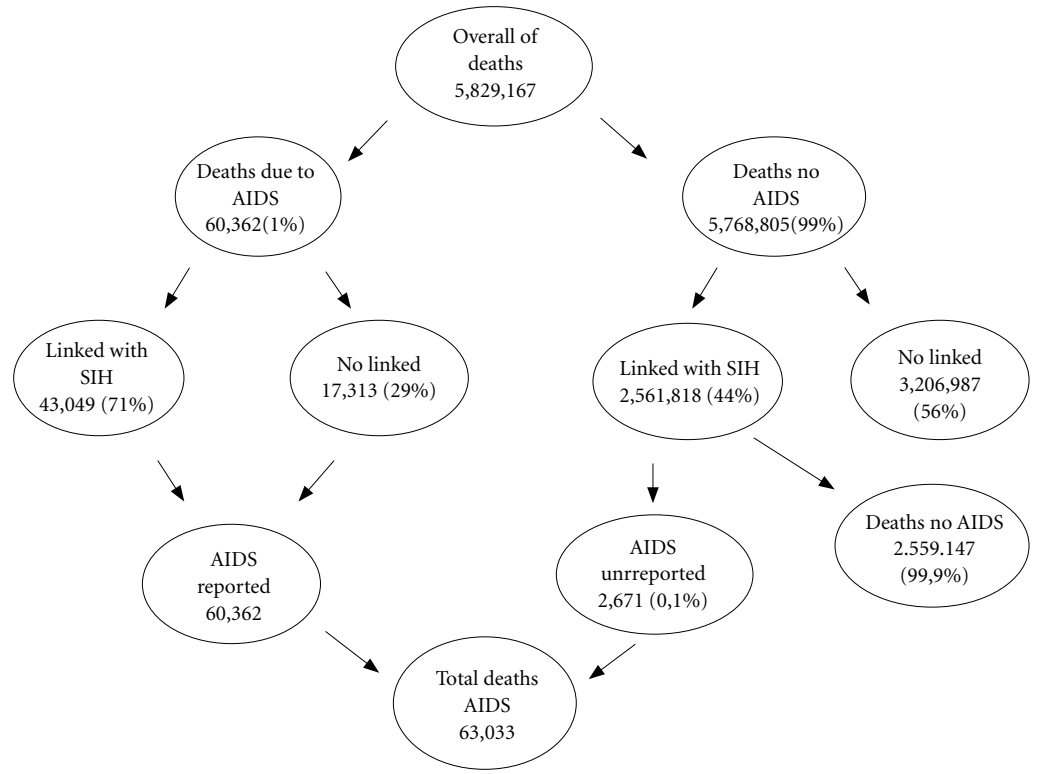

Figure 1. Linkage process performed for the study of AIDS deaths underreporting.

Source: SIM - SIH, 2008 to 2012.

Table 1. Distribution of reported and underreported AIDS deaths per year and region, Brazil.

	Aids deaths		
	Reported * N (%)	**Underreported ** N (%)**	**Total**
	Year		
2008	11,842 (96.5)	425 (3.5)	12,267 (100.0)
2009	12,136 (95.5)	570 (4.5)	12,706 (100.0)
2010	12,158 (95.4)	591 (4.6)	12,749 (100.0)
2011	12,153 (95.8)	531 (4.2)	12,684 (100.0)
2012	12,073 (95.6)	554 (4.4)	12,627 (100.0)
	Region		
Center-West	3,688 (97.5)	95 (2.5)	3,783 (100.0)
South	12,915 (95.9)	559 (4.1)	13,474 (100.0)
North	4,340 (95.8)	192 (4.2)	4,532 (100.0)
Northeast	10,595 (95.7)	472 (4.3)	11,067 (100.0)
Southeast	28,824 (95.5)	1,353 (4.5)	30,177 (100.0)
Total	60,362 (95.8)	2,671 (4.2)	63,033 (100.0)

Source: SIM - SIH, 2008 to 2012.
* Reported: SIM records with AIDS as the underlying cause of death. ** Underreported: SIM records that represent the sum of the 4 types of AIDS underreported deaths defined in the study.

Table 2. Distribution of AIDS deaths per year and types of underreporting. Brazil.

Year	Types of underreported AIDS deaths				
	Type A* N (%)	Type B** N (%)	Type C*** N (%)	Type D**** N (%)	Total N (%)
2008	142 (33.4)	16 (3.8)	44 (10.3)	223 (52.5)	425 (100.0)
2009	203 (35.5)	16 (2.8)	66 (11.6)	285 (50.0)	570 (100.0)
2010	229 (38.7)	9 (1.5)	66 (11.2)	287 (48.6)	591 (100.0)
2011	200 (37.7)	18 (3.4)	67 (12.6)	246 (46.3)	531 (100.0)
2012	232 (41.9)	18 (3.2)	76 (13.7)	228 (41.2)	554 (100.0)
Total	1006 (37.7)	77 (2.9)	319 (11.9)	1269 (47.5)	2.671 (100.0)

Source: SIM - SIH, 2008 to 2012.
* Type A (death with an AIDS predictive disease listed as the UCD and a hospitalization for HIV / AIDS); ** Type B (death with an AIDS presumptive disease listed as the UCD and a hospitalization for HIV / AIDS); *** Type C (death with an ill-defined UCD and a hospitalization for HIV / AIDS); **** Type D (death with a current hospitalization due to AIDS, but not type A, B or C).

during the period from 2008 to 2012, which corresponds to a percentage of 4.2% of the total of AIDS deaths in the period. When the data were detailed by Brazilian regions, the highest underreporting of deaths was recorded in the Southeast region, with 4.5% and the lowest in the Center-West region, with 2.5%. These differences by region can be explained by regional differences in care capacity for people with HIV observed in large urban centers of the country, generally higher in the municipalities of more developed regions. In contrast, municipalities in less developed regions have a growing burden of morbidity and mortality and a relatively low capacity to diagnose and assist HIV patients. The disease is often unknown to the patient and / or health professionals, due to lack of diagnosis or due to the strong social stigma[29]. In the Center-West, most cases are concentrated in the cities of Brasilia and Goiânia, with good public health care networks and good surveillance services. Among the four types of underreporting considered in the study methodology, those that had the highest proportions were type D (47.5%), followed by type A (37.7%). The increasing trends of types A and C and decreasing type D need to be interpreted with caution, due to the constraints of the analyzed period, that is, the linkage was only done within the study period and, therefore, did not include the adjacent years. For types A, B, and C, the more years a patient had of hospitalizations analysed prior to death, the more likely it would be to find previous AIDS hospitalizations, which is not true for type D. In any case, increased underreporting of types A and C, in the presence of AIDS predictive or ill-defined conditions, serve as a warning for certifying physicians to consider the possibility of AIDS as the underlying cause, indicating that they should actively seek information on the presence of this disease in the patient's history and medical records.

In Brazil, there are no studies with a similar methodology at the moment, but there are some studies using different methodologies that aim to quantify the underreporting of AIDS deaths. Fazito et al.[16], using a redistribution method based on SIM data and adjusting for the overall level of completeness, found 25% of underreporting for the period from 2000 to 2006. Pacheco et al.[14], linking SIM data with two AIDS patients' cohort registries from Rio de Janeiro state for the same period, found 23% of underreporting. Ciríaco[30], linking SIM data with the databases of SINAN, SICLOM and SISCEL in Maceió, the state capital of Alagoas, followed by an investigation of the linked records, found 9.8% of underreporting out of 204 total AIDS deaths of individuals from 20 to 49 years of age.

For the AIDS mortality rates in the period from 2008 to 2012, the average was 6.3 using the deaths reported in the SIM. However, when we recalculated the rates in our study adding the underreported deaths, the average for the period increased to 6.6 (per 100,000 inhabitants). The study of Guimarães et al.[31] compared the official mortality rates used by the Brazilian Ministry of Health and the estimated rates produced in the GBD-2015 study, which are based on complex modelling using multiple data sources. The mortality rates presented by the Ministry of Health ranged from 6.3 in 2000 to 5.7 in 2014, and the GBD-2015 study ranged from 9.6 in 2000 to 9.5 in 2015 (per 100,000 inhabitants). Our values, although still far from the values estimated by the

GBD, are higher than the official mortality rates used by the Ministry of Health.

There are methodological issues that may explain, at least partially, the lower percentage of underreporting found in our study as compared to the studies mentioned above. The underreporting of AIDS deaths refers to those who had been registered in the SIM as a wrongly classified cause, not to those who had not been registered. As for the SIH database, although it is a national database, it does not include about 20% of the total number of hospitalizations attributed to private hospitals not agreed by SUS. In this study, it was observed that 8.4% of deaths due to AIDS in the SIM occurred outside the hospital, which suggests that there may be underreported deaths occurring outside the hospital throughout the country[32]. Linkage procedures are never perfect[23]. For this process to be ideal, it would require a unique identifier as well as the integrity of the databases to be linked. Another point to consider is that there are limitations in the ICD-10 coding rules. For example, visceral leishmaniasis (VL) is a disease classically associated with AIDS immunodeficiency and is not included in the coding rules of ICD-10[33]. Deaths due to VL that were found had previous hospitalizations due to AIDS and could be counted as underreported AIDS deaths, if the rules were in fact inclusive. The rules also do not mention what to do when the code R75 ICD-10 (laboratory evidence of HIV) is mentioned in the lines of causes of the DC. This code should not be listed but, in reality, it was listed in some DC and the coders ended up using this information, along with other causes, to define AIDS as the underlying cause of death. In other words, the rules do not cover all possibilities.

Deaths that we classified as type D underreporting (death with a current hospitalization due to AIDS, but not type A, B or C), can be considered quite inclusive, since, with the exception of external causes, all other causes were classified as AIDS. This is because the main AIDS diagnosis in the hospitalization registry in which the death occurred was considered important enough to be selected in the study methodology as the underlying cause of death[34,35].

There may be some explanations contributing for the underreporting of AIDS deaths in the country. First, AIDS may not have been diagnosed in the hospitalization where the death occurred, only in previous hospitalizations, so the certifying physician did not have this information and therefore did not place it in the DC. This may have occurred in patients who had not yet been diagnosed with the infection/disease, or due to lack of specific HIV/AIDS data in the patient's medical records. Not all people living with HIV in Brazil know their HIV status, only about 80-85%. By 2020, one of the UNAIDS targets is that 90% of all people living with HIV should have their HIV status know[36]. Some studies point to a worrying tendency to achieve a reduction in AIDS deaths in the country, explained in part by the late diagnosis of HIV, with over 10% of HIV diagnoses being made only at death, in addition to low adherence to treatment[37]. Second, the certifying physician may have been aware that the patient died of AIDS but deliberately decided to omit this information from the DC. There may be a variety of reasons for this. The decedents' family members may have requested their doctors not to do so, because of the strong social stigma attached to HIV/AIDS, or for health or life insurance issues[38]. Third, the certifying physician may have been aware that the patient died of AIDS but neglected or simply was not able to adequately complete the DC. It is unfortunately rather common for the physician to list only the mechanism of death rather than the UCD (for example, cardiorespiratory arrest). Incomplete or inaccurate records of the cause of death can also occur due to lack of time, lack of interest of the doctors in the patient's history and lack of knowledge of these professionals on the correct completion of the DC and on its importance for establishing public health priorities[34,38].

In conclusion, this study demonstrated that, using the linkage in between the SIH and the SIM, it is possible to find underreported AIDS deaths in Brazil. Our results serve as an alert for the need of actions to promote better certification of causes of death among HIV/AIDS patients and to continue working towards increasing the reliability of the SIM records. The study can be repeated over time and complemented with other databases in addition to the SIH, such as private hospital or insurance databases.

Collaborations

The authors RA Carmo and AL Bierrenbach worked on the research, on the methodology and in obtaining the results, GM Policena in the linkage process for obtaining the research data, GP Alencar in the discussion of the results and in the final wording and EB França in the approval of the project for the development of research.

References

1. Marins JR, Jamal LF, Chen SY, Barros MB, Hudes ES, Barbosa AA, Chequer P, Teixeira PR, Hearst N. Dramatic improvement in survival among adult Brazilian AIDS patients. *AIDS (London, England)* 2003; 17(11):1675-1682.
2. Tancredi MV, Waldman EA. Survival of AIDS patients in Sao Paulo-Brazil in the pre- and post-HAART eras: a cohort study. *BMC infectious diseases* 2014; 14:599.
3. Luz PM, Girouard MP, Grinsztejn B, Freedberg KA, Veloso VG, Losina E, Struchiner CJ, MacLean RL, Parker RA, Paltiel AD, Walensky RP. Survival benefits of antiretroviral therapy in Brazil: a model-based analysis. *J Int AIDS Soc* 2016; 19(1):20623.
4. da Saúde. Boletim Epidemiológico HIV Aids. Brasília, DF: O Ministério; 2017.
5. Mathers CD, Fat DM, Inoue M, Rao C, Lopez AD. Counting the dead and what they died from: an assessment of the global status of cause of death data. *Bulletin of the World Health Organization* 2005; 83(3):171-177.
6. Bittencourt SA, Camacho LA, Leal Mdo C. Hospital Information Systems and their application in public health [in Portuguese]. *Cad Saude Publica* 2006; 22(1):19-30.
7. França E, Abreu DX, Rao C, Lopez AD. Evaluation of cause-of-death statistics for Brazil, 2002-2004. *Int J Epidemiol* 2008; 37(4):891-901.
8. Barchielli A, Buiatti E, Galanti C, Giovanetti L, Acciai S, Lazzeri V. Completeness of AIDS reporting and quality of AIDS death certification in Tuscany (Italy):a linkage study between surveillance system of cases and death certificates. *Eur J Epidemiol* 1995; 11(5):513-517.
9. Electronic record linkage to identify deaths among persons with AIDS--District of Columbia, 2000-2005. *MMWR Morbidity and mortality weekly report* 2008; 57(23):631-634. [acessado 2018 Set 2]. Disponível em: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5723a4.htm
10. Bessa Ferreira VM, Portela MC. [Evaluation of under-reporting of AIDS cases in the city of Rio de Janeiro based on data from the hospital information system of the Unified Health System]. *Cad Saude Publica* 1999; 15(2):317-324.
11. Fonseca MGP, Coeli CM, Lucena FFA, Veloso VG, Carvalho MS. Accuracy of a probabilistic record linkage strategy applied to identify deaths among cases reported to the Brazilian AIDS surveillance database. *Cad Saude Publica* 2010; 26(7):1431-1438.
12. Rique J, Silva MD. [Study of the under-reporting of AIDS cases in Alagoas (Brazil), 1999-2005]. *Cien Saude Colet* 2011; 16(2):599-603.
13. Krentz HB, Kliewer G, Gill MJ. Changing mortality rates and causes of death for HIV-infected individuals living in Southern Alberta, Canada from 1984 to 2003. *HIV medicine* 2005; 6(2):99-106.
14. Pacheco AG, Saraceni V, Tuboi SH, Lauria LM, Moulton LH, Faulhaber JC, Golub JE, Durovni B, Cavalcante S, Harrison LH, Chaisson RE, Schechter M. Estimating the extent of underreporting of mortality among HIV-infected individuals in Rio de Janeiro, Brazil. *AIDS research and human retroviruses* 2011; 27(1):25-28.
15. Birnbaum JK, Murray CJ, Lozano R. Exposing misclassified HIV/AIDS deaths in South Africa. *Bull World Health Organization* 2011; 89(4):278-285.
16. Fazito E, Cuchi P, Fat DM, Ghys PD, Pereira MG, Vasconcelos AMN, Pascom ARP. Identifying and quantifying misclassified and under-reported AIDS deaths in Brazil: a retrospective analysis from 1985 to 2009. *Sex Transm Infect* 2012; 88(Supl. 2):i86-94.
17. Health Organization (WHO). *Assessing tuberculosis under reporting through inventory studies*. Geneva WHO; 2012.
18. World Health Organization (WHO). *ICD-10 version: 2010. International Statistical Classification of Diseases and Related Health Problems, 10th Revision*. Geneva: WHO; 2010.
19. Pinheiro CE, Santo AH. Batch processing of causes of death based on the underlying cause of death selection system. *Rev Saude Publica* 1998; 32(1):72-73.
20. GBD 2016 Brazil Collaborators. Burden of disease in Brazil, 1990-2016: a systematic subnational analysis for the Global Burden of Disease Study 2016. *Lancet* 2018; 392(10149):760-775.
21. Schnell R, Bachteler T, Reiher J. Privacy-preserving record linkage using Bloom filters. *BMC Med Inform Decis Mak* 2009; 9(41):1-11.
22. Pacheco AG, Saraceni V, Tuboi SH, Lauria LM, Moulton LH, Faulhaber JC, Golub JE, Durovni B, Cavalcante S, Harrison LH, Chaisson RE, Schechter M. Validation of a hierarchical deterministic record-linkage algorithm using data from 2 different cohorts of human immunodeficiency virus-infected persons and mortality databases in Brazil. *Am J Epidemiol* 2008; 168(11):1326-1332.
23. Oliveira GP, Bierrenbach AL, Camargo Júnior KR, Coeli CM, Pinheiro RS. Accuracy of probabilistic and deterministic record linkage: the case of tuberculosis. *Rev Saude Publica* 2016; 50:49.
24. Machado JP, Martins M, Leite ID. Quality of hospital databases in Brazil: some elements. *Rev Bras Epidemiol* 2016; 19(3):567-581.
25. Bittencourt SA, Camacho LA, Leal Mdo C. Hospital Information Systems and their application in public health. *Cad Saude Publica* 2006; 22(1):19-30.
26. Ministério da Saúde (MS). *A experiência brasileira em sistemas de informação em saúde*. Brasília: MS; 2009. Vol. 1.
27. Centro Brasileiro de Codificação de Doenças (CBCD). *Manual de Treinamento de Codificação em Mortalidade (Aluno)*. São Paulo: Universidade de São Paulo; 2012.
28. Brasil. Ministério da Saúde (MS). *Guia de Vigilância em Saúde: volume 2*. Brasília: MS; 2017.
29. Grangeiro A, Escuder MML, Castilho EA. Evaluation of strategies by the Brazilian Ministry of Health to stimulate the municipal response to AIDS. *Cad Saude Publica* 2011; 27(Supl 1):114-128.
30. Ciríaco DL. *Subrregistro de óbitos por AIDS: investigação de óbitos por causas indeterminadas ou com diagnósticos sugestivos de imunodeficiência adquirida na região metropolitana de Maceió, Alagoas* [dissertação]. Recife: Centro de Pesquisas Aggeu Magalhães; 2010.

31. Guimarães MDC, Carneiro M, Abreu DMX, França EB. HIV/AIDS Mortality in Brazil, 2000-2015: Are there reasons for concern? *Revista brasileira de epidemiologia - Brazilian journal of epidemiology* 2017; 20 Suppl 01(Suppl 01):182-190.
32. Rezende EL, Vasconcelos AM, Pereira MG. Causes of death among people living with HIV/AIDS in Brazil. *Braz J Infect Dis* 2010; 14(6):558-563.
33. Martins-Melo FR, Lima MS, Alencar CH, Ramos AN Jr, Heukelbach J. Epidemiological patterns of mortality due to visceral leishmaniasis and HIV/AIDS co -infection in Brazil, 2000-2011. *Trans R Soc Trop Med Hyg* 2014; 108(6):338-347.
34. Cunha CC, Teixeira R, Franca E. Assessment of the investigation of ill-defined causes of death in Brazil in 2010. *Epidemiologia e Servicos de Saude: Revista do Sistema Unico de Saude do Brasil* 2017; 26(1):19-30.
35. Fazito E, Vasconcelos AM, Pereira MG, Rezende DF. Trends in non-AIDS-related causes of death among adults with HIV/AIDS, Brazil, 1999 to 2010. *Cad Saude Publica* 2013; 29(8):1644-1653.
36. Joint United Nations Programme on HIV/AIDS (UNAIDS). UNAIDS data 2017. [acessado 2018 Set 2]. Disponível em: http://www.unaids.org/en/resources/documents/2017/90-90-90
37. Bonolo PF, Machado CJ, César CC, Ceccato MGB, Guimarães MDC. Vulnerability and non-adherence to antiretroviral therapy among HIV patients, Minas Gerais State, Brazil. *Cad Saude Publica* 2008; 24(11):2603-2613.
38. Rampatige R, Mikkelsen L, Hernandez B, Riley I, Lopez A. Hospital cause-of-death statistics: what should we make of them? *Bull World Health Organization* 2014; 92(1):3-3A

Article submitted 10/12/2018
Approved 27/06/2019
Final version submitted 29/06/2019

Chief Editors: Romeu Gomes, Antônio Augusto Moura da Silva