ABSTRACT: **Objective:** To propose a method for improving mortality estimates from noncommunicable diseases (NCD), including the redistribution of garbage codes in Brazilian municipalities. **Methods:** Brazilian Mortality Information System (MIS) was used as a data source to estimate age standardized mortality rates, before and after correction, for NCD (cardiovascular, chronic respiratory, diabetes, and neoplasms). The treatment for the correction of data addressed missing data, under-registration, and redistribution of garbage codes (GCs). Three-year periods 2010–2012 and 2015–2017, and the Bayesian method were used to estimate mortality rates, reducing the effect of fluctuation caused by small numbers at the municipal level. **Results:** GCs redistribution step showed greater weight in corrections, about 40% in 2000 and roughly 20% as from 2007, with stabilization starting in this year. Throughout the historical series, the quality of information on causes of death has improved in Brazil, with heterogeneous results being observed among municipalities. **Conclusion:** Methodological studies that propose correction and improvement of the MIS are essential for monitoring mortality rates due to NCD at regional levels. The methodological proposal applied, for the first time in real data from Brazilian municipalities, is challenging and deserves further improvements. Improving the quality of the data is essential in order to build more accurate estimates based on the raw data from the SIM. **Keywords:** Noncommunicable diseases. Data accuracy. Cause of death. Small-area analysis.
INTRODUCTION

Noncommunicable diseases (NCD) are responsible for approximately 40 million annual deaths worldwide, and the vast majority occur in low and middle income countries, largely affecting premature mortality, under 70 years of age.\(^1\) Despite the decrease in standardized mortality rates, the scenario in Brazil is no different from that observed in the rest of the world, and NCD represent about 75% of total deaths that occurred in the country.\(^2,3\) Because of the magnitude of NCD, target 3.4, regarding health and well-being, was included in the Sustainable Development Goals (SDG) in 2015, to be achieved by 2030: to reduce premature mortality from NCD (diabetes, cardiovascular diseases, respiratory diseases, and neoplasms), with prevention and treatment, and foster mental health.\(^4\) Therefore, monitoring these causes of death is essential.

For tracking NCD, continuous follow-up with epidemiological surveys and health information systems for variables such as risk factors, morbidity, and social determinants is recommended. Vital statistics on causes of death are essential for the epidemiological knowledge of the population’s health situation and, especially, for tracking NCD. Information generated from such data serves as a subsidy for public management through analyzes for health planning, monitoring, and evaluation.
Even though they are aware of the relevance of these data, few low- and middle-income countries have mortality systems with desirable coverage and quality, which are essential characteristics to generate information reliability.\(^5\)

In Brazil, Mortality Information System (MIS) is responsible for capturing, storing, and making this data available in the country. Created in 1975 by the Brazilian Ministry of Health, despite being considered a consolidated system, MIS has heterogeneous characteristics with regard to the quality of its data, especially in the analysis of coverage indicators and causes of death. Thus, usual indicators of underreporting of deaths and the high ratio of ill-defined causes and garbage codes present different scenarios in the regions and states of Brazil, in addition to differences according to socioeconomic and demographic levels.\(^6\)\(^,\)\(^8\)

Death is the result of a chain of events that must be considered in their logical sequence. However, for public health mortality statistics, a death is represented by a cause of death, which must be defined by the cause that started the chain of events, called the basic cause. The declaration of the basic cause in a death certificate (DC) must be made by a physician, a technically qualified professional and capable of giving accuracy to the definition of the cause that culminated in death. Analyzes on mortality are conducted in countries based on this declaration.

For this reason, the concern with the quality of information on the cause of death is a widely discussed issue. In this sense, new definitions of non-specific causes of death have been addressed.\(^9\) Garbage causes (GCs) represent a group of nonspecific causes with no relevance to public health, since identifying actions for prevention or control is not feasible.\(^10\)

In addition to chapter 18 of the tenth revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10), all other chapters have codes considered to be GCs. An example are deaths from malignant neoplasms without a specified location (C80), which is not considered a basic cause of death, as the disease that started the chain of events culminating in death is not defined.

Due to the limitations of mortality databases in the country, in order to obtain estimates for more precise causes, correcting the underreporting of deaths, as well as the redistribution of GCs, is recommended to be done in a regionalized manner and over time.\(^11\)\(^,\)\(^12\) There are methodological proposals for working with these causes, including the proportional redistribution of ill-defined causes (IDC) or the total of GCs.\(^11\)\(^,\)\(^13\) In the Global Burden of Disease (GBD) study, this step is fundamental for the treatment of the causes of death and uses weights generated by statistical models and redistributed by algorithms among the group of defined causes (target) as a basis.\(^14\)

Since there are no analyzes that consider correction methods for mortality data at municipal levels, the present study aimed to propose a method for improving the estimates of causes of death from NCD, including the redistribution of GCs in Brazilian municipalities, which will support the country’s effort to monitor the targets of reducing such diseases.
METHODS

DATABASE

Open data from MIS between 2000 and 2017 were used, which do not contain personal information, making it feasible to prepare studies without the need for approval by the Research Ethics Committee Involving Human Beings.

DATABASE TREATMENT

In order to standardize and improve the accuracy of mortality data, the correction method for the suggested MIS data was carried out in steps.

Step 1: Redistribution of missing data

Despite the completeness of MIS being currently considered adequate,\textsuperscript{15} it varies among municipalities over a longer period and at the municipal level. Thus, a proportional redistribution between deaths was made with information on the following missing or incomplete variables: municipality of residence, age, and sex. Deaths without information as to municipality of residence contained information from their Federative Unit (FU).

Step 2: Correction of MIS underreporting

State correction coefficients of the 2017 GBD study were used, according to age and sex. Each municipality was corrected for its respective state coefficient, which was applied only in those municipalities with a general mortality rate of less than five deaths per 1,000 inhabitants.\textsuperscript{2}

Step 3: Redistribution of GCs

Definitions of the groups of causes followed the list of the 2017 GBD study.\textsuperscript{14} GBD classification considers three major groups of diseases:

- NCD;
- communicable, maternal, neonatal, and nutritional diseases;
- external causes.

All groups are hierarchical and can be broken down to level 4. The classification of GCs is conceptually divided into four levels. Level 1 considers GCs to be redistributed among
any of the three major groups of GBD causes, such as septicemia, which may be due to an accident or even pneumonia. Accordingly, level 2 GCs can be redistributed at most to a level 2 group, such as unspecified gastrointestinal tract hemorrhage, which must be redistributed in the group of NCD. On the other hand, level 3 GCs refer to causes such as unspecified cancer, which will be attributed to cancer and will be redistributed to the level 3 group of specific causes of cancer. Finally, level 4 GCs refer to a defined level 4 cause, such as unspecified stroke, which can be redistributed into ischemic or hemorrhagic, and diabetes, which can be redistributed as type 1 or 2 diabetes.

The process of redistributing GCs followed the concept of GCs levels. An analysis of the codes was carried out to identify which are the GCs specifically related to NCD (diabetes, cardiovascular diseases, chronic respiratory diseases, and neoplasms) and the other specific causes. This process consisted of four steps, as detailed in Figure 1.

**Step 3.1: Definition of codes among groups studied**

In this stage, deaths whose basic causes were specified codes of the ICD-10 and defined by the 2017 GBD study were identified, considered as the target causes, for which GCs will be redistributed. The present study had NCD as its focus, composed of the four major groups of causes, as used in the target 3.4 of the SDGs, as previously mentioned. The analyzes considered the four groups of causes together and separately. However, no detailed breakdown for the subgroups in the considered causes was adopted. The other causes were grouped into other NCD, transmissible, maternal, neonatal and nutritional, and external causes.

---

**Figure 1. Flowchart of the redistribution of GCs to NCD.**

GCs: garbage causes; NCD: noncommunicable diseases; GBD: Global Burden Disease; NS: not specified.
Step 3.2: Redefinition of the groups of causes with the inclusion of levels 3 and 4 GCs

As these are codes with more specific information on the causes of death, CGs deaths from levels 3 and 4 were included in the large groups defined in step 3.1. For example, malignant neoplasm without specifying the location, a level 3 GC, was classified as a neoplasm in the study. This cause does not define the location of the primary tumor, hiding the real basic cause of death, and, according to the GBD study, it is redistributed to different specific types of cancers, such as breast, liver, prostate, among others. The same process was considered for the other levels 3 and 4 GCs that were classified as one of the other three major groups of causes of the analysis, except for unspecified pneumonias, of which 50% were redistributed into NCD, and 50%, into transmissible, maternal, neonatal, and nutritional causes.

Step 3.3: Redistribution of level 2 GCs

In this step, level 2 GCs were redistributed proportionally to the target cause groups, defined in step 3.1. The exceptions to this step were hypertension (I10 and I15), cor pulmonale (I27), atherosclerosis (I70), and arterial embolism (I74), which were classified as cardiovascular diseases.

Step 3.4: Redistribution of level 1 GCs

At this stage, level 1 GCs were redistributed proportionally among the groups of causes studied, with the exception of chronic respiratory failure (J96.1), included in the group of chronic respiratory patients. This step includes ill-defined causes (chapter 18 of ICD-10).

The list of GCs selected for each group is found in Appendix 1, as well as the exceptions in relation to the GC classification levels.

DATABASE TREATMENT: GEOGRAPHIC LEVEL OF ANALYSIS

The geographical unit considered for the study was municipalities, aiming to contemplate local characteristics for data correction.

In order to smooth the estimates of mortality rates for small areas, the analyzes of rate distributions were based on the local empirical Bayesian estimator (EBS), which considers the neighborhood structure to estimate municipal rates. In addition, the analysis by municipalities evaluated the three-year periods from 2010 to 2012 and 2015 to 2017. All rates considered in the study, both those that used raw data and those that used corrected data, were standardized by age.

RESULTS

Figures 2A, 2B, and 2C show the proportional increase in the number of deaths after data treatment according to the correction steps, missing data, under-registration, and...
redistribution of GCs in comparison with the total gross deaths from NCD in Brazil, from 2000 to 2017. The country has significantly improved as to the quality of mortality data over the years, since the ratio decreases over time in all stages. In addition, redistribution of GCs is the stage in which there was a greater increase in mortality after data treatment. A stabilization of 19% from 2007 onwards could be observed, for this correction stage.

Figure 2. Historical series of the proportional variation in the number of deaths after treatment, according to stages of redistribution of missing data, correction of under-registration and redistribution of GCs in deaths from chronic diseases. Brazil, 2000 to 2017.
Diabetes was the cause with the highest increase ratio (more than 40% in 2000). Respiratory diseases became the second most corrected cause in 2006, surpassing cardiovascular diseases. Neoplasms present the smallest corrections in the entire period (Figure 3).
Spatial distribution of the percentage changes in mortality rates with corrected and uncorrected data, according to three-year periods, is shown in Figure 4. In general, there is a clearing on the map for the most recent period, 2015 to 2017, which shows a decrease in variation. There are clusters with the highest ratios in the Northeastern and Northern regions. High correction is also seen in the Northern region of Minas Gerais State.

Figure 4. Maps of percentage variations in mortality rates due to NCD standardized with raw data and corrected data in municipalities. Brazil, (A) 2010 to 2012, (B) 2015 to 2017.
DISCUSSION

In the present study, the impact on the final estimates of mortality due to NCD was demonstrated after the application of a methodology for the redistribution of GCs in Brazilian data. NCD are the leading causes of death in the world and in Brazil,\textsuperscript{20,21} accounting for roughly two-thirds of total deaths. The inclusion of the target to reduce premature deaths from NCD by 30% by 2030, set in the 2030 Agenda of the United Nations, results in an important priority for monitoring these causes at more disaggregated levels, seeking to identify inequalities and act to correct these inequities.\textsuperscript{4,22,23} Therefore, methodological studies that propose the correction and improvement of the MIS become essential for tracking mortality rates at different geographic levels.

The appropriate methodology for analyzing small areas, which includes Bayesian methods and the use of triennial periods, allows using data at more disaggregated geographic levels. This is opportune, as it considers information about the local reality. Therefore, minimizing differences in the quality of data generated by death information systems in small areas and over time is feasible, thus allowing the obtention of more accurate estimates.

Although the quality of mortality data has been improved in the country, the present study shows the importance of comparing rates before and after data correction. Caution in the use of mortality estimates from direct methods with raw data is of utmost importance, since proportional increases of more than 20% were found in estimates comparison before and after the correction of NCD.

Data analyzed herein pointed to an improvement in the quality of mortality data in Brazilian municipalities over the last decade, evidenced by the reduction in percentage variations between mortality rates, with the use of raw and corrected data (missing data, under-registration and redistribution of GCs). These advances were the result of efforts by the Brazilian Ministry of Health in partnership with federative units and municipalities to better track deaths through MIS, such as the project to reduce ill-defined causes in 2005 and the project to reduce regional inequalities and infant mortality in the states of the Northeastern Region and the Legal Amazon.\textsuperscript{24-28} The proactive search of deaths project is highlighted, which made it possible to define methodologies for redistributing underreported deaths.\textsuperscript{29,30} This commitment, together with the corrections, was essential for the most adequate interpretation and comparability of the historical series in the different regions of the country.

In the analysis of the increase in rates after correction according to NCD causes, diabetes showed the highest correction ratio, reaching more than 40% in the year 2000, perhaps because many deaths were originally declared as unspecified diabetes (level 4-GC). On the other hand, since these are diseases that require specialized and highly complex care after diagnosis, neoplasms had the smallest increase in rates after correction.

Despite the improvement in the quality of the data,\textsuperscript{31} large regional increases were still observed in mortality rates estimates from NCD after corrections. In addition to the importance of continuously searching to improve the quality of information, in order to obtain...
more accurate estimates, less quality in filling out the causes of death in the death certificate may signal less access to health services and the quality of medical care.\textsuperscript{5,32}

The accuracy of mortality data is essential, since they serve to support the planning of health actions, the monitoring of disease trends, the evaluation of public policies, the identification of the most vulnerable populations, among others.

GCs are a relevant indicator for assessing the quality of the death information system, since high ratios can impair analyzes, especially when it comes to more stratified analyzes, such as specific causes, according to age and sex or in small areas. The analysis of time series must be conducted carefully especially when using raw data, seen that quality can vary over time, as pointed out in the present study, with progressive improvement of the MIS and consequent reduction of GCs, under-registration of deaths, etc.

The proportional distribution of ill-defined causes (chapter 18 of ICD-10) is used in several mortality studies.\textsuperscript{12,33,34} Other, more robust methods are available and provide better estimates. Reclassification of deaths based on death investigations, for example, is a widely used strategy.\textsuperscript{8,35} With this methodology, obtaining more accurate information about the individual’s cause of death is possible, because it uses information from medical records, cross-checking data from different sources and including home interviews with close relatives. Despite being considered a great strategy, death investigations require high investment for good results to be obtained.

Despite that, as well as the method proposed in this study, there are other methodological ways to treat GCs in mortality data without making local investigations. The GBD study proposes the redistribution of this group of causes in a robust manner, with an analytical methodology based on standardized algorithms for all regions studied. Its main step is the redistribution of GCs, considering redistribution weights generated according to sex, age, and specific cause in each of the locations studied.\textsuperscript{8,35} Other approaches aim to recover inaccurate information with simpler redistributions, in a proportional way, as is the case of R99 – other ill-defined and unspecified causes of mortality, and redistributions with the use of machine learning techniques, which have regression models, multiple cause redistribution techniques, among others.\textsuperscript{10,36}

Estimates generated by the GBD study have been used by researchers around the world. However, replicating this methodology is hampered by the absence of structure, computational and human resources, and knowledge generated by the group of the Institute of Metrics and Health Assessment. For this reason, looking for alternative methods applicable in the local context is important.

The applied methodology represents an advance in relation to the redistribution of ill-defined causes in a \textit{pro rata} way in the treatment of raw data to improve the quality of vital statistics, but it is still not enough. Among the study limits, under-registration of deaths due to the use of secondary bases is pointed out, as well as the validation of this correction without field research, making it exceedingly difficult to obtain real local information. The definition of the basic cause of death is extremely important information, registered by mortality systems. Therefore, the ideal would be filling out the death certificate with well-defined causes, by the corresponding doctors.
Another important limitation population estimates in the country may be subject to errors, in addition to the difficulties of working with numerators, that is, deaths, for small areas. However robust the demographic methods applied in population estimates are, distance from the last census makes validations more complex, since a series of assumptions must be made based on the distance in-between two censuses.

The methodological proposal applied, for the first time in real data from Brazilian municipalities, is challenging and subject to improvement. Comparisons with other proposals should be encouraged and tested. The application of GBD correction methods at the municipal level, the redistribution of ill-defined causes, and the analysis of investigations of deaths from GCs deserve further analysis so that the knowledge of several proposals can lead to the best final estimate of the country’s vital statistics at the level of small areas.

Finally, the results of the present study highlight the importance of redistributing GCs, especially those related to NCD, to obtain more accurate estimates of the risk of death in the country. NCD are a global priority and are included in the SDGs. Thus, in order to achieve the United Nations motto “Leave No One Behind (LNOB)”, investing in quality of data and advance in estimates that allow correction and analysis in municipalities and other smaller geographic areas is essential, because there are still large inequalities, difficulty in accessing services, and high mortality rates in these places.

By proposing a methodology based on local information, with empirical data, in a replicable way and tested at the municipal level, the present study proved to be an important management tool. Although the development of this methodology represents an advance in the use of mortality data in the country, efforts to universalize death records and improve the definition of causes of deaths must continue progressing. Improving the quality of mortality data is essential so that, in the future, building more accurate estimates based on the original MIS data is possible.

ACKNOWLEDGMENTS

The authors would like to thank Professor Elisabeth França, from the Postgraduate Program in Public Health at the Faculty of Medicine, for her guidance and support in the study.

REFERENCES


Received on: 11/10/2020
Revised on: 12/04/2020
Accepted on: 12/10/2021
Preprint on: 15/12/2020

Authors’ contributions: RAT, LHI, FM, and DCM conceived and designed the study. RAT developed the management, exploration, and analysis of the data, the elaboration and interpretation of results, and the discussion. EPJ assisted in writing and final proofreading. All authors critically reviewed the manuscript and contributed to the entire process. All authors read, contributed to the final manuscript, and approved it.