DOI: 10.1590/1980-5497201600040003

ORIGINAL ARTICLE / *ARTIGO ORIGINAL*

Use of primary diagnosis during hospitalization in the Unified Health System (Sistema Único de Saúde) to qualify information regarding the underlying cause of natural deaths among the elderly*

Uso do diagnóstico principal das internações do Sistema Único de Saúde para qualificar a informação sobre causa básica de mortes naturais em idosos

Angela Maria Cascão[I,II], Maria Helena Prado de Mello Jorge[III], Antonio José Leal Costa[IV], Pauline Lorena Kale[IV]

ABSTRACT: ***Introduction:*** Ill-defined causes of death are common among the elderly owing to the high frequency of comorbidities and, consequently, to the difficulty in defining the underlying cause of death. ***Objective:*** To analyze the validity and reliability of the "primary diagnosis" in hospitalization to recover the information on the underlying cause of death in natural deaths among the elderly whose deaths were originally assigned to "ill-defined cause" in their Death Certificate. The hospitalizations occurred in the state of Rio de Janeiro, in 2006. ***Methods:*** The databases obtained in the Information Systems on Mortality and Hospitalization were probabilistically linked. The following data were calculated for hospitalizations of the elderly that evolved into deaths with a natural cause: concordance percentages, Kappa coefficient, sensitivity, specificity, and the positive predictive value of the primary diagnosis. Deaths related to "ill-defined causes" were assigned to a new cause, which was defined based on the primary diagnosis. ***Results:*** The reliability of the primary diagnosis was good, according to the total percentage of consistency (50.2%), and fair, according to the Kappa coefficient ($k = 0.4$; $p < 0.0001$). Diseases related to the circulatory system and neoplasia occurred with the highest frequency among the deaths and the hospitalizations and presented a higher consistency of positive predictive values per chapter and grouping of the International Classification of Diseases. The recovery of the information on the primary cause occurred in 22.6% of the deaths with ill-defined causes ($n = 14$). ***Conclusion:*** The methodology developed and applied for the recovery of the information on the natural cause of death among the elderly in this study had the advantage of effectiveness and the reduction of costs compared to an investigation of the death that is recommended in situations of non-linked and low positive predictive values. Monitoring the mortality profile by the cause of death is necessary to periodically update the predictive values.

Keywords: Elderly. Mortality. Primary cause of death. Hospitalization. Information systems. Reproducibility of tests.

*This paper was part of the first author's PhD thesis in Public health, titled "Mortality among the Elderly: the development of a methodology to recover information regarding underlying cause," defended in April 2014 at the Institute of Public Health Studies at the *Universidade Federal do Rio de Janeiro*.

[I]Consultancy of Vital Data at the State Department of Health in Rio de Janeiro – Rio de Janeiro (RJ), Brazil.

[II]PhD program in Public Health at the Institute of Public Health Studies at the *Universidade Federal do Rio de Janeiro* – Rio de Janeiro (RJ), Brazil.

[III]School of Publich Health at the *Universidade de São Paulo* – São Paulo (SP), Brazil.

[IV]Institute of Public Health Studies at the *Universidade Federal do Rio de Janeiro* – Rio de Janeiro (RJ), Brazil.

Corresponding author: Angela Maria Cascão. *Universidade Federal do Rio de Janeiro*. Institute of Public Health Studies. Praça Jorge Machado Moreira, Ilha do Fundão, Cidade Universitária, CEP: 21941-598, Rio de Janeiro, RJ, Brazil. E-mail: angelacascao@uol.com.br

Conflict of interests: nothing to declare – **Financial support:** This study is a part of the research project "Health Information" from the School of Public Health at the *Universidade de São Paulo*, a project financed by the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq) (National Council of Scientific and Technological Development), process number 2006/8.

RESUMO: ***Introdução:*** Causas mal definidas de morte destacam-se entre idosos devido à alta frequência de comorbidades e consequente dificuldade de definir a causa básica. ***Objetivo:*** Analisar a validade e a confiabilidade da informação "diagnóstico principal" da internação hospitalar para recuperação da causa de morte natural de idosos que tiveram originalmente "causas mal definidas" como causa básica nas Declarações de Óbito, ocorridas no estado do Rio de Janeiro, em 2006. ***Métodos:*** As bases de dados obtidas dos Sistemas de Informações sobre Mortalidade e de Internação Hospitalar foram relacionadas probabilisticamente. Foram calculados percentuais de concordância, coeficiente Kappa, sensibilidade, especificidade e valor preditivo positivo do diagnóstico principal da internação de idosos que evoluíram para óbito por causas naturais. Óbitos por "causas mal definidas" tiveram uma nova causa definida a partir do diagnóstico principal. ***Resultados:*** A confiabilidade do diagnóstico principal foi boa, segundo a concordância percentual total (50,2%), e razoável, conforme o coeficiente Kappa ($k = 0,4$; $p < 0,0001$). Doenças do aparelho circulatório e neoplasias ocorreram com maior frequência entre os óbitos e as internações e apresentaram maior concordância e valores preditivos positivos por capítulo e agrupamento da Classificação Internacional de Doenças. A recuperação da causa básica ocorreu em 22,6% dos óbitos por causas mal definidas ($n = 14$). ***Conclusão:*** A metodologia desenvolvida e aplicada para recuperação da causa natural de óbito entre idosos neste estudo tem como vantagens a efetividade e a redução dos custos provenientes de uma investigação do óbito. Esta é recomendada nas situações de registros não relacionados e de baixo valor preditivo positivo. O monitoramento do perfil de mortalidade por causas é necessário para atualização periódica dos valores preditivos.

Palavras-chave: Idoso. Mortalidade. Causa básica de morte. Hospitalização. Sistemas de informação. Reprodutibilidade dos testes.

INTRODUCTION

The steady rise of the elderly population growth rate in Brazil calls for public social and health policies aimed at this population contingent[1,2], which in itself justifies a more detailed study regarding the mortality that affects the elderly. In 2010, the population contingent of elderly people in the country was 10.8% and, in the state of Rio de Janeiro, approximately 13.0%[1].

With the population aging, the higher rate of morbidity and mortality moves from the youngest part of the population to the elderly[2]. There is a significant increase in the occurrence of cardiovascular disease, neoplasias, and respiratory diseases, the main causes of hospitalizations and deaths among the elderly[3-5]. Ill-defined causes can also be highlighted among the main causes of deaths[4,6], partially owing to the difficulty in identifying the underlying cause of the morbidity process, because of the complex chain of events that lead to death in the elderly[7] population and the high frequency of comorbidities in this age stratum[8]. Socioeconomic, demographic, and geographic inequalities such as education level, race/color, size of the municipality, and per capita gross domestic product also have shown to be linked to ill-defined causes of death in elderly people[9]. Strategies to qualify the causes of death have been developed and utilized to improve the description of the health status[10-13], subsidizing measures of prevention[14], particularly for the elderly[15].

In Brazil, the Mortality Information System (SIM) is universal and has a broad coverage. The Death Certificate is the basic form that provides, among other informations, the cause of death. The official source for the hospital morbidity is the Hospitalization System from the Unified Health System (SIH-SUS), which, therefore, does not include private hospitalization services. In the Authorization for Hospitalization (AIH) — the basic document from the SIH-SUS, the main affection being treated or investigated is informed, and it is understood as the "primary diagnosis." The SIH-SUS is an important source of information and it has been increasingly used in studies concerning the qualification of the SIM and in monitoring indicators for the analysis of the health status in the population.

The aim of this study is to assess the reliability and the validity of the primary diagnosis of the AIH to retrieve the underlying cause of natural death of the elderly individuals who originally had ill-defined cause assigned as their underlying cause of death in the Death Certificate of the state of Rio de Janeiro, in 2006.

METHODS

This study is linked to the research "Health Information" from the School of Public Health of the *Universidade de São Paulo*, with the participation of technicians from the State Department of Health in Rio de Janeiro (SESRJ).

This is a case series study covering deaths of elderly individuals (aged 60 and older)[16] owing to natural causes and which occurred during hospitalization in the SUS healthcare facilities. The maximum duration of the hospitalization when the discharge was due to "death" was equivalent to one year in this study. Data were obtained from the information system SIM and SIH-SUS, both from the SESRJ, and were probabilistically linked using the software RecLink II[17]. The following record linkages were conducted: standardization of common fields (name of individual, age, date of birth, date of death, and discharge due to death, home address); definition of the strategy for blocking the databases to optimize the comparison of records (blocking expressions: SOUDEX (PBL OCO) + SOUDEX (UBLOCO) + SEXO + CODMUNIRES + C NES); application of algorithms for the approximate comparison of the chain of characters, which take into consideration possible phonetic and typing errors; manual review of probable pairs (matching or non-matching). After the process for probabilistic linkage, 15,804 records were considered as probable matching pairs[18].

Distribution of related records according to underlying cause and primary diagnosis by groups of causes were described (chapters from the 10th revision of the International Statistical Classification of Diseases and Related Health Problems, ICD 10)[19].

Agreement and validity of the primary diagnosis were analyzed (chapters from ICD 10 – Chap). The following was excluded from the analysis: records of hospitalizations that evolved into death due to underlying causes classified as "external causes of morbidity and mortality," which correspond to unnatural causes, accidents or violence (Chap. XX),

analyzed in another paper[18]; "pregnancy, childbirth, and the puerperium" (Chap. XV), "certain conditions originating in the perinatal period" (Chap. XVI), causes which are incompatible with the age group analyzed; "injury, poisoning and certain other consequences of external causes" (Chap. XIX); and "factors influencing health status and contact with health services" (Chap. XXI), as they cannot be certified with an underlying cause of death.

For the reliability and validity analyses, binary variables that refer to the underlying cause of death and the primary diagnosis of the hospitalization were created: yes, classified in a chapter or specific grouping of the ICD 10; no, classified in other chapters or groupings.

Initially, for the reliability analyses concerning the primary diagnosis and the underlying cause of death according to the chapter from ICD 10, the following was calculated: percentage and positive percentage agreement[20] (< 25% – poor; from 25.0 to 49.0% – regular; from 50.0 to 74.0% – good; and ≥ 75.0% – very good)[21] and the Kappa coefficient (< 0.00 – poor; 0.00 to 0.20 – superficial; 0.21 to 0.40 – fair; 0.41 to 0.60 – moderate; 0.61 to 0.80 – substantial; 0.81 to 1.00 – nearly perfect)[22]. For the causes (chapter from ICD 10) that presented the highest agreement between the primary diagnosis and underlying cause, the percentage and positive percentage agreement were calculated by group of cause (ICD 10).

The validity of the primary diagnosis by cause (chapter from the ICD 10) was assessed by sensitivity and specificity[23], with the gold standard being the original underlying cause in the Death Certificate. Positive predictive value (PPV) was calculated to assess the quality of the primary diagnosis in recognizing the underlying cause of death.

Later, for the deaths with an underlying cause that was originally certified as an ill-defined cause (Chap. XVIII — "Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified — 'ill-defined'") and whose primary diagnosis informed for the hospitalization belongs to the group of causes (chapters from ICD) that, in the previous analysis, presented acceptable reliability, validity, and PPV had their underlying cause modified for the purpose of analysis in this study.

The computer program Stata (version 12) was utilized. This study was approved by the Research Ethics Committee from the Institute of Public Health Studies at the *Universidade Federal do Rio de Janeiro*, under protocol number 114.623, on October 3, 2012.

RESULTS

In 2006, in the state of Rio de Janeiro, 119,513 non-fetal deaths occurred, with 73,323 (61.3%) being among elderly people, of which 54.9% (40,227) had been hospitalized in a public facility or through SUS. As for the SUS hospitalizations, 25.4% (178,190) were elderly individuals, and in 12.7% (22,580) of them, the discharge in the AIH (Hospitalization Authorization) was due to death. After comparing the databases, it was possible to identify 70.0% (15,807) of common records (related records). The 6,773 records unrelated to those from SIM represented 30.0% of the hospitalizations of elderly individuals with a discharge related to "death."

Table 1 contains a description of the related and unrelated records by age, sex, and primary diagnosis. Deaths related to natural and external causes were included. Taking into consideration that data tabulation was conducted according to the information from the primary diagnosis, deaths related to external causes are contained within the information on the group of causes: "injury, poisoning and certain other consequences of external causes" (Chap. XIX).

Related and unrelated records present a higher frequency in the age group 70 to 79 years (Table 1). Males were predominant among the related records (50.2%) and females among the unrelated records (52.4%). Diseases of the circulatory (Chap. IX) and the respiratory systems (Chap. X) held the first two positions in the ranking of primary diagnosis, regardless of the success of the link. The third and fourth positions were held by neoplasia (Chap. II) and certain infectious and parasitic diseases (Chap. I), respectively, with this order being inverted among the non-linked records (Table 1).

For the reliability and validity analysis, after the exclusions, the study population corresponded to 15,074 linked records, whose underlying causes of death were classified between Chapters I and XVIII of the ICD 10. For the underlying cause as well as the primary diagnosis of the related records, the three groups of causes with the highest frequency were diseases of the circulatory system (Chap. IX), neoplasias (Chap. II), and diseases of the respiratory system (Chap. X). The following were in fourth place: endocrine, nutritional, and metabolic diseases (Chap. IV) and certain infectious and parasitic diseases (Chap. I), for the underlying cause and the primary diagnosis, respectively (Table 2).

The total agreement between the underlying cause (SIM) and the primary diagnosis (SIH), according to the ICD chapters, was 50.2%; and the Kappa coefficient was around 0.4 ($p < 0.0001$) and considered, respectively, good (criteria used by Mello Jorge et al.[21]) and fair (Landis e Koch *apud* Szklo and Nieto)[22,23]. The highest values of positive percentage agreement (not considering the records for negative concordance) between the underlying cause and the primary diagnosis, not including diseases of the eye and adnexa (Chap. VII) due to its low frequency, correspond to the two main groups of causes, diseases of the circulatory system (Chap. IX) at 49.6% and neoplasias (Chap. II) at 49.2% (Table 2). As for the validity of the primary diagnosis, the sensitivity varied from 7.0 to 100.0% and the specificity from 83.8 to 99.8%. The specificity of the primary diagnosis presented values higher than the sensitivity for all of the ICD chapters. The primary diagnosis was equal to the underlying cause of death in 100.0% of records, when they were certified as diseases of the eye and adnexa (Chap. VII); however, they represent only two matching pairs pairs in a total of 15,074 records (0.01%). The specificity was 100.0% for the primary diagnosis when the causes belonged to the same chapter. The sensitivity of the primary diagnosis for the remaining chapters was low, under 62%. Specifically for the deaths due to ill-defined causes (Chap. XVIII), a low positive percentage agreement and low sensitivity represent good results, that is, the positive percentage agreement is a better-defined cause.

For the neoplasms (Chap. II) and the diseases of the cardiovascular system (Chap. IX), the sensitivities were, respectively, 51.5 and 61.6%; and the specificities also high, although not

Table 1. Hospitalization records, linked and non-linked to the death record by demographic and hospitalization characteristics, Rio de Janeiro, 2006.

Demographic and hospitalization characteristics	Linked (%) (n = 15.807)	Non-linked (%) (n = 6.773)
Age (years)		
60 to 69	28.8	32.7
70 to 79	37.3	36.1
80 and over	34.0	31.2
Sex		
Female	49.8	52.4
Male	50.2	47.6
Primary diagnosis (ICD 10 chapters)		
I – Certain infectious and parasitic diseases	9.9	10.4
II – Neoplasms	10.3	9.8
III – Diseases of the blood and immune system disorders	1.1	1.0
IV – Endocrine, nutritional and metabolic diseases	8.3	7.9
V – Mental and behavioral disorders	0.6	0.7
VI – Diseases of the nervous system	2.5	2.2
VII – Diseases of the eye and adnexa	–	–
IX –Diseases of the circulatory system	29.7	30.3
X –Diseases of the respiratory system	22.0	20.5
XI – Diseases of the digestive system	5.8	6.0
XII – Diseases of the skin and subcutaneous tissue	0.6	0.5
XIII – Diseases of the musculoskeletal system and connective tissue	0.4	0.4
XIV – Diseases of the genitourinary system	3.3	3.5
XVII – Congenital malformations	0.1	0.1
XVIII – Ill-defined causes	2.5	2.4
XIX – Injuries/poisoning/ other external consequences	2.8	3.9
XX – External causes	–	–
XXI – Factors influencing health status and contact with health services	0.1	0.2

Source: Hospitalization System from the Unified Health System (Sistemas de Informações sobre Mortalidade e de Internação Hospitalar do Sistema Único de Saúde - SIH-SUS).

Table 2. Underlying cause and primary diagnosis for the hospitalization of elderly individuals, matching pairs, positive total concordance, sensitivity, specificity, and positive predictive value, Rio de Janeiro, 2006.

Chapters (ICD 10)	UC n	PD n	CP n	PC %	TC %	SE %	SP %	PPV %
I – Certain infectious and parasitic diseases	686	1,528	252	12.8	88.7	36.7	91.1	16.5
II – Neoplasms	2,883	1,624	1,486	49.2	89.8	51.5	98.9	91.5
III – Diseases of the blood and immune system disorders	144	167	29	10.3	98.3	20.1	99.1	17.4
IV – Endocrine, nutritional, and metabolic diseases	1,641	1,289	473	19.3	86.8	28.8	93.9	36.7
V – Mental and behavioral disorders	84	90	13	8.1	99.0	15.5	99.5	14.4
VI – Diseases of the nervous system	230	383	41	10.7	96.5	17.8	97.7	10.7
VII – Diseases of the eye and adnexa	2	3	2	66.7	100.0	100.0	99.9	66.7
VIII – Diseases of the ear and mastoid process	2	–	–	–	–	–	–	–
IX – Diseases of the circulatory system	5,391	4,623	3,320	49.6	77.6	61.6	86.5	71.8
X – Diseases of the respiratory system	2,391	3,407	1,356	305	79.5	56.7	83.8	39.8
XI – Diseases of the digestive system	778	903	425	33.8	94.5	54.6	96.7	47.1
XII – Diseases of the skin and subcutaneous tissue	109	80	8	4.4	98.9	7.3	99.5	10.0
XIII – Diseases of the musculoskeletal system and connective tissue	40	62	7	7.4	99.4	17.5	99.6	11.3
XIV – Diseases of the genitourinary system	627	507	145	14.7	94.4	23.1	97.5	28.6
XVII – Congenital malformations	4	16	0	–	–	–	–	–
XVIII – Ill-defined causes	62	392	11	2.5	97.1	17.7	97.5	2.8
Total	15,074	15,074	7,568	–	50.21	–	–	–

UC: underlying cause; PD: primary diagnosis; CP: concordant pairs; TC: total concordance; PC: positive concordance; SE: sensitivity; SP: specificity; PPV: positive predictive value.
Source: Hospitalization System from the Unified Health System (*Sistemas de Informações sobre Mortalidade e de Internação Hospitalar do Sistema Único de Saúde* - SIH-SUS)

the highest, respectively, 98.9 and 86.5%. For every 100 hospitalizations with a primary diagnosis classified as neoplasia (Chap. II) and diseases of the cardiovascular system (Chap. IX), around 92 and 72, respectively, had a medical certificate of cause of death which declared underlying causes from the same groups of causes (ICD chapters). The direct relation between PPV and frequency of deaths in the study population can be observed in Table 2.

Considering the frequency of the causes of death and hospitalizations classified as neoplasias and diseases of the cardiovascular system, and the respective values of sensitivity, specificity, and PPV at the ICD chapter level, the grouping analysis for neoplasias (Chap. II) and diseases of the circulatory system (Chap. IX) were repeated. The results are presented, respectively, in Tables 3 and 4.

Among the 1,486 matching pairs classified as neoplasia, 18 groupings were identified, corresponding to a total concordance of approximately 81.0% (Table 3). Overall, the positive percentage agreement were high, with the values being higher for neoplasias of the lymphatic and hematopoietic tissue, of breasts and male genital organs. The sensitivity and the specificity by groupings were considered good, with few exceptions. The two groupings with the highest frequency for the underlying cause of death and primary diagnosis were malignant neoplasms of the digestive organs (C15-C26) and malignant neoplasms of the respiratory system (C30-C39), which presented the highest percentage of agreement, respectively, 27.2 e 11.5%. The sensitivity and the predictive value of the primary diagnosis of these groupings were high. The primary diagnosis classified in the grouping of malignant neoplasms of the lymphatic, hematopoietic, and related tissues (C81-C96) presented the highest sensitivity (94.1%) and a high PPV (93.1%). The following groupings presented high values of sensitivity and PPV: malignant neoplasms of breast (C50), malignant neoplasms of female genital organs (C51-C58), and malignant neoplasms of the male genital organ (C60-C63) with percentages of 89,6; 80,3 e 85,5%, and 90,9; 94,4 e 92,6%, respectively.

With regard to the 3,320 matching pairs which were classified as diseases of the circulatory system, nine groupings were found with a total concordance of 72.1% (Table 4). The agreement, sensitivity, and PPV percentage values per grouping from the chapter of diseases of the circulatory system were well below those found for the neoplasms. However, the cardiovascular diseases grouping (I60-I69) stands out by showing higher positive percentage of agreement, sensitivity, and specificity and a fair PPV. The PPV of the ischemic heart diseases (I20-I25) was the second highest of this chapter, with the remaining groupings showing values below 62% (Table 4).

Among the 62 linked records with deaths related to ill-defined causes (Chap. XVIII), 14 had a primary diagnosis classified within the two groups of causes with the highest PPV: 3 in neoplasias (Chap. II) and 11 in diseases of the circulatory system (Chap. IX). Considering only the large groups of causes (ICD chapter), it would be possible to reclassify the underlying cause as "well-defined" in 22.6% of deaths due to ill-defined causes (14 deaths) of the study population (Table 5).

With regard to the groupings with higher PPV among these groups of causes (chapters), it would be possible to recover information of three deaths: one related to malignant

Table 3. Underlying cause and primary diagnosis for the hospitalization of elderly individuals owing to neoplasia, matching pairs, positive total concordance, sensitivity, specificity, and positive predictive value, Rio de Janeiro, 2006.

Chapter II Groupings – Neoplasias (ICD 10)	UC n	PD n	CP n	PC %	TC %	SE %	SP %	PPV %
Malignant neoplasms, lip, oral cavity, and pharynx (C00-C14)	70	83	64	71.9	98.3	91.4	98.7	77.1
Malignant neoplasms, digestive organs (C15-C26)	482	442	404	77.7	92.2	83.8	96.2	91.4
Malignant neoplasms, respiratory system and intrathoracic organs (C30-C39)	208	190	171	75.3	96.2	82.2	98.5	90.0
Malignant neoplasms, bone and articular cartilage (C40-C41)	5	7	3	33.3	99.6	60.0	99.7	42.9
Malignant neoplasms, skin (C43-C44)	35	30	24	58.5	98.9	68.6	99.6	80.0
Malignant neoplasms, connective and soft tissue (C45-C49)	7	10	2	13.3	99.1	28.6	99.5	20.0
Malignant neoplasms, breast (C50)	134	132	120	82.2	98.3	89.6	99.1	90.9
Malignant neoplasms, female genital organs (C51-C58)	127	108	102	76.7	97.9	80.3	99.6	94.4
Malignant neoplasms of male genital organs (C60-C63)	117	108	100	80.0	98.3	85.5	99.4	92.6
Malignant neoplasms, urinary organs (C64-C68)	59	49	45	71.4	98.8	76.3	99.7	91.8
Malignant neoplasms, eye, brain, and central nervous system (C69-C72)	45	40	33	63.4	98.7	73.3	99.5	82.5
Malignant neoplasms, endocrine glands and related structures (C73-C75)	7	6	5	62.5	99.8	71.4	99.9	83.3
Malignant neoplasms, secondary and ill-defined (C76-C80)	69	119	28	17.5	91.2	40.6	93.6	23.5
Malignant neoplasms, lymphoid/ haematopoietic/related tissue (C81-C96)	101	102	95	88.0	99.1	94.1	99.5	93.1
Malignant neoplasms of independent (primary) multiple sites (C97)	10	4	–	–	99.1	–	99.7	–
In situ neoplasms (D00-D09)	1	7	–	–	99.5	–	99.5	–
Benign neoplasms (D10-D36)	1	14	–	–	99.0	–	99.1	–
Neoplasms of uncertain or unknown behavior (D37-D48)	8	35	6	–	97.9	75	98	17.1
Total	1,486	1,486	1,202	–	80.9	–	–	–

UC: underlying cause; PD: primary diagnosis; CP: concordant pairs; TC: total concordance; PC: positive concordance; SE: sensitivity; SP: specificity; PPV: positive predictive value.
Source: Hospitalization System from the Unified Health System (*Sistemas de Informações sobre Mortalidade e de Internação Hospitalar do Sistema Único de Saúde* - SIH-SUS).

neoplasms in the feminine genital organs (C51-C58), one related to the malignant neoplasms in the male genital organs (C60-C63), and one related to benign neoplasms (D10-D36). For diseases of the circulatory system, 11 information on deaths would be recovered, with 2 being related to ischemic heart diseases (I20-I25), 1 related to pulmonary heart disease and diseases of pulmonary circulation, 5 related to other forms of heart diseases (I30-I52), and 3 as cerebrovascular diseases (I60-I69), as seen in Table 5.

DISCUSSION

In this study, 70.0% of elderly deaths with hospitalization in SUS and discharge owing to "death" were linked (n = 22.580). Some hypothesis that were raised to explain the lack of a linkage between records in the analysis of unnatural causes from these

Table 4. Underlying cause and primary diagnosis for the hospitalization of elderly individuals owing to circulatory system diseases, matching pairs, positive total concordance, sensitivity, specificity and positive predictive value, Rio de Janeiro, 2006.

Chapter IX Groupings – Circulatory system diseases (ICD 10) Circulatory System Diseases	UC n	PD n	CP n	PC %	TC %	SE %	SP %	PPV %
Chronic rheumatic heart diseases (I05-I09)	18	13	8	34.8	99.5	44.4	99.9	61.5
Hypertensive diseases (I10-I15)	378	170	52	10.5	86.6	13.8	96.0	30.6
Ischemic heart diseases (I20-I25)	605	479	409	60.6	92.0	67.6	97.4	85.4
Pulmonary heart disease and diseases of pulmonary circulation (I26-I28)	24	22	5	12.2	98.9	20.8	99.5	22.7
Other forms of heart disease (I30-I52)	664	943	495	44.5	81.4	74.6	83.1	52.5
Cerebrovascular diseases (I60-I69)	1510	1546	1359	80	89.8	90.0	89.7	87.9
Diseases of arteries, arterioles and capillaries (I70-I79)	99	125	61	37.4	96.9	61.6	98.0	48.8
Diseases of veins, lymphatic vessels and lymph nodes (I80-I89)	18	22	4	11.1	99.0	22.2	99.5	18.2
Other and unspecified disorders of the circulatory system (I95-I99)	4	0	0	0	99.9	0.0	100.0	–
Total	3,320	3,320	2,393	–	72.1	–	–	–

UC: underlying cause; PD: primary diagnosis; CP: concordant pairs; TC: total concordance; PC: positive concordance; SE: sensitivity; SP: specificity; PPV: positive predictive value.
Source: Hospitalization System from the Unified Health System (*Sistemas de Informações sobre Mortalidade e de Internação Hospitalar do Sistema Único de Saúde* - SIH-SUS).

same databases[18] are applied to this study: the inclusion of health facilities in selecting the deaths from the SIM in 2006, which might no longer have an agreement with SUS; typographical errors in at least one of the databases for the information on the main variable for the link (name of the deceased), which were not considered as linked records; very common names with the remaining variables selected for the link being not enough consistent to be considered a matching link; and information error for "type of discharge," originally informed as being "death," when actually the discharge was unrelated to the death of the elderly individual. The related and non-related records did not show a significant difference in terms of demographic characteristics, regarding sex, age, and primary diagnosis.

Table 5. Number of deaths with an ill-defined underlying cause according to the primary diagnosis for the hospitalization of elderly individuals, Rio de Janeiro, 2006.

Primary Diagnosis – ICD 10 Chapter and Grouping	Signs and symptoms regarding the circulatory and respiratory systems ICD 10: R00-R09	General signs and symptoms ICD 10: R50-R69	Ill-defined and unknown causes ICD 10: R95-R99	Total	
				n	%
Chapter II – Neoplasms	–	1	2	3	4.8
Malignant neoplasms, female genital organs (C51-C58)	–	1	–	1	–
Malignant neoplasms of male genital organs (C60-C63)	–	–	1	1	–
Benign (D10-D36)	–	–	1	1	–
Chapter IX – Diseases of the circulatory system	3	1	7	11	17.7
Ischemic heart diseases (I20-I25)	1	–	1	2	–
Pulmonary heart disease and diseases of pulmonary circulation (I26-I28)	–	–	1	1	–
Other forms of heart disease (I30-I52)	–	1	4	5	–
Cerebrovascular diseases (I60-I69)	2	–	1	3	–
Remaining causes	4	22	22	48	77.4
Total	7	24	31	62	100.0

Source: Hospitalization System from the Unified Health System (Sistemas de Informações sobre Mortalidade e de Internação Hospitalar do Sistema Único de Saúde - SIH-SUS).

Based on the comparison of the SIM and SIH-SUS databases, the researchers observed a predominance of diseases of the cardiovascular system, neoplasms, and diseases of the respiratory system, among the deaths and hospitalizations, which are the most frequent causes of hospitalization and hospital deaths among the elderly[3,5,6].

The percentage of agreement between the underlying cause and the primary diagnosis for the groups of causes (ICD chapter) was considered good and, when the agreement by chance (Kappa coefficient) was eliminated, it was considered fair. The reliability between the primary diagnosis and the underlying cause varied between the groups of causes (ICD chapters) and, therefore, substituting the underlying ill-defined causes for the primary diagnosis should not be indiscriminate.

For all of the primary diagnosis, the agreement with the natural underlying causes of death is not expected to be elevated. The principles that guide the definition of the primary diagnosis and the underlying cause of death are different. The primary diagnosis of the AIH is the one that motivated the hospitalization, transcribed directly from the medical records, under the responsibility of the record services conducted at the service units. The underlying cause of natural death is defined as the disease or injury that initiated the chain of events leading to death[19], with the completion of the Death Certificate being under the responsibility of the doctor. Taking these definitions into consideration, it is possible to assume that the primary diagnosis is more reliable to define the natural terminal cause of death instead of the underlying cause, unlike the external causes[18].

In this study, the agreement between the primary diagnosis and the underlying cause varied according to the ICD chapter, being outstandingly higher for diseases of the circulatory system, when compared to the remaining groups of causes, although the value is only approximately 22.0%. The quality of the AIH data should also be carefully analyzed, as there are no explicit rules for their issuance and completion, nor standardized or regular training for the teams[13]. Problems regarding the quality of information in medical records and the codification of primary diagnosis also affect the quality of data in the SIH-SUS[24]. Another issue is the occurrence of comorbidities among the elderly, which partly justifies the difficulty in defining the underlying cause of death and can also justify the lack of agreement between the underlying cause and the primary reason for hospitalization.

As for the accuracy of the primary diagnosis, low sensitivity and high specificity levels were found, except for diseases of the eye and adnexa (which had a sensitivity and specificity of nearly 100.0%), whose frequency of deaths is insignificant.

The PPV can be considered a good criterion to recommend the replacement of the original ill-defined cause by the primary diagnosis. In this study, the PPV represents the total number of individuals with the same underlying cause, among the records containing a primary diagnosis that has been classified in one of the groups of causes, chapter or specific grouping of the ICD. If the cause of death of an elderly individual that was hospitalized in a SUS facility was ill-defined, it would be possible to use the primary diagnosis for the hospitalization in which the death occurred, and if this corresponded to one of the groups of causes with elevated PPV, it would be possible to replace the original underlying cause

of death. The PPV, in addition to being directly influenced by the frequency of deaths with the same underlying cause in the population, is also positively linked to the specificity[23]. Based on the results of this study, it is recommended that the state of Rio de Janeiro employ this substitution for the primary diagnosis of the cardiovascular disease and neoplasias.

A limitation of the study was to ensure the quality of the information concerning the primary diagnosis and the defined underlying cause, these problems can range from completing the respective forms to the codification and typing of the data. Additionally, the selection of the age group 60 years and over in the SIH-SUS database in 2006 excludes elderly individuals who, during the hospitalization, were aged 59 years, but whose death occurred at the age of 60 years, which represents a loss of information when linking the database to the SIM data.

The methodology that was developed and applied in this study to recover the information on the natural cause of death among the elderly has the advantage of effectiveness and reduction of costs when compared to the type of investigation on the deaths that would be recommended in situations of non-linkage or low PPV of the primary diagnosis. Monitoring of the mortality profile by the causes of death is necessary for the periodic update of the PPV.

However, it must be reinforced that the correct completion of the underlying cause in the medical certificate of cause of death and more precise information concerning the causes of death enable the construction of more reliable health indicators and more effective interventions, in addition to sparing qualification methodologies of the causes of death. Training on how to properly complete the Death Certificate should be included and/or reinforced in medical schools, and specific norms for filling out the primary and secondary diagnosis should be communicated to the professionals responsible for inputting this information into the SIH/SUS.

CONCLUSIONS

The primary diagnosis of the AIH is reliable and valid for the purpose of recovering the information on the underlying cause of natural death among the elderly whose deaths were initially assigned to ill-defined causes. The simple and cost-effective methodology for the databases linkage enables to obtain more qualified health information and indicators. However, the most effective way to obtain valid health information and indicators is to ensure the quality of information on the primary diagnosis and the underlying defined cause. Its main issues may range from completing the respective forms to the codification and typing of data.

ACKNOWLEDGMENTS

The authors would like to thank DATASUS for making the SIH/SUS database available and the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq) for the financial support under the process number 2006/8.

REFERENCES

1. Instituto Brasileiro de Geografia e Estatística. Censo Demográfico 2010. [cited 12 Nov 2013]. Disponível em: http://www.ibge.gov.br/home/estatistica/populacao/censo2010/caracteristicas_da_populacao/caracteristicas_da_populacao_tab_brasil_zip_ods.shtm
2. Chaimowicz F. A saúde dos idosos brasileiros às vésperas do século XXI: problemas, projeções e alternativas. Rev Saúde Pública 1997; 31(2): 184-200.
3. Góis AL, Veras RP. Informações sobre a morbidade hospitalar em idosos nas internações do Sistema Único de Saúde do Brasil. Ciênc Saúde Coletiva 2010; 15(6): 2859-69.
4. Jorge MH, Laurenti R, Lima-Costa, MF, Gotlieb SL, Chiavegatto Filho AD. A mortalidade em idosos no Brasil: a questão das causas mal definidas. Epidemiol Serv Saúde 2008; 17(4): 271-81.
5. Lima-Costa MF, Peixoto SV, Giatti, L. Tendências da mortalidade entre idosos brasileiros (1980 - 2000). Epidemiol Serv Saúde 2004; 13(4): 217-28.
6. Abreu DM, Sakurai E, Campos LN. A evolução da mortalidade por causas mal definidas na população idosa em quatro capitais brasileiras, 1996-2007. Rev Bras Estud Popul 2010; 27(1): 75-88.
7. Vasconcelos AM. Estatísticas de mortalidade por causas: uma avaliação da qualidade da informação. In: Anais do X Encontro Nacional de Estudos Populacionais. Caxambu: ABEP; 1996; 1: 151-66. [CD-ROM]
8. Martin GB, Cardoni Junior L, Bastos YG. Aspectos demográficos do processo de envelhecimento populacional em cidade do sul do Brasil. Epidemiol Serv Saúde 2005; 14(3): 151-8.
9. Kanso S, Romero DE, Leite IC, Moraes EN. Diferenciais geográficos, socioeconômicos e demográficos da qualidade da informação da causa básica de morte dos idosos no Brasil. Cad Saúde Pública 2011; 27(7): 1323-39.
10. Cascão AM, Costa, AJ, Kale PL. Qualidade da informação sobre mortalidade numa coorte de diabéticos - Estado do Rio de Janeiro, 2000 a 2003. Rev Bras Epidemiol 2012; 15(1): 134-42.
11. Barbucia DM, Rodrigues Júnior AL. Completude da informação nas Declarações de Nascidos Vivos e nas Declarações de Óbito, neonatal precoce e fetal, da região de Ribeirão Preto, São Paulo, Brasil, 2000-2007. Cad Saúde Pública 2011; 27(6): 1192-200.
12. Teixeira CL, Block KV, Pinheiro RS, Klein CH. Concordâncias entre causas de óbito nas declarações de óbito e diagnósticos nas autorizações de internação hospitalar em regiões do Estado do Rio de Janeiro, em 1998, usando relacionamento probabilístico de dados. Cad Saúde Colet (Rio J) 2006; 14(2): 405-10.
13. Melo EC, Travassos C, Carvalho MS. Qualidade dos dados sobre óbitos por infarto agudo do miocárdio, Rio de Janeiro. Rev Saúde Pública 2004; 38(3): 385-91.
14. Laurenti R, Mello Jorge MH, Gotlieb SL. A confiabilidade dos dados de mortalidade por doenças crônicas não transmissíveis. Ciênc Saúde Coletiva 2004; 9(4): 909-20.
15. Nasri F. O envelhecimento populacional no Brasil. Einstein 2008; 6(Suppl 1): S4-6.
16. Nogueira SL, Geraldo JM, Machado JC, Ribeiro RC. Distribuição espacial e crescimento da população idosa nas capitais brasileiras de 1980 a 2006: um estudo ecológico. Rev Bras Estud Popul 2008; 25(1): 195-8.
17. Camargo Junior KR, Coeli CM. RecLink: aplicativo para relacionamento de banco de dados implementando o método probabilístico record linkage. Cad Saúde Pública 2000; 16(2): 439-47.
18. Jorge MH, Cascão AM, Reis AC, Laurenti R. Em busca de melhores informações sobre a causa básica do óbito por meio de *linkage*: um recorte sobre causas externas em idosos – Estado do Rio de Janeiro, Brasil, 2006. Epidemiol Serv Saúde 2012; 21(3): 407-18.
19. Organização Mundial da Saúde. Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde. 10ª revisão. v I, II e III. São Paulo: Editora da Universidade de São Paulo; 1995.
20. Gordis L. Assessing the validity and reliability of diagnostic and screening tests. In: Gordis L (editor). Epidemiology. 4th edition. Philadelphia: W.B. Saunders Elsevier; 2009. p. 85-108.
21. Mello Jorge MH, Laurenti R, Gotlieb SL. Avaliação dos Sistemas de Informações em Saúde no Brasil. Cad Saúde Pública 2010; 18(1): 7-18.
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33(1): 159-74.
23. Szklo M, Nieto F. Epidemiology: beyond the basics. 2nd ed. Boston: Jones and Barllett Publishers; 2007.
24. Rehem TC, Oliveira MR, Ciosak SI, Egry EY. Registro das internações por condições sensíveis à atenção primária: validação do sistema de informação hospitalar. Rev Latino-Am Enferm 2013; 21(5): 1159-64.

Received on: 08/27/2015
Final version presented on: 02/21/2016
Accepted on: 05/16/2016