Hospital-Based Cancer Registry: Methodological Proposal for Correction of Lung Cancer Staging

Registros Hospitalares de Câncer: Proposta Metodológica para Correção do Estadiamento de Câncer de Pulmão

Registros Hospitalarios de Cáncer: Propuesta Metodológica para la Corrección de la Estadificación del Cáncer de Pulmón

Tayná Sequeira Valerio[1]; Mario Jorge Sobreira da Silva[2]; Mônica Rodrigues Campos[3]; Debora Castanheira Pires[4]; Carlos Henrique Dantas Cardoso[5]; Lara Vinhal Faria[6]; Silvio Rodrigues Campos[7]; Luisa Arueira Chaves[8]; Aline Pinto Marques[9]; Jéssica Muzy[10]; Isabel Cristina Martins Emmerick[11]

ABSTRACT

Introduction: A major challenge to utilize the registries and secondary databases is the quality of the data and the percentage of losses in strategic and necessary variables for better effectiveness of the database. **Objective:** To propose a correction method for the cancer staging variable of the Hospital-Based Cancer Registry (HBCR), to improve its completeness and quality. **Method:** HBCR-based descriptive analysis covering Brazil's Federation Units from January 2013 to December 2019. Due to its high mortality in Brazil and worldwide, lung cancer was selected as case for database correction. The analyzes were performed with the software SAS Studio for statistical analyzes and the data were organized in Excel®. **Results:** The total number of cases registered at the HBCR was 86,026, and 32% of the variable of interest, staging, were missed. At the end of the correction process, the missed data reached 9.8%, corresponding to a recovery of 22.2%. **Conclusion:** The proposed methodology is an advance for the correction of the HBCR database on the treatment of lung cancer, allowing a more extensive use, with better representativeness of different country regions, and potential utilization in other topographies.
Key words: hospital records; electronic health records; database management systems; lung neoplasms; neoplasm staging.

RESUMO

Introdução: Um grande desafio para a utilização de registros e bases de dados secundárias é a qualidade do registro e o percentual de perdas em variáveis estratégicas e necessárias à plena utilização do banco. **Objetivo**: Propor um método de correção para a variável de estadiamento no âmbito dos Registros Hospitalares de Câncer (RHC), a fim de aprimorar sua completude e qualidade. **Método:** Estudo descritivo, abrangendo as Unidades da Federação, utilizando-se como fonte de informação o RHC, de janeiro de 2013 a dezembro de 2019. O câncer de pulmão foi escolhido como caso para a correção do banco, em razão da sua alta taxa de mortalidade no Brasil e no mundo. As análises foram realizadas com o *software* de análises estatísticas *SAS Studio* e a base de dados organizada em *Excel®*. **Resultados**: O total de casos registrados no RHC foi de 86.026, e a variável de interesse, o estadiamento, teve um total de 32,0% de perda. Ao final de todas as etapas de correção, a perda foi de 9,8%, correspondendo a 22,2% de recuperação. **Conclusão:** A metodologia proposta representa um avanço na correção do banco do RHC, possibilitando a utilização dos dados de base secundária, com melhor representatividade das diferentes Regiões do país, sobre o tratamento de câncer de pulmão, com possibilidade de expansão de seu uso para outras topografias.
Palavras-chave: registros hospitalares; registros eletrônicos de saúde; sistemas de gerenciamento de base de dados; neoplasias pulmonares; estadiamento de neoplasias.

RESUMEN

Introducción: Un gran desafío para el uso de registros y bases de datos secundarias es la calidad del registro en sí, el porcentaje de pérdidas en variables estratégicas y necesarias para el pleno uso de la base de datos. **Objetivo:** Proponer un método de corrección de la variable estadificación en el ámbito de los Registros Hospitalarios de Cáncer (RHC), con el fin de mejorar su exhaustividad y calidad. **Método:** Análisis descriptivo, abarcando las Unidades de la Federación. Se utilizó el RHC como fuente de información, de enero de 2013 a diciembre de 2019. El cáncer de pulmón fue elegido como caso para la corrección de la base de datos, debido a su alta tasa de mortalidad en el Brasil y en el mundo. Los análisis se realizaron con el *software* de análisis estadístico *SAS Studio* y los datos se organizaron en *Excel®*. **Resultados:** El total de casos registrados en el RHC fue de 86 026, y la variable de interés, la estadificación, tuvo una pérdida total del 32,0% Al final de todas las etapas esta fue de 9,8%, es decir el 22,2% de recuperación. **Conclusión:** La metodología propuesta representa un avance en la corrección del RHC, permitiendo una mejor utilización de la base de datos, con una mejor representatividad de las diferentes regiones del país, sobre el tratamiento del cáncer de pulmón, con la posibilidad de expandir su uso a otras topografías.
Palabras clave: registros de hospitales; registros electrónicos de salud; sistemas de administración de bases de datos; neoplasias pulmonares; estadificación de neoplasias.

[1,2]Instituto Nacional de Câncer (INCA), Coordenação de Ensino (Coens). Rio de Janeiro (RJ), Brazil. E-mails: tayna.sequeira@gmail.com; mario.silva@inca.gov.br. Orcid iD: https://orcid.org/0000-0002-9757-9076; Orcid iD: https://orcid.org/0000-0002-0477-8595
[3]Fundação Oswaldo Cruz (Fiocruz), Escola Nacional de Saúde Pública (Ensp), Departamento de Ciências Sociais. Rio de Janeiro (RJ), Brazil. E-mail: monicarodriguescampos@gmail.com. Orcid iD: https://orcid.org/0000-0002-7443-5977
[4]Fiocruz, Instituto Nacional de Infectologia Evandro Chagas, Laboratório de Pesquisa Clínica em DST e Aids. Rio de Janeiro (RJ), Brazil. E-mail: deboracastanheira1@gmail.com. Orcid iD: https://orcid.org/0000-0001-6995-1259
[5-7]Fiocruz/Ensp, Departamento de Administração e Planejamento em Saúde. Rio de Janeiro (RJ), Brazil. E-mails: henrique.dantas2030@gmail.com; laravinhalf@gmail.com; silvio.rodrigues.campos@gmail.com. Orcid iD: https://orcid.org/0000-0002-6175-2666; Orcid iD: https://orcid.org/0000-0003-3158-702X; Orcid iD: https://orcid.org/0000-0003-2401-1911
[8]Universidade Federal do Rio de Janeiro (UFRJ), Instituto de Ciências Farmacêuticas. Macaé (RJ), Brazil. E-mail: luisa.arueira@gmail.com. Orcid iD: https://orcid.org/0000-0002-9597-3651
[9,10]Fiocruz, Instituto de Comunicação Científica e Tecnológica em Saúde, Laboratório de Informações em Saúde. Rio de Janeiro (RJ), Brazil. E-mails: alinepmarques.fiocruz@gmail.com; jehmuzy@gmail.com. Orcid iD: https://orcid.org/0000-0002-9072-5333; Orcid iD: https://orcid.org/0000-0003-2526-2317
[11]*UMass Chan Medical School, Department of Surgery, Division of Thoracic Surgery*. Worcester, Massachusetts, United States. E-mail: isabel.emmerick@umassmemorial.org. Orcid iD: https://orcid.org/0000-0002-0383-2465
Corresponding author: Lara Vinhal Faria. Fiocruz, Ensp, Departamento de Administração e Planejamento em Saúde. Rua Leopoldo Bulhões, 1480 – Manguinhos. Rio de Janeiro (RJ), Brazil. CEP 21041-210. E-mail: laravinhalf@gmail.com

Valerio TS, Silva MJS, Campos MR, Pires DC, Cardoso CHD, Faria LV, Campos SR, Chaves LA, Marques AP, Muzy J, Emmerick ICM

INTRODUCTION

Health information systems are vital for health management[1]. Brazil relies on a set of information systems covering mortality, live births, compulsory notifiable diseases[2], and registries for specific diseases as cancer[3].

The cancer monitoring and surveillance systems differ from others since their finality is to estimate the disease burden and incidence-related risk factors. Accurate data about tumor anatomic and morphologic locations are necessary to classify the extent of cancer and its main outcomes that should encompass information about cure, recurrence or relapse, survival and mortality[4].

Hospital-based cancer registries (HBCR) are important tools to improve the quality of care provided to the individual with cancer[5]. Its implementation is easy since it collects, stores, processes, and analyzes systematically and continuously information about patients consulted in a hospital with confirmed diagnoses. To consolidate most of the HBCR, Brazil utilizes the SisRHC, a computer-based data system developed and offered by the National Cancer Institute (INCA)[6].

The databases consolidated according to the year of the first consultation reported by the hospital are sent to form the national base of HBCR under INCA's supervision through *IntegradorRHC*[7]. Keeping the continuous operation of HBCR and sending data to *IntegradorRHC* routinely are mandatory actions for the accredited high-complexity cancer hospitals of the National Health System (SUS) and optional for non-accredited hospitals[8].

The HBCR has 46 variables divided into sociodemographic, epidemiologic, and clinical data[9]. The great challenge to utilizing secondary registries and databases is the registry's quality due to the percentage of losses of strategic variables and the necessity to use these bases fully. HBCR is no exception, as completeness and inconsistencies issues have been identified in the system[10-12].

The development of strategies to minimize the impact of these problems and improve the quality of the information analyzed is challenging. Lima et al.[13] had to exclude 24.06% of the cases while analyzing the spatial distribution of the diagnosis of advanced stage and mortality by lung cancer and association with the offer of services and socioeconomic indicators of the country due to missed information about patients' staging, furthermore for being concentrated in specific regions of Brazil.

This article aims to propose a method to correct the staging variable in HBCR to improve its completeness and quality.

METHOD

HBCR[7] was the source of information from January 2013 to December 2019, with descriptive analyses covering all the Brazilian Federative Units (FU) and Regions.

Lung cancer was selected as a case to correct the database due to its high mortality rate in Brazil and worldwide[14-16]. Operationally, the cases were selected according to the International Classification of Diseases and Related Health Problems, version 10 (ICD-10 = "C34", based on the variable "LOCTUDET"[6].

The statistic software SAS Studio was utilized for the analyses and the database was organized in Excel®.

The HBCR lung cancer staging variable is categorized into: IA, IIA, IB, IIB, IIIA, IIIB and IV. Given the impossibility of retrieving information about losses at this level of detail, the staging was recategorized into:

a. "Initial" (IA, IIA, IB, IIB), the disease is located in the lung and has a better prognosis, consequently.

b. "Advanced" (IIIA, IIIB and IV), compromise of lymph nodes and/or metastases to other organs and lower global survival.

Considering that the information about the first treatment is fully available in HBCR, the correction proposal is based on the assumption that this registry is defined according to the staging at the diagnosis. According to lung cancer management protocols, most of the patients at initial staging will undergo surgical treatment, and those at advanced staging, systemic treatment[9,17]. In addition, according to the HBCR consistency manual, surgical procedures only for diagnosis should not be filled in this variable because it refers specifically to procedures performed as first treatment.

It has been decided to use the patient's residence and not where the treatment was performed. The total losses for the FU of residence was 532 (0.61% of the total registries), for which no correction or redistribution was done, and these were not included in the correction tables.

The proposal for correction of the losses of staging ('ESTADIAM' at HBCR) considered the following phases sequentially:

Phase 1: the variable TNM was utilized – defined as "Assigning a letter or number to describe the tumor according to the TNM System" – Classification of Malignant Tumors[18]. Based in the characteristics of the primary tumor (T), T1 and T2 are classified as the initial stage and T3 and T4, advanced stage.

Phase 2: after Phase 1, the remaining staging losses were corrected with the variable "RZNTR", the main reason for not performing the treatment. If the main reason was advanced disease without the possibility of treatment or other associated diseases, the registry was considered advanced stage (stage 3 or 4).

Phase 3: after Phase 2, the remaining losses were corrected according to the variable "PRITRATH", the first treatment received at the hospital. If chemotherapy has been registered at this variable, the losses were considered

advanced stage, and if surgery was registered, the losses were determined as initial stage.

Phase 4: the remaining losses after Phase 3 were corrected utilizing the staging variable "ESTDFIMT" at the end of the first treatment. If death was registered, consider advanced staging. This phase was not performed for the State of São Paulo because the State does not fill this variable at the HBCR.

Phase 5: the remaining losses after Phase 4 were corrected utilizing the variable "RZNTR", the main reason for not performing the antineoplastic treatment. If death was registered for not treating, the registry of loss was considered as advanced staging.

In order to evaluate the consistency of the proposal of correction, an interrater reliability analysis was performed, for which the cases where the staging variable was filled were selected. A variable titled "new staging recovered" was created after the proposal phases had been completed and compared with the original variable through the calculation of the percent of agreement and adjusted Kappa by the prevalence with its respective confidence interval.

The review by the Institutional Review Board (IRB) was waived because only deidentified secondary data were utilized according to Directive 466[19], December 12, 2012, of the National Health Council.

RESULTS

The percent of losses of the variables utilized in the proposal of correction of the staging variable for Brazil, Regions, and FU is portrayed in Table 1. The total number of cases of lung cancer in Brazil registered at the HBCR[7] was 86,026 and a total loss of 32.0% for the staging variable. The highest percent of loss among macroregions was found in the North Region, where 53.3% of the cases had missing data for the variable and the lowest was in the Southeast Region, with a loss of 22.2% of the cases. Considering the FU, the State of Amazonas presented the highest percent of loss (71.7%,) and the State of Piauí, the lowest (5.0%).

Given the variables utilized for the correction, including TNM, a variable that determines the staging, the losses were high and followed the same profile as the variable "ESTADIAM". On the contrary, for the variables related to the treatment provided, the percentage of losses was much lower, being higher than 10% only in the North and Northeast Regions and approximately 2% in the Southeast Region. For the States of Paraíba, Rio Grande do Norte and Minas Gerais, the percent was lower than 1% (Table 1).

The high national percent of loss for the variable "Stage of the disease at the end of the first treatment at the hospital" (41.9%) occurs because the State of São Paulo, which accounts for nearly 25% of the cases in Brazil, does not collect this information.

The percent of total recovery is the sum of the percent of recovery at each phase of the correction (Table 2). The variable "first treatment" (phase 3) contributed the most to the recovery of the loss of staging. Nationwide, the total loss was 32% and, at the end of all phases, 9.8%, a percent of recovery of 22%, for which phase 3 accounted for 12.7%.

The proposal presented for Brazil resulted in a loss reduction of 69.4%. The Region with the most significant decrease was Southeast (71.3%), and the lowest was Midwest (60.8%). The percent of recovery of the staging variable was higher than 60% for all the States, except Rondônia (44.5%) and Mato Grosso (55%). Overall, according to the FU, the final percent of loss continued above 15% for the States of Acre, Rondônia, Alagoas, Pará, Paraíba, Sergipe, Goiás, Mato Grosso, Santa Catarina and Federal District (Table 3).

Given the national-based distribution of initial and advanced staging, 9.2% of the cases were diagnosed as initial and 58.8% as advanced staging, with 32% of losses before the correction. After the correction, these results were revised to 15.8%, 74.4%, and 9.8%, respectively. The percent of initial staging, only for the valid cases, initially at 13.5%, raised to 17.6%, and the advanced staging consequently declined from 86.5% to 82.4% with high variation among FU.

The reliability of the proposal (Table 4) reached 81.3% of the agreement for Brazil and Kappa-adjusted of 0.63%. The percent of agreement varied from 79.1% in the North to 84.7% in the Midwest Regions; the FU of Maranhão presented the lowest percent of agreement of 59.7% and the highest in Roraima with perfect agreement, followed by Sergipe with 95% of agreement. The median of the percent of agreement was 81.9%, and the interquartile range was 79.4%-87.9.

DISCUSSION

The present proposal of correction of the staging variable reached a high percent of recovery (69%). The treatment variables were collected as trustworthily as possible and allowed improved accuracy when the disease's initial staging (stage I or II) or advanced staging (stages II or IV) was determined.

Despite the limitations of the variables utilized, the first treatment, which is the variable that contributed most to the correction, follows strictly clinical criteria. Additionally, the proposal is conservative because it corrects the staging only if the systemic treatment

Table 1. Number of lung cancer cases, number and percent of losses of the staging variable, TNM, reason-to-not-treat, first treatment received at the hospital and stage of the disease at the end of the first hospital treatment. Brazil, Regions and FU of residence, 2013 to 2019

FU	Cases	Staging		TNM		Reason-to-not-treat		First treatment received at the hospital		Stage of the disease at the end of the first treatment at the hospital	
	N	N	%	N	%	N	%	N	%	N	%
Brazil	**86,026**	**27,714**	**32.0**	**46,998**	**54.3**	**5,396**	**6.2**	**4,485**	**5.2**	**36,265**	**41.9**
North	**3,407**	**1,819**	**53.4**	**2,212**	**64.9**	**262**	**7.7**	**349**	**10.2**	**472**	**13.9**
AC	337	177	52.5	179	53.1	29	8.6	11	3.3	94	27.9
AM	544	390	71.7	443	81.4	28	5.1	6	1.1	42	7.7
AP	81	42	51.8	56	69.1	3	3.7	7	8.6	14	17.3
PA	1,265	698	55.2	919	72.6	175	13.8	24	1.9	124	9.8
RO	592	357	60.3	393	66.4	23	3.9	172	29.0	119	20.1
RR	79	32	40.5	49	62.0	2	2.5	2	2.5	21	26.6
TO	509	123	24.2	173	34.0	2	0.4	127	25.0	58	11.4
Northeast	**17,241**	**7,286**	**42.3**	**12,149**	**70.5**	**2,169**	**12.6**	**1,891**	**11.0**	**4,446**	**25.8**
AL	1,212	782	64.5	902	74.4	69	5.7	145	12.0	254	21.0
BA	3,049	1,743	57.2	2,212	72.5	197	6.5	204	6.7	559	18.3
CE	5,236	1,782	34.0	4,225	80.7	1,323	25.3	1,137	21.7	1,512	1.7
MA	1,471	408	27.7	960	65.3	136	9.2	82	5.6	291	19.8
PB	1,245	710	57.0	921	74.0	14	1.1	153	0.2	889	71.4
PE	2,095	855	40.8	1,509	72.0	167	8.0	57	2.7	325	15.5
PI	1,029	51	5.0	169	16.4	134	13.0	31	3.0	282	27.4
RN	1,277	436	34.1	704	55.1	52	4.1	9	0.7	65	5.1
SE	627	519	82.8	547	87.2	77	12.3	73	11.6	269	42.9
Midwest	**3,388**	**1,286**	**38.0**	**1,865**	**55.0**	**247**	**7.3**	**257**	**7.6**	**1,245**	**36.7**
DF	746	344	46.1	480	64.3	59	7.9	76	10.2	144	19.3
GO	814	343	42.1	445	54.7	67	8.2	97	11.9	514	63.1
MS	979	290	29.6	450	46.0	60	6.1	48	4.9	273	27.9
MT	849	309	36.4	490	57.7	61	7.2	36	4.2	314	37.0
Southeast	**38,577**	**8,557**	**22.2**	**16,356**	**42.4**	**801**	**2.1**	**842**	**2.2**	**23,905**	**62.0**
ES	1,851	748	40.4	1,501	81.1	180	9.7	118	6.4	340	18.4
MG	11,009	3,050	27.7	5,550	50.4	267	2.4	90	0.8	2,819	25.6
RJ	5,575	2,273	40.8	3,583	64.3	333	6.0	212	3.8	658	11.8
SP	20,142	2,486	12.3	5,722	28.4	21	0.1	422	2.1	20,088	99.7
South	**23,413**	**8,548**	**36.5**	**14,133**	**60.4**	**1,816**	**7.8**	**1,058**	**4.5**	**5,927**	**25.3**
PR	5,704	1,453	25.5	2,904	50.9	191	3.5	75	1.3	526	0.60
RS	12,580	4,404	35.0	7,374	58.6	1,439	11.4	743	5.9	4,090	32.5
SC	5,129	2,691	52.5	3,855	75.2	186	3.6	240	4.7	1,311	25.6

Caption: FU = Federative Unit.

"chemotherapy" is applied. This decision is based on the treatment protocol that does not utilize systemic treatment for initial staging in addition to not being recommended surgical treatment for advanced disease[9,16,17].

The proposal herein overestimates the cases diagnosed as initial because it assumes that surgical procedures as the first treatment would be utilized only in cases in which it is specifically indicated. For Brazil, nevertheless, it was noticed that 10% of individuals diagnosed with late staging in 2019 were submitted to surgery as the first treatment. Still, regardless of this finding, the authors opted to keep the correction as such, since it follows the

Table 2. Initial and final percent of losses and total percent of recovery and per phase of correction of the staging variable. Brazil, Regions, and FU of residence, 2013 to 2019

	Staging				Percent of recovery per phase			
	Initial loss	Final loss	Recovery	TNM	Main reason (advanced disease)	First treatment	End of the first treatment	Main reason (death)
UF	%	%	%	%	%	%	%	%
Brazil	**32.0**	**9.8**	**22.2**	**1.9**	**2.9**	**12.7**	**1.5**	**3.2**
North	**53.4**	**17.4**	**36.0**	**2.9**	**7.2**	**16.0**	**3.9**	**6.0**
AC	52.5	20.2	32.3	0.3	5.9	13.4	1.5	11.2
AM	71.2	11.4	59.8	0.2	23.0	29.3	6.2	1.1
AP	51.8	13.6	38.2	1.2	2.5	33.3	1.2	0.0
PA	55.2	17.1	38.1	5.6	6.8	11.5	1.7	12.5
RO	60.3	33.4	26.9	2.4	0.8	18.1	5.6	0.0
RR	40.5	15.2	25.3	5.1	6.3	12.6	0.0	1.3
TO	24.2	4.5	19.7	0.4	0.6	10.4	7.7	0.6
Northeast	**42.3**	**13.0**	**29.3**	**2.1**	**4.1**	**17.8**	**2.6**	**2.7**
AL	64.5	16.9	47.6	4.0	3.7	35.1	2.5	2.3
BA	57.2	13.5	43.7	3.5	3.1	26.8	5.2	5.1
CE	34.0	13.5	20.5	0.2	3.9	13.2	1.6	1.6
MA	27.7	8.2	19.5	0.9	6.6	8.4	2.3	1.3
PB	57.0	20.6	36.4	3.7	0.1	28.5	2.2	1.9
PE	40.8	11.7	29.1	2.0	7.9	14.1	1.3	3.8
PI	5.0	1.5	3.5	0.8	0.1	1.8	0.8	0.0
RN	34.1	11.8	22.3	1.3	7.8	8.7	0.9	3.6
SE	82.8	21.7	61.1	9.3	0.5	38.1	9.9	3.3
Midwest	**38.0**	**14.9**	**23.1**	**2.2**	**4.2**	**11.6**	**2.0**	**3.1**
DF	46.1	16.9	29.2	2.8	6.4	13.2	4.4	2.4
GO	42.1	16.5	25.6	1.8	2.5	14.5	2.2	4.6
MS	29.6	10.7	18.9	1.6	6.4	8.2	0.8	1.9
MT	36.4	16.4	20.0	2.8	1.3	11.3	1.0	3.6
Southeast	**22.2**	**6.4**	**15.8**	**1.2**	**2.2**	**8.1**	**0.7**	**3.7**
ES	40.4	8.2	32.2	1.1	1.2	23.4	1.6	4.9
MG	27.7	7.4	20.3	2.2	2.5	10.8	0.9	3.9
RJ	40.8	11.5	29.3	3.5	6.9	12.0	2.6	4.3
SP	12.3	4.2	8.1	0.0	0.7	4.1	0.0	3.3
South	**36.5**	**11.1**	**25.4**	**2.9**	**2.4**	**16.2**	**1.6**	**2.4**
PR	25.5	7.1	18.4	1.2	3.5	8.7	0.6	4.4
RS	35.0	11.3	23.7	3.8	2.1	14.9	1.1	1.8
SC	52.5	15.3	37.2	2.5	2.0	27.7	3.6	1.4

Caption: FU = Federative Unit.

protocol of lung cancer treatment. These results show that data collection and compliance with the protocol should undergo a deep analysis.

The consistency of the results of the correction proposal is closely related to the quality of the information found in the databases. The authors deemed as satisfactory the overall result of the analysis of agreement. For some FU, the proposal of correction reached an agreement below the expected, requiring improvement of the quality of the registry and investigation of other possibilities of imputation of the losses. It is also necessary that future studies address the validation of the proposal when compared with multiple proposals of correction and imputation of losses. The authors believe that the

Table 3. Number of lung cancer cases, number and percent of losses of the staging variable before and after the correction, per classification of staging. Brazil, Regions and FU of residence, 2013 to 2019

FU	Total	Loss		Initial staging		Advanced staging		Loss		Initial staging		Advanced staging		% correction
		Before the correction						After the correction						
	N	N	%	N	%	N	%	N	%	N	%	N	%	
Brazil	86,026	27,496	32.0	7,914	9.2	50,616	58.8	8,404	9.8	13,630	15.8	63,992	74.4	69.4
North	3,407	1,819	53.4	185	5.4	1,403	41.2	591	17.3	388	11.4	2,428	71.3	67.5
AC	337	177	52.5	17	5.0	143	42.4	68	20.2	19	5.6	250	74.2	61.6
AM	544	390	71.7	29	5.3	125	23.0	62	11.4	83	15.3	399	73.4	84.1
AP	81	42	51.9	5	6.2	34	42.0	11	13.6	9	11.1	61	75.3	73.8
PA	1,265	698	55.2	50	4.0	517	40.9	217	17.2	112	8.9	936	74.0	68.9
RO	592	357	60.3	49	8.3	186	31.4	198	33.5	83	14.0	311	52.5	44.5
RR	79	32	40.5	3	3.8	44	55.7	12	15.2	5	6.3	62	78.5	62.5
TO	509	123	24.2	32	6.3	354	69.6	23	4.5	77	15.1	409	80.4	81.3
Northeast	17,241	7,286	42.3	928	5.4	9,027	52.4	2,246	13.0	2,312	13.4	12,683	73.6	69.2
AL	1,212	782	64.5	50	4.1	380	31.4	205	16.9	261	21.5	746	61.6	73.8
BA	3,049	1,743	57.2	150	4.9	1,156	37.9	412	13.5	606	19.9	2,031	66.6	76.4
CE	5,236	1,782	34.0	254	4.9	3,200	61.1	706	13.5	381	7.3	4,149	79.2	60.4
MA	1,471	408	27.7	106	7.2	957	65.1	120	8.2	164	11.2	1,187	80.7	70.6
PB	1,245	710	57.0	76	6.1	459	36.9	256	20.6	309	24.8	680	54.6	63.9
PE	2,095	855	40.8	117	5.6	1,123	53.6	245	11.7	209	10.0	1,641	78.3	71.3
PI	1,029	51	5.0	100	9.7	878	85.3	15	1.5	111	10.8	903	87.8	70.5
RN	1,277	436	34.1	66	5.2	775	60.7	151	11.8	137	10.7	989	77.5	65.4
SE	627	519	82.8	9	1.4	99	15.8	136	21.7	134	21.4	357	56.9	73.8
Midwest	3,388	1,286	38.0	268	7.9	1,834	54.1	504	14.9	419	12.4	2,465	72.8	60.8
DF	746	344	46.1	73	9.8	329	44.1	126	16.9	123	16.5	497	66.6	63.4
GO	814	343	42.1	64	7.9	407	50.0	134	16.5	97	11.9	583	71.6	60.9
MS	979	290	29.6	79	8.1	610	62.3	105	10.7	109	11.1	765	78.1	63.8
MT	849	309	36.4	52	6.1	488	57.5	139	16.4	90	10.6	620	73.0	55.0
Southeast	38,577	8,557	22.2	4,598	11.9	25,422	65.9	2,456	6.4	6,277	16.3	29,844	77.4	71.3
ES	1,851	748	40.4	89	4.8	1,014	54.8	151	8.2	354	19.1	1,346	72.7	79.8
MG	11,009	3,050	27.7	922	8.4	7,037	63.9	818	7.4	1,653	15.0	8,538	77.6	73.2
RJ	5,575	2,273	40.8	283	5.1	3,019	54.2	639	11.5	594	10.7	4,342	77.9	71.9
SP	20,142	2.486	12.3	3,304	16.4	14,352	71.3	848	4.2	3,676	18.3	15,618	77.5	65.9
South	23,413	8,548	36.5	1,935	8.3	12,930	55.2	2,607	11.1	4.234	18.1	16,572	70.8	69.5
PR	5,704	1,453	25.5	474	8.3	3,777	66.2	403	7.1	708	12.4	4,593	80.5	72.3
RS	12,580	4,404	35.0	1,150	9.1	7,026	55.9	1,420	11.3	2,295	18.2	8,865	70.5	67.8
SC	5,129	2,691	52.5	311	6.1	2,127	41.5	784	15.3	1,231	24.0	3,114	60.7	70.9

Caption: FU = Federative Unit.

correction of the staging following the phases mirrors the reality better than the simple imputation of data given the proportional distribution of the valid cases of the original variable (staging).

The correction is an advance in utilizing information on HBCR compared to other studies[13,20], which decided to exclude the losses while analyzing lung cancer. These studies have an important selection bias since the quality of the information registry varies significantly across the FU. For instance, 70% of the cases of lung cancer in the State of Amazonas would be eliminated if staging losses were excluded.

Tabela 4. Number of lung cancer cases, number and percent of losses of the staging variable before the correction, percent of agreement of the variable new staging recovered, Kappa-adjusted by the prevalence and confidence interval. Brazil, Regions and FU of residence, 2013 to 2019

FU	Total	Staging (missing)		Agreement	Kappa	CI (95%)	
	N	N	%				
Total	**86,026**	**27,496**	**32.0**	**81.3**	**0.63**	**0.61**	**0.63**
North	**3,407**	**1,819**	**53.4**	**79.1**	**0.58**	**0.53**	**0.63**
AC	337	177	52.5	93.4	0.87	0.77	0.97
AM	544	390	71.7	89.3	0.79	0.67	0.91
AP	81	42	51.8	75.0	0.50	0.12	0.88
PA	1,265	698	55.2	85.0	0.70	0.63	0.77
RO	592	357	60.3	78.3	0.57	0.41	0.72
RR	79	32	40.5	100.0	1.00	1.00	1.00
TO	509	123	24.2	62.1	0.24	0.13	0.35
Northeast	**17,241**	**7,286**	**42.3**	**80.6**	**0.61**	**0.59**	**0.63**
AL	1,212	782	64.5	79.6	0.59	0.51	0.68
BA	3,049	1,743	57.2	74.6	0.49	0.44	0.55
CE	5,236	1,782	34.0	88.5	0.77	0.74	0.80
MA	1,471	408	27.7	59.7	0.16	0.12	0.27
PB	1,245	710	57.0	66.2	0.32	0.23	0.42
PE	2,095	855	40.8	87.9	0.76	0.71	0.80
PI	1,029	51	5.0	87.6	0.75	0.70	0.80
RN	1,277	436	34.1	82.2	0.64	0.58	0.71
SE	627	519	82.8	95.3	0.91	0.80	1.00
Midwest	**3,388**	**1,286**	**38.0**	**84.7**	**0.69**	**0.66**	**0.74**
DF	746	344	46.1	87.3	0.75	0.66	0.83
GO	814	343	42.1	87.8	0.76	0.67	0.84
MS	979	290	29.6	79.2	0.58	0.51	0.66
MT	849	309	36.4	87.6	0.75	0.69	0.82
Southeast	**38,577**	**8,557**	**22.2**	**81.9**	**0.64**	**0.63**	**0.65**
ES	1,851	748	40.4	90.7	0.81	0.77	0.95
MG	11,009	3,050	27.7	80.6	0.61	0.59	0.63
RJ	5,575	2,273	40.8	91.6	0.83	0.81	0.85
SP	20,142	2,486	12.3	79.7	0.59	0.58	0.61
South	**23,413**	**8,548**	**36.5**	**80.8**	**0.62**	**0.60**	**0.63**
PR	5,704	1,453	25.5	79.5	0.59	0.56	0.62
RS	12,580	4,404	35.0	80.9	0.62	0.57	0.64
SC	5,129	2,691	52.5	82.8	0.66	0.62	0.69

Captions: FU = Federative Unit; CI = Confidence interval.

This methodology of loss recovery can be replicated and matched to other topographies[21], expanding the utilization of HBCR in future studies.

The quality of the staging registry and the variables utilized in the correction varied widely per State and are similar to those presented by other information systems[2,22].

A perfect proposal of correction or imputation does not exist. It is, therefore, vital to invest in strengthening the country's information system and the health professional's correct filling of the data – these are essential actions to construct reliable indicators of health status. Lung cancer staging information is critical to planning actions and health services, especially in designing care networks to reinforce reference and counter-reference systems to ensure equitable access by the population.

CONCLUSION

The methodology is an advance in correcting the HBCR database since it produces reliable results and robust analyses that effectively mirror the differences in treating lung cancer by each country Region, possibly expanding its use to other topographies.

CONTRIBUTIONS

All the authors contributed substantially to the study design, analysis, interpretation of the data, wording and/or critical review. They approved the final version to be published.

DECLARATION OF CONFLICT OF INTERESTS

There is no conflict of interest to declare.

FUNDING SOURCES

REFERENCES

1. AbouZahr C, Boerma T. Health information systems: the foundations of public health. Bull World Health Organ [Internet]. 2005 [cited 2023 Feb 18];83(8):578-83. Available from: https://apps.who.int/iris/handle/10665/269465

2. Coelho Neto GC, Chioro A. Afinal, quantos Sistemas de Informação em Saúde de base nacional existem no Brasil? Cad Saúde Pública. 2021;37(7):e00182119. doi: https://doi.org/10.1590/0102-311X00182119

3. Piñeros M, Abriata MG, Mery L, et al. Cancer registration for cancer control in Latin America: a status and progress report. Rev Panam Salud Publica. 2017;41:e2. doi: https://doi.org/10.26633/RPSP.2017.2

4. Piñeros M, Znaor A, Mery L, et al. A global cancer surveillance framework within noncommunicable disease surveillance: making the case for population-based cancer registries. Epidemiol Rev. 2017;39(1):161-9. doi: https://doi.org/10.1093/epirev/mxx003

5. Mohammadzadeh Z, Ghazisaeedi M, Nahvijou A, et al. Systematic Review of Hospital Based Cancer Registries (HBCRs): necessary tool to improve quality of care in cancer patients. Asian Pac J Cancer Prev. 2017;18(8):2027-33. doi: https://doi.org/10.22034/APJCP.2017.18.8.2027

6. Instituto Nacional de Câncer [Internet]. Rio de Janeiro: INCA; [data desconhecida]. Registros Hospitalares de Câncer (RHC); 2022 jun 23 [atualizado 2022 jul 11; acesso 2022 nov 3]. Disponível em: https://www.gov.br/inca/pt-br/assuntos/cancer/numeros/registros/rhc

7. Integrador RHC: Registros Hospitalares de Câncer [Internet]. Rio de Janeiro: INCA. [2012] – [acesso 2022 set 20]. Disponível em: https://irhc.inca.gov.br/RHCNet

8. Thuler LCS, Hartz Z. Big Data em pesquisas populacionais sobre câncer: desafios e oportunidades. Rev Bras Cancerol. 2019;65(4):e-03718. doi: https://doi.org/10.32635/2176-9745.RBC.2019v65n4.718

9. Instituto Nacional de Câncer. Registros Hospitalares de Câncer [Internet]. 2. ed. Rio de Janeiro: INCA; 2010 [acesso 2023 fev 19]. Disponível em: https://www.inca.gov.br/sites/ufu.sti.inca.local/files/media/document/registros-hospitalares-de-cancer-2010.pdf

10. Oliveira JCS, Azevedo EFS, Caló RS, et al. Registros Hospitalares de Câncer de Mato Grosso: análise da completitude e da consistência. Cad Saúde Colet. 2021;29(3):330-43. doi: https://doi.org/10.1590/1414-462X202129030230

11. Lopes-Júnior LC, Dell'Antonio LS, Pessanha RM, et al. Completeness and consistency of epidemiological variables from Hospital-Based Cancer Registries in a Brazilian state. Int J Environ Res Public Health. 2022;19(19):12003. doi: https://doi.org/10.3390/ijerph191912003

12. Oliveira NPD, Cancela MC, Martins LFL, et al. Completeness of cervical cancer staging information in Brazil: a national hospital-based study. Cancer Epidemiol. 2022;79:102191. doi: https://doi.org/10.1016/j.canep.2022.102191

13. Lima KYN, Cancela MC, Souza DLB. Spatial assessment of advanced-stage diagnosis and lung cancer mortality in Brazil. PLoS One. 2022;17(3):e0265321. doi: https://doi.org/10.1371/journal.pone.0265321

14. Miranda-Filho A, Charvat H, Bray F, et al. A modeling analysis to compare eligibility strategies for lung cancer screening in Brazil. EClinicalMedicine. 2021;42:101176. doi: https://doi.org/10.1016/j.eclinm.2021.101176

15. Araujo LH, Baldotto C, Castro Jr G, et al. Lung cancer in Brazil. J Bras Pneumol. 2018;44(1):55-64. doi: https://doi.org/10.1590/S1806-37562017000000135

16. Mathias C, Prado GF, Mascarenhas E, et al. Lung cancer in Brazil. J Thorac Oncol. 2020;15(2):170-5. doi: https://doi.org/10.1016/j.jtho.2019.07.028

17. Ministério da Saúde (BR), Secretaria de Atenção Especializada à Saúde. Portaria nº 957, de 26 de setembro de 2014. Aprova as Diretrizes Diagnósticas e Terapêuticas do Câncer de Pulmão [Internet]. Diário Oficial da União, Brasília, DF. 2014 set 29 [acesso 2022 set 20]; Seção 1:59. Disponível em: http://bvsms.saude.gov.br/bvs/saudelegis/sas/2014/prt0957_26_09_2014.html

18. Sobin LH, Wittekind C, editors. TNM: Classificação de Tumores Malignos [Internet]. Eisenberg ALA, tradução. 6. ed. Rio de Janeiro: INCA; 2004 [acesso 2022 nov 3]. Disponível em: https://bvsms.saude.gov.br/bvs/publicacoes/inca/tnm2.pdf

19. Conselho Nacional de Saúde (BR). Resolução nº 466, de 12 de dezembro de 2012. Aprova as diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos. Diário Oficial da União, Brasília, DF. 2013 jun 13; Seção 1:59.

20. Souza MC, Vasconcelos AGG, Rebelo MS, et al. Profile of patients with lung cancer assisted at the National Cancer Institute, according to their smoking status, from 2000 to 2007. Rev Bras Epidemiol. 2014;17(1):175-88. doi: https://doi.org/10.1590/1415-790X201400010014ENG

21. Faria LV, Emmerick ICM, Silva MJS. Effect of municipal and state regulation on access and outcomes for cervical cancer patients in Rio de Janeiro, Brazil: an interrupted time series analysis. J Cancer Policy. 2022;33:100339. doi: https://doi.org/10.1016/j.jcpo.2022.100339

22. Ministério da Saúde (BR); Organização Pan-Americana da Saúde; Fundação Oswaldo Cruz. A experiência brasileira em sistemas de informação em saúde [Internet]. Vol 1, Produção e disseminação de informações sobre saúde no Brasil. Brasília (DF): Ministério da Saúde; 2009 [acesso 2023 jan 5]. (Série B. Textos Básicos de Saúde). Disponível em: https://bvsms.saude.gov.br/bvs/publicacoes/experiencia_brasileira_sistemas_saude_volume1.pdf