Observational Study

Medicine

OPEN

Incompleteness trends of epidemiological variables in a Brazilian high complexity cancer registry

An ecological time series study

Wesley Rocha Grippa, MSEE[a], Larissa Soares Dell'Antonio, RN, PhD[a], Luciane Bresciani Salaroli, PhD[a], Luís Carlos Lopes-Júnior, PhD[a,*]

Abstract

Hospital Cancer Registries serve as a vital source of information for clinical and epidemiological research, allowing the evaluation of patient care outcomes through therapeutic protocol analysis and patient survival assessment. This study aims to assess the trend of incompleteness in the epidemiological variables within the Hospital Cancer Registry of a renowned oncology center in a Brazilian state. An ecological time-series study was conducted using secondary data from the Hospital Santa Rita de Cássia Cancer Registry in Espírito Santo between 2000 and 2016. Data completeness was categorized as follows: excellent (<5%), good (5%–10%), fair (10%–20%), poor (20%–50%), and very poor (>50%), based on the percentage of missing information. Descriptive and bivariate statistical analyses were performed using the free software RStudio (version 2022.07.2) and R (version 4.1.0). The Mann–Kendall test was used to assess temporal trends between the evaluated years, and the Friedman test was employed to evaluate quality scores across the years. Among the variables assessed, birthplace, race/color, education, occupation, origin, marital status, history of alcohol and tobacco consumption, previous diagnosis and treatment, the most important basis for tumor diagnosis, tumor-node-metastasis staging (TNM) staging, and clinical tumor staging by group (TNM) showed the highest levels of incompleteness. Conversely, other epidemiological variables demonstrated excellent completeness, reaching 100% throughout the study period. Significant trends were observed over the years for history of alcohol consumption ($P < .001$), history of tobacco consumption ($P < .001$), TNM staging ($P = .016$), clinical tumor staging by group (TNM) ($P = .002$), first treatment received at the hospital ($P = .012$), disease status at the end of the first treatment at the hospital ($P < .001$), and family history of cancer ($P < .001$), and tumor laterality ($P = .032$). While most epidemiological variables within the Hospital Santa Rita de Cássia Cancer Registry exhibited excellent completeness, some important variables, such as TNM staging and clinical staging, showed high levels of incompleteness. Ensuring high-quality data within Cancer Registries is crucial for a comprehensive understanding of the health-disease process.

Abbreviations: ES = Espírito Santo, HCR = hospital cancer registry, HSRC = Hospital Santa Rita de Cássia, INCA = National Cancer Institute.

Keywords: cancer epidemiology, health surveillance, hospital cancer registry, oncology, prostate neoplasms.

1. Introduction

Globally, non-communicable diseases are the leading causes of illness and death in the population.[1] Among non-communicable diseases, malignant neoplasms hold a significant position and contribute to a substantial number of annual deaths worldwide (41 million deaths yearly, corresponding to 71%), accounting for 9.3 million cases, second only to cardiovascular diseases.[2,3] In Brazil, the National Cancer Institute (INCA) estimates a yearly occurrence of 704,080 new cancer cases during the 2023 to 2025 triennium. Among men, prostate cancer is projected to be the most prevalent type (30%), followed by colon and rectal cancers (9.2%).[4]

Importantly, population studies reveal an unequal geographic distribution in the incidence and aggressiveness of cancer,

This research received funding by Fundação de Amparo à Pesquisa e Inovação do Espírito Santo—FAPES, Edital FAPES/CNPq/Decit-SCTIE-MS/SESA No 09/2020—PPSUS. Termo de Outorga: 155/2021. Process Number: 2021-F0436.

The authors have no conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are not publicly available, but are available from the corresponding author on reasonable request.

This study involves human participants and was approved by an Ethics Committee or Institutional Board - Centro de Ciências da Saúde da Universidade Federal do Espírito Santo - CEP/CCS/UFES and approved under opinion no. 5.433.541 of July 18, 2022, in accordance with the relevant guidelines from the Declaration of Helsinki and the ethical principles in the National Health Council of Brazil.

[a] Graduate Program in Public Health, Health Sciences Center at the Federal University of Espírito Santo (UFES), Vitoria, ES, Brazil.

**Correspondence: Luís Carlos Lopes-Júnior, Graduate Program in Public Health at the Federal University of Espírito Santo (UFES), Av. Marechal Campos, 1468 – Maruípe, Zip Code: 29.043-900, Vitória, ES, Brazil (e-mail: lopesjr.lc@gmail.com).*

How to cite this article: Grippa WR, Dell'Antonio LS, Salaroli LB, Lopes-Júnior LC. Incompleteness trends of epidemiological variables in a Brazilian high complexity cancer registry: An ecological time series study. Medicine 2023;102:31(e34369).

Received: 21 May 2023 / Received in final form: 23 June 2023 / Accepted: 26 June 2023

http://dx.doi.org/10.1097/MD.0000000000034369

suggesting the influence of hereditary characteristics and lifestyle habits on the risk of developing the disease.[2,5–7]

The hospital cancer registry (HCR) has been implemented in many low- and middle-income countries, particularly in Asia and Latin America, to fulfill various objectives, including providing information on patient diagnosis and treatment as well as specific tumor characteristics and clinical outcomes within the hospital setting. However, the data collected may be subject to potential bias depending on the organization of the healthcare system, capturing a more or less representative subgroup of cancer patients.[8]

In Brazil, the INCA has developed the HCR to standardize technical offerings and provide national-level training to enhance hospital management for cancer patients' care. HCRs compile data on cancer cases diagnosed or treated within an institution.[9] Furthermore, HCRs serve as a valuable resource for clinical-epidemiological research, allowing the evaluation of results of therapeutic protocols and the analysis of patient survival rates.[8,10–14] The objective of the present study is to assess the trend of incompleteness in the epidemiological variables within the HCR of a renowned oncology center in a Brazilian state.

2. Methods

2.1. Study design

This is an ecological time-series study utilizing secondary data from the HCR database at Hospital Santa Rita de Cássia (HSRC) in Espírito Santo (ES) between 2000 and 2016. The data were obtained from comprehensive databases maintained by the Health Secretariat of ES.

The definition of quality dimensions proposed by Lima et al (2009)[15] was used, where completeness is determined by the proportion of fields containing non-null values. Additionally, for the completeness analysis, the classification proposed by Romero and Cunha (2006)[16] was adopted. The percentage of missing data was categorized as follows: excellent (<5%), good (5%–10%), fair (10%–20%), poor (20%–50%), or very poor (≥50%). Therefore, completeness refers to the extent to which the analyzed fields are filled, measured by the proportion of notifications with a category other than those indicating missing data. In this study, a field filled with the category "ignored," the numeral zero, an unknown date, or a term indicating missing data were considered incomplete.[16,17]

2.2. Study population and data collection procedures

The study utilized secondary data from the state of ES, situated in southeastern Brazil. The ES Oncology Care Network encompasses 3 health regions: North/Center, Metropolitan, and South.[14] Within this network, the oncology hospital unit at HSRC maintains a well-structured and operational HCR. In addition, the HSRC oncology hospital unit, which is the only Highly Complex Oncology Center in ES, is a reference in Oncology for the entire state of ES, Brazil. The hospital maintains a philanthropic characteristic and allocates 60% of its services to patients of the Brazilian Unified Health System-SUS. As a result, it also receives patients from other Brazilian states, such as: from the south of the state of Bahia, east of the state of Minas Gerais and north of the state of Rio de Janeiro. It has a structured HCR that has been operating since 2000, with its databases annually forwarded to the Integrating System of the HCR (SIS-RHC). Furthermore, HSRC operates the Oncological Care Line, which establishes a systematic framework for cancer care across the state of ES. The objectives of the care line include reducing neoplasm-related mortality, enhancing accessibility to diagnostic and treatment procedures for cancer, and improving overall healthcare accessibility throughout the state.[18]

For this study, a total of 6545 observations for prostate cancer spanning the period from 2000 to 2016 were extracted from the HSRC HCR database. All cases recorded as either analytic or non-analytic were included in the analysis. Data collection for this study occurred between August 2021 and December 2021 at Health Secretariat of ES. The period from 2000 to 2016 was chosen for analysis due to the consolidation of data from all hospitals within the Oncology Care Network in the state of ES. Until December 2016, these hospitals submitted their respective HCR data, which were processed by the Epidemiological Surveillance of the state. It is important to note that the COVID-19 pandemic presented challenges and delays for hospitals in processing the submission of HCR to the Epidemiological Surveillance network, mainly due to operational reasons. To ensure the consistency and reliability of the data for this study, it was decided to maintain the standardization of the historical series, allowing for the utilization of consolidated data from this particular hospital.

2.3. Variables

The tumor registration form of the Brazilian hospital registry integrator[9] encompasses several epidemiological variables, including: gender; age; place of birth; race/skin color; schooling; occupation; provenance; marital status; history of alcohol consumption; history of tobacco consumption; date of first hospital visit; date of first tumor diagnosis; previous diagnosis and treatment; most important basis for tumor diagnosis; primary tumor location; detailed primary tumor location; histological type of primary tumor; tumor-node-metastasis staging (TNM) staging; clinical tumor staging by group (TNM); date of initiation of treatment; main reason for not performing antineoplastic treatment in the hospital; first treatment received at the hospital; disease status at the end of the first hospital treatment; date of death; family history of cancer; forwarding source; tumor laterality; occurrence of more than 1 primary tumor; first clinic attended; first treatment clinic; and relevant exams for tumor diagnosis and treatment planning.

The HCR tumor registration form serves multiple purposes, including gathering information from medical records, providing a concise summary of the case, and serving as a data entry document for inputting information into the computerized databases of the Brazilian hospital registry integrator.[9] The content of this form is designed to meet the information requirements of hospitals with a cancer registry, adhering to standardization guidelines established by the World Health Organization and the International Agency for Research on Cancer. These guidelines have been validated through expert consensus at meetings coordinated by the INCA.[9]

2.4. Statistical analysis

Statistical analyses were conducted using the free software RStudio (version 2022.07.2) and R (version 4.1.0). The completeness of the data was described based on the observed relative frequency and their corresponding completeness scores. The Friedman test[19] was employed to compare score classifications across different years. Additionally, the Mann–Kendall test[20,21] was utilized to assess the presence of statistically significant temporal trends over the evaluated years. A significance level of 5% was adopted for all analyses.

2.5. Ethical considerations

This study received ethical approval from the Research Ethics Committee of the Federal University of ES Health Science Center, with opinion number 5433,541. Furthermore, approval and authorization were obtained from the ES State Health Department, based in Vitória, the capital city, to collect secondary data and access to restricted data associated with this research.

3. Results

3.1. Sociodemographic variables in the HSRC HCR: description of incompleteness and classification of completeness

During the study period, a total of 6545 cases of prostate cancer were registered in the HSRC HCR. The number of cases per year is as follows: 63 (2000), 168 (2001), 212 (2002), 225 (2003), 260 (2004), 288 (2005), 281 (2006), 407 (2007), 357 (2008), 377 (2009), 556 (2010), 417 (2011), 646 (2012), 520 (2013), 632 (2014), 506 (2015), and 630 (2016).

The data completeness of the sociodemographic variable "place of birth" in 2000 was classified as fair, with 14.25% missing data. However, for the following years, it was classified as excellent for 13 years and good for 3 years. The variable "race/skin color" presented 17.69% (regular) incompleteness in 2013, which was a discrepant year compared to other years, where it had excellent indexes for 14 years and good for 2 years. The variable "schooling" was classified as poor in 2009 and 2010, with 24.14% and 38.31% missing data, respectively. In other years, it was classified as excellent (8 years) or good (7 years). The variable "occupation" was classified as poor in 2006, 2007, and 2010, with 10.68%, 11.30%, and 11.69% missing data, respectively. In other years, it fluctuated between excellent (10 years) and good (4 years).

Except for the year 2003, which recorded 0.44% missing observations, the variable "origin" obtained 100% completeness and was classified as excellent in all years. The variable "marital status" was also classified as excellent in all years, with its highest incompleteness rate in 2001, with only 2.38% missing data. The variables "history of alcohol consumption" and "history of tobacco consumption" were classified as very poor in most years, with 100% and 98.41% missing data, respectively, in the year 2000. However, both variables showed excellent levels of completeness from 2010 to 2013 before returning to being classified as poor and very poor in the following years. The variables "gender" and "age" showed 100% completeness for all years evaluated. For a detailed breakdown of the completeness ratings year by year, please refer to Table 1.

3.2. Description of incompleteness and classification of completeness of the clinical variables of the HSRC HCR

Regarding the clinical variables, the TNM staging variable was consistently classified as very poor for all the years analyzed. It had the lowest completeness in 2009, with 97.35% missing information. The variable representing clinical tumor staging by group (TNM) also exhibited high levels of incompleteness. It was classified as very poor from 2000 to 2013, except for 2006, with a poor index with 25.62% missing information. In the most recent years analyzed, this variable showed a progressive decrease in incompleteness, with 43.67%, 36.96%, and 31.90% of missing data for 2014, 2015, and 2016, respectively. The treatment provided in the hospital was classified as very poor from 2000 to 2009, with incompleteness ranging from 72.48% to 93.45%. However, starting in 2010, the rating changed to poor, with missing data ranging between 21.35% and 38.39%.

The variable "date of start of treatment" exhibits a range of incompleteness varying from 0.27% to 16.61%, with classifications fluctuating between excellent, good, and fair. Starting from 2010, the completeness classifications shifted from excellent to

Table 1

Percentage of incompleteness and classification of completeness of the sociodemographic variables of the Hospital Santa Rita de Cássia (HSRC) hospital cancer registry (HCR) between 2000 and 2016.

Variables		2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Sex	Incompleteness (%)	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Scoring	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Age	Incompleteness (%)	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Scoring	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Place of birth	Incompleteness (%)	14,29	0,00	2,36	6,22	9,23	4,51	2,49	0,98	0,84	0,27	0,36	0,24	2,79	0,77	3,8	4,15	6,35
	Scoring	3	1	1	2	2	1	1	1	1	1	1	1	1	1	1	1	2
Race/skin color	Incompleteness (%)	1,59	2,38	0,47	0,00	0,38	0,35	0,36	0,00	0,56	0,00	0,18	2,88	5,11	17,69	6,49	4,15	1,43
	Scoring	1	1	1	1	1	1	1	1	1	1	1	1	2	3	2	1	1
Schooling	Incompleteness (%)	3,17	2,38	2,83	2,22	3,08	13,89	14,59	5,90	2,24	24,14	38,31	6,95	9,13	9,81	9,34	3,36	3,65
	Scoring	1	1	1	1	1	2	2	2	1	4	4	2	2	2	2	1	1
Occupation	Incompleteness (%)	0,00	1,19	2,83	3,56	3,85	2,78	10,68	11,3	5,04	5,84	11,69	1,44	5,26	7,12	4,11	2,17	3,65
	Scoring	1	1	1	1	1	1	3	3	2	2	3	1	2	2	1	1	1
Provenance	Incompleteness (%)	0,00	0,00	0,00	0,44	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,15	0,00	0,00	0,00	0,00
	Scoring	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Marital status	Incompleteness (%)	0,00	2,38	0,00	0,89	0,00	0,35	0,00	0,74	0,00	0,00	1,26	0,96	0,15	0,19	0,32	0,79	0,16
	Scoring	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
History of alcohol consumption	Incompleteness (%)	100,00	90,48	96,70	95,11	93,85	91,32	88,97	83,05	90,2	84,08	0,54	0,72	0,15	0,96	21,99	69,37	66,98
	Scoring	5	5	5	5	5	5	5	5	5	5	1	1	1	1	4	5	5
History of tobacco consumption	Incompleteness (%)	98,41	88,1	94,34	94,22	86,15	81,25	76,87	72,73	73,39	65,52	0,18	0,24	0,31	0,96	20,41	64,43	60
	Scoring	5	5	5	5	5	5	5	5	5	5	1	1	1	1	4	5	5

1 = Excellent (<5%), 2 = Good (5%–10%), 3 = Fair (10%–20%), 4 = Poor (20%–50%), 5 = Very poor (≥50%).

regular and remained so until 2016, reaching the highest level of incompleteness in 2014.

The variable "main reason for not performing the antineoplastic treatment in the hospital" was rated as excellent or good for 13 out of the 17 years evaluated, with incompleteness ranging from 0.44% to 9.27%. However, there was a drastic change in its rating, sharply transitioning to very poor for 2006, 2008, 2009, and 2010, with 100% of observations missing each year. The variable "family history of cancer" was consistently classified as very poor throughout the study, with missing information ranging from 64.62% to 100% in all years.

The variable "source of referral" was classified as poor and regular throughout the study period, with percentages of missing data ranging from 10.87% to 30.16%. On the other hand, the variables "first treatment received at the hospital," "tumor laterality," "exams relevant to the diagnosis and planning of tumor therapy," "diagnosis and previous treatment," and "most important basis for tumor diagnosis" were consistently classified as excellent, with mean percentages of incompleteness of 0%, 15%, 0.14%, 0.67%, 0.54%, and 0.29%, respectively, for the years 2000 to 2016.

Several other clinical variables, including "type of case," "date of the first consultation," "date of diagnosis," "primary location," "detailed primary location," "histological type of primary tumor," "date of death," "occurrence of more than one primary tumor," "clinic of first visit," and "clinic from the start of treatment," achieved 100% completeness and were consistently classified as excellent in all years studied. Table 2 presents an overview of the incomplete details of clinical variables from 2000 to 2016.

When comparing the scores of the epidemiological variables of the HSRC HCR, no significant difference was observed ($P = .9076$) in the classification of the scores. In other words, the classification remained similar from 2000 to 2016.

3.3. Trend of incompleteness of sociodemographic and clinical variables of the HSRC HCR

The Mann–Kendall test was employed to examine whether there were significant trends for sociodemographic and clinical variables over the years. The results of the test are provided in Table 3. Notably, significant decreasing trends were observed for the variables history of alcohol consumption, history of tobacco consumption, TNM staging, clinical tumor staging by group (TNM), first treatment received in the hospital, disease status at the end of the first treatment in the hospital, family history of cancer, and significant upward trends for the variable tumor laterality.

The variables that achieved 100% completeness in all years studied were not included in the Mann–Kendall test and, as a result, are not presented in Table 3. Figure 1 displays the graphs of the historical series depicting the percentage of incompleteness for the variables that exhibited significant trends according to the Mann–Kendall test from 2000 to 2016. The time series with incomplete data are represented by black lines, while the blue line represents the time trend. Among the variables analyzed, only handedness displayed a positive trend. However, as indicated in Table 3 and Figure 1 (tumor laterality), the percentage values of incompleteness for handedness are very small, ranging from 0% to 0.44%.

4. Discussion

Our findings revealed that several crucial variables for understanding the health-disease process were classified as excellent. These variables include sex, age, date of diagnosis, primary location, and histological type of the primary tumor. However, despite many variables being classified as good or excellent, many clinical-epidemiological variables of relevance contained missing information. Some variables had an average of <10% missing data, such as race/skin color, education, marital status, occupation, previous diagnosis and treatment, and the most important basis for tumor diagnosis. Other variables had an incompleteness rate exceeding 50%, such as history of alcohol and tobacco consumption, TNM staging, clinical tumor staging by group (TNM), and family history. In line with our findings, a study conducted in Mato Grosso, Brazil, which examined the quality of information and assessed the completeness and consistency of the HCR, identified TNM and education as the most incomplete variables.[22]

The gender variable consistently demonstrated excellent performance throughout the study period, and its significance was acknowledged.[6] This finding can be attributed to the minimal subjectivity involved in recording this information, contributing to the results' robustness. Other studies have reported similar findings, highlighting the excellent completeness of the gender variable.[22,23]

The completeness of the race/skin color variable fluctuated between 2011 and 2013, with an increasing trend of non-completeness during that period. However, it returned to values below 2% of missing information in 2016. Except for 2013, this variable demonstrated excellent and good completeness ratings. In the context of prostate cancer studies, the race/skin color variable holds immense significance due to the higher incidence and lower survival rates observed in individuals of African and Asian origin.[5,24] It should be noted that this variable encompasses more than just biological differences; it represents a complex variable that encapsulates socioeconomic and cultural factors, highlighting inequities in accessing healthcare, particularly in cancer diagnosis and treatment. Our findings align with other research conducted in Brazil.[14,22,23,25,26] The incompleteness and inaccuracies in recording this variable pose challenges in accurately assessing the need for programs focused on health promotion and disease prevention in vulnerable populations.[26] Furthermore, the race/skin color variable contributes to broader discussions on social inclusion, individual and political vulnerabilities, and programmatic approaches.[14,27]

Except for 2009 and 2010, where the average rate of missing observations exceeded 30%, the education variable demonstrated excellent and good scores. This variable exhibited a different pattern than other studies on HCRs in Brazil.[14,22,25,26] The education variable holds significant clinical and epidemiological relevance as it greatly influences a patient prognosis.[14] Its low incompleteness is crucial as it enables various comparisons such as early diagnosis, treatment adherence, survival evaluation, and disease recurrence.[28] Studying the level of education also allows for insights into socioeconomic situations where precise income information may be lacking.[14,28] Moreover, this variable can be associated with late diagnosis, highlighting how lower levels of education present obstacles to accessing early diagnosis, treatment, and prognosis.[22,26]

The occupation variable in the HSRC HCR exhibited varying levels of incompleteness, ranging from 0% to 11.69%, and remained classified as excellent or good for most of the years studied. Compared to other studies involving HCRs in Brazil, this variable demonstrated good completeness quality. A study assessing the completeness of occupation information in an HCR in Brazil reported a lack of information in 46% of cases (40% for men), with only a slight improvement over the years of the study.[29] Numerous studies have established associations between specific occupations and an increased likelihood of developing cancer or experiencing cancer-related mortality, underscoring the importance of capturing detailed information regarding work activities.[22,30,31] Furthermore, occupation serves as a significant diagnostic marker, particularly

Table 2

Percentage of incompleteness and classification of completeness of the variables related to diagnosis and treatment of the Hospital Santa Rita de Cássia (HSRC) hospital cancer registry (HCR) between 2000 and 2016.

Variables		2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Type of case	Incompleteness (%)	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Scoring	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Date of first hospital visit	Incompleteness (%)	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Scoring	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Date of first tumor diagnosis	Incompleteness (%)	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Scoring	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Previous diagnosis and treatment	Incompleteness (%)	4,76	0,00	0,00	0,00	0,00	0,00	0,00	0,49	0,00	0,00	0,72	0,48	0,77	0,58	0,32	0,4	0,63
	Scoring	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Most important basis for tumor diagnosis	Incompleteness (%)	1,59	0,00	0,00	0,44	0,77	0,00	0,00	0,00	0,00	0,00	0,00	0,24	0,00	0,96	0,32	0,2	0,48
	Scoring	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Primary tumor location	Incompleteness (%)	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Scoring	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Detailed primary tumor location	Incompleteness (%)	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Scoring	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Histological type of primary tumor	Incompleteness (%)	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Scoring	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
TNM staging	Incompleteness (%)	95,24	92,26	94,34	89,33	84,62	89,24	42,55	96,31	95,24	97,35	87,95	88,73	85,29	76,73	71,15	71,54	74,76
	Scoring	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	5
Clinical tumor staging by group (TNM)	Incompleteness (%)	79,37	85,12	89,62	80,44	74,69	79,17	25,62	82,06	80,95	85,41	69,42	63,55	65,48	53,27	43,67	36,96	31,9
	Scoring	5	5	5	5	5	5	4	5	5	5	5	5	5	5	4	4	4
Date of initiation of treatment	Incompleteness (%)	6,35	6,55	7,08	6,22	0,77	0,69	0,36	0,49	0,28	0,27	12,41	11,27	14,24	15,77	16,61	10,67	16,19
	Scoring	2	2	2	2	1	1	1	1	1	1	3	3	3	3	3	3	3
Main reason for not performing antineoplastic treatment in the hospital	Incompleteness (%)	3,17	4,17	2,83	0,44	1,92	4,51	100,00	3,19	100,00	100,00	100,00	3,6	5,11	9,27	9,02	5,93	7,94
	Scoring	1	1	1	1	1	1	5	1	5	5	5	1	2	2	2	2	2
First treatment received at the hospital	Incompleteness (%)	1,59	0,00	0,47	0,44	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Scoring	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Disease status at the end of the first hospital treatment	Incompleteness (%)	76,19	93,45	86,79	76,89	74,62	73,26	72,95	72,48	78,99	81,7	34,53	22,54	38,39	21,35	36,71	37,35	25,4
	Scoring	5	1	5	5	5	5	5	5	5	5	4	4	4	4	4	4	4
Date of death	Incompleteness (%)	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Scoring	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Family history of cancer	Incompleteness (%)	100,00	97,02	99,53	96,44	97,69	97,57	95,37	96,56	93,28	87	87,77	79,62	77,24	64,62	78,96	77,67	85,87
	Scoring	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Forwarding source	Incompleteness (%)	30,16	37,5	24,06	13,33	11,92	15,97	20,69	14,25	17,93	13,53	18,35	19,42	17,49	17,69	13,13	10,87	19,68
	Scoring	4	4	4	3	3	3	4	3	3	3	3	3	3	3	3	3	3

(*Continued*)

Table 2

(Continued)

Variables		2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Tumor laterality	Incompleteness (%)	0,00	0,00	0,00	0,00	0,38	0,00	0,00	0,25	0,28	0,00	0,18	0,00	0,15	0,19	0,32	0,4	0,16
	Scoring	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Occurrence of more than 1 primary tumor	Incompleteness (%)	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Scoring	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
First clinic attended	Incompleteness (%)	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,24	0,00	0,00	0,00	0,00	0,00
	Scoring	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
First treatment clinic	Incompleteness (%)	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Scoring	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Relevant exams for tumor diagnosis and treatment planning	Incompleteness (%)	0,00	0,60	0,00	0,00	4,23	1,39	0,00	0,49	4,76	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Scoring	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

1 = Excellent (<5%), 2 = Good (5%–10%), 3 = Fair (10%–20%),; 4 = Poor (20%–50%), 5 = Very poor (≥50%).

Table 3

Evaluation of the incompleteness trend of the epidemiological variables of the Hospital Santa Rita de Cássia (HSRC) hospital cancer registry (HCR) between 2000 and 2016.

Variable	S	*P* value	Trend
Place of birth	–10	.710	Non significant
Race/skin color	35	.160	Non significant
Schooling	36	.149	Non significant
Occupation	34	.174	Non significant
Provenance	–3	.881	Non significant
Marital status	15	.554	Non significant
History of alcohol consumption	**–86**	**<.001***	**Decrease**
History of tobacco consumption	**–90**	**<.001***	**Decrease**
Previous diagnosis and treatment	40	.088	Non significant
Most important basis for tumor diagnosis	14	.559	Non significant
TNM staging	**–59**	**.016***	**Decrease**
Clinical tumor staging by group (TNM)	**–74**	**.002***	**Decrease**
Date of initiation of treatment	40	.108	Non significant
Main reason for not performing antineoplastic treatment in the hospital	46	.061	Non significant
First treatment received at the hospital	**–41**	**.012***	**Decrease**
Disease status at the end of the first hospital treatment	**–82**	**<.001***	**Decrease**
Family history of cancer	**–102**	**<.001***	**Decrease**
Forwarding source	–34	.174	Non significant
Tumor laterality	**50**	**.032***	**Increase**
Occurrence of more than 1 primary tumor	6	.609	Non significant
Relevant exams for tumor diagnosis and treatment planning	–28	.164	Non significant

In bold are the variables that were statistically significant $P < 0.05$.

*P value < .05.

in cases such as lung cancer, where the work environment can be a potential source of exposure to carcinogenic agents.[32]

When it comes to recording the variables related to the history of alcohol consumption and tobacco consumption, their measurement is subjective, leading to high rates of incompleteness. This can introduce bias when quantifying the use of these substances or even result in the omission of this information.[33] However, an interesting finding emerged for the years 2010 to 2013, where these variables exhibited atypical completion rates, reaching excellent levels of completeness. This finding is particularly relevant considering the well-established carcinogenic potential of alcohol and tobacco.[34]

The completeness of clinical variables, specifically TNM staging and clinical tumor staging by group (TNM), was very poor. Similar findings have been reported in studies conducted in Brazil[14,28,30] that analyzed HCR data and classified completeness levels as "poor." However, excellent completeness for the TNM variable has been observed in another state within the country.[26] TNM staging is a globally used classification system that holds tremendous importance in understanding the extent of the disease. This information is crucial in defining the appropriate therapeutic plan, evaluating the treatment outcomes for individuals with cancer, and facilitating standardized procedures and exchange of experiences among cancer treatment institutions.[14,22,35,36] Moreover, knowledge of staging contributes to assessing the quality of care provided to cancer patients and aids in implementing public policies focused on early diagnosis.[9,36] The staging information is also highly relevant as a widely utilized prognostic factor in survival studies.

Family history of cancer serves as a valuable marker for comprehending the health-disease process[26,37–41] and plays a significant role in early cancer detection.[42–44] Unfortunately, for the HSRC HCR, this variable exhibited a classification of very poor, with an average of 88.95% missing information during the studied period. This level of incompleteness is concerning, considering that a family history of cancer is a known risk factor for prostate cancer.[2,5,6]

Undoubtedly, there is a pressing need for high-quality data due to political commitments towards cancer information and achieving health goals related to cancer. However, it is concerning to note that only one in 3 countries worldwide possess high-quality incidence data, and only 1 in 4 has high-quality mortality statistics.[45]

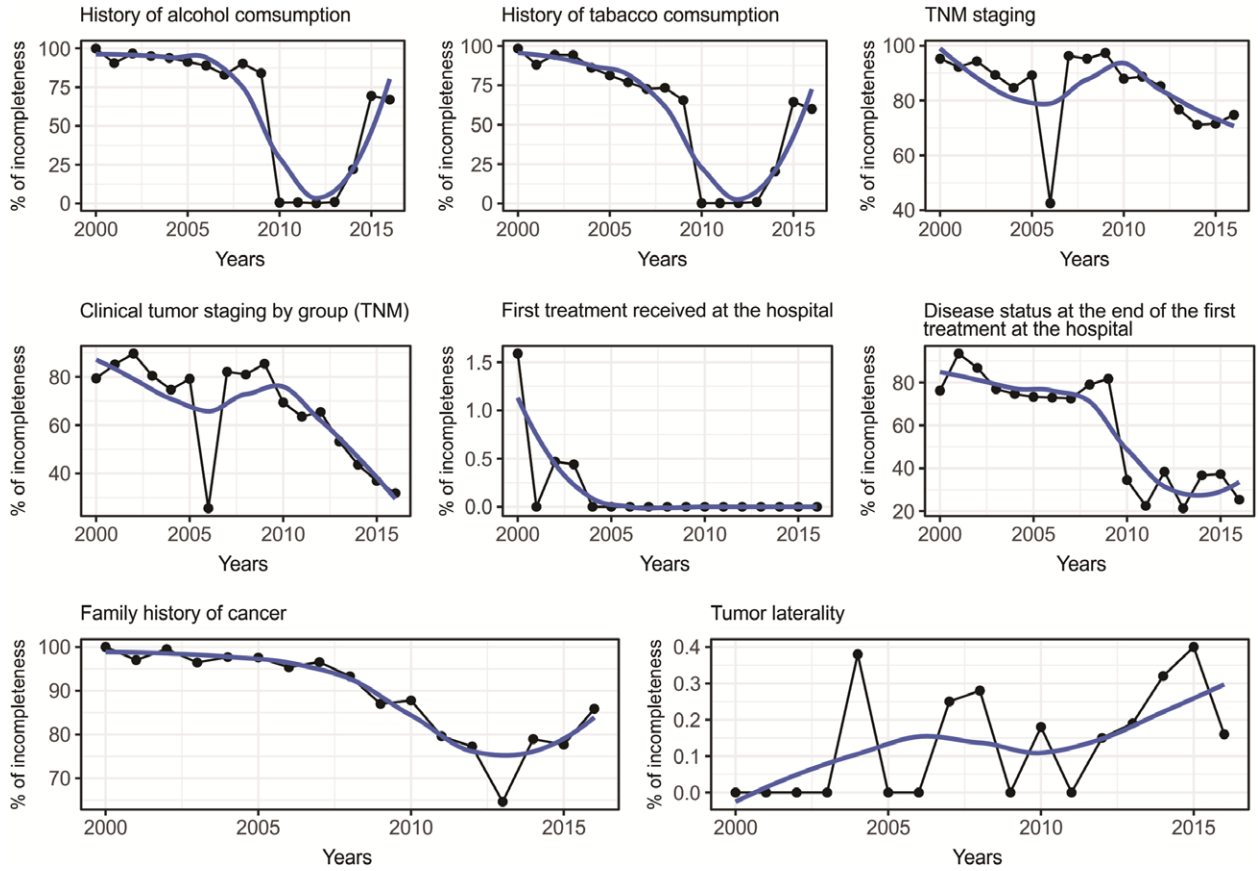

Figure 1. Trend of incompleteness of variables: history of alcohol consumption; history of tobacco consumption; TNM staging; clinical tumor staging by group (TNM); first treatment received at the hospital; disease status at the end of the first hospital treatment; family history of cancer and tumor laterality for HSRC HCR data between 2000 and 2016. HCR = hospital cancer registry, HSRC = Hospital Santa Rita de Cássia.

5. Limitations

The present study has some limitations. Firstly, it should be noted that the study exclusively utilized data obtained from a single HCR located within a specific Brazilian state. As a result, caution must be exercised when interpreting the findings regarding their external validity and generalizability to broader populations. Secondly, although the HSRC HCR provides valuable insights into the quality of services offered, it does not comprehensively represent the underlying local, regional, or national cancer epidemiology. The data collected within the HCR is derived from patient care within a specific hospital or the number of biopsied cancers within the system, which means that the inclusion of cases is contingent upon the resources and expertise available within the respective institutions. Consequently, the aggregated cases recorded within the HCR represent only a subset of the total cases in the broader context.

6. Conclusion

Most epidemiological variables examined from the HSRC HCR in ES, Brazil, demonstrated excellent completeness. However, notable variables like TNM staging and clinical tumor staging by group (TNM) exhibited high rates of incompleteness between 2000 and 2016. Ensuring high-quality data within HCRs is crucial for enhancing our understanding of the health-disease process and evaluating the quality of hospital care delivered. Additionally, reliable and comprehensive data collection within HCRs is indispensable, as these records serve as the foundation for planning public policies and conducting research focused on cancer surveillance.

Acknowledgments

The authors gratefully acknowledge the strong support of the Secretaria de Estado da Saúde do Espírito Santo, Vitória, ES, Brazil.

Author contributions

Conceptualization: Wesley Rocha Grippa, Luís Carlos Lopes-Júnior.
Data curation: Wesley Rocha Grippa, Larissa Soares Dell'Antonio, Luciane Bresciani Salaroli, Luís Carlos Lopes-Júnior.
Formal analysis: Wesley Rocha Grippa, Luís Carlos Lopes-Júnior.
Funding acquisition: Luís Carlos Lopes-Júnior.
Investigation: Wesley Rocha Grippa and Luís Carlos Lopes-Júnior.
Methodology: Wesley Rocha Grippa and Luís Carlos Lopes-Júnior.
Project administration: Luís Carlos Lopes-Júnior.
Resources: Luís Carlos Lopes-Júnior.
Software: Wesley Rocha Grippa, Luís Carlos Lopes-Júnior.
Supervision: Luís Carlos Lopes-Júnior.
Validation: Wesley Rocha Grippa, Larissa Soares Dell'Antonio, Luciane Bresciani Salaroli, Luís Carlos Lopes-Júnior.
Visualization: Wesley Rocha Grippa, Larissa Soares Dell'Antonio, Luciane Bresciani Salaroli, Luís Carlos Lopes-Júnior. **Writing – original draft:** Wesley Rocha Grippa, Larissa Soares Dell'Antonio, Luciane Bresciani Salaroli, Luís Carlos Lopes-Júnior.
Writing – review & editing: Luís Carlos Lopes-Júnior.

References

[1] World Health Organization. Health Statistics and Information Systems: Disease Burden and Mortality Estimates; WHO: Geneva, Switzerland, 2022.

[2] Siegel RL, Miller KD, Wagle NS, et al. Cancer statistics, 2023. CA Cancer J Clin. 2023;73:17–48.

[3] Lopes-Júnior LC, De Lima RAG. Cancer care and interdisciplinary practice. Cad Saude Publica. 2019;35:e00193218.

[4] Brasil. Instituto Nacional de Câncer José Alencar Gomes da Silva. Coordenação de Prevenção e Vigilância. Estimativa 2023: Incidência do Câncer no Brasil. 2022.

[5] Garbin JRT, Leite FMC, Lopes-Júnior LC, et al. Analysis of survival of patients hospitalized with COVID-19 in Espírito Santo, Brazil. Int J Environ Res Public Health. 2022;19:8709.

[6] Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71:209–49.

[7] Dell'Antonio LS, Leite FMC, Dell'Antonio CSDS, et al. COVID-19 mortality in public hospitals in a Brazilian state: an analysis of the three waves of the pandemic. Int J Environ Res Public Health. 2022;19:14077.

[8] Bray F, Znaor A, Cueva P, et al. Planning and developing population-based cancer registration in low- or middle-income settings. In The Role and Status of Population-Based Cancer Registration; IARC Technical Report, No. 43; Chapter 2; International Agency for Research on Cancer: Lyon, France, 2014.

[9] Brasil. Instituto Nacional de Câncer. Registros Hospitalares de Câncer: Planejamento e Gestão/Instituto Nacional de Câncer, 2ª ed.; INCA: Rio de Janeiro, Brasil, 2010; 536p.

[10] 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.

[11] Bhatia A, Victora CG, Beckfield J, et al. "Registries are not only a tool for data collection, they are for action": cancer registration and gaps in data for health equity in six population-based registries in India. Int J Cancer. 2020;148:2171–83.

[12] Ahn Y-O. Cancer registration in Korea: the present and furtherance. J Prev Med Public Health. 2007;40:265–72.

[13] Lopes-Júnior LC. Os registros de câncer no contexto da vigilância em saúde pública: ensaio teórico. Int J Develop Res. 2021;11:45693–6.

[14] 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;19:12003.

[15] Lima CR, Schramm JM, Coeli CM, et al. Review of data quality dimensions and applied methods in the evaluation of health information systems. Cad Saude Publica. 2009;25:2095–109.

[16] Romero DE, Da Cunha CB. Quality of socioeconomic and demographic data in relation to infant mortality in the Brazilian Mortality Information System (1996/2001). Cad Saude Publica. 2006;22:673–81.

[17] Dell'Antonio LS, Leite FMC, Dell'Antonio CSDS, et al. Completeness and quality of information about death from COVID-19 in a Brazilian state: a descriptive population-based register study. Medicine (Baltim). 2023;102:e33343.

[18] Secretaria do Estado do Espírito Santo (SESA). Secretaria do Estado do Espírito Santo. Informativo Vigilância do Câncer. Gerência Estratégica de Vigilância em Saúde (GEVS). Núcleo Especial de Vigilância Epidemiológica(Neve) Doenças e Agravos Não Transmissíveis (DANT`s). Vigilância do Câncer. Coordenação Estadual dos Registros de Câncer do estado do Espírito Santo. 2017.

[19] Pettersson A, Robinson D, Garmo H, et al. Age at diagnosis and prostate cancer treatment and prognosis: a population-based cohort study. Ann Oncol. 2018;29:377–85.

[20] Kendall MG. Rank Correlation Methods. London: Griffin; 1975.

[21] Mann HB. Nonparametric tests against trend. Econometrica. 1945;13:245–59.

[22] Oliveira JCS, Azevedo EFS, Caló RS, et al. Hospital-based cancer registries of Mato Grosso, Brazil: analysis of completeness and consistency. Cad Saude Colet. 2021;29:330–43.

[23] Felix JD, Zandonade E, Amorim MHC, et al. Evaluation of the plenitude of epidemiological variables of the Information System on Mortality of women with deaths from breast cancer in the Southeast Region: Brazil (1998–2007). Cien Saude Colet. 2012;17:945–53.

[24] Wu D, Yang Y, Jiang M, et al. Competing risk of the specific mortality among Asian-American patients with prostate cancer: a surveillance, epidemiology, and end results analysis. BMC Urol. 2022;22:42.

[25] Pinto IV, Ramos DN, Costa MCE, et al. Completeness and consistency of data in hospital-based cancer registries in Brazil. Cad Saude Colet. 2012;20:113–20.

[26] Brandão-Souza C, Amorim MHC, Zandonade E, et al. Completude dos prontuários de idosas com câncer de mama: estudo de tendência. Acta Paul Enferm. 2019;32:416–24.

[27] Ayres, JRCM, França Júnior, I, Calazans, GJ, et al. O conceito de vulnerabilidade e as práticas de saúde: novas perspectivas e desafios. In Czeresnia D, Freitas CM, Organizadores. Promoção da Saúde—Conceitos, Desafios, Tendencias; Fiocruz: Rio de Janeiro, Brazil, 2003; pp. 117–138.

[28] Rebelo, PAP, Lima, RGM, Souto Rebelo, M. Registros Hospitalares de Cancer—Rotinas e Procedimentos; INCA: Rio de Janeiro, Brazil, 2000.

[29] Grabois MF, Souza MC, Guimarães RM, et al. Completeness of information "Occupation" in hospital cancer records in Brazil: basis for surveillance of work-related cancer. Rev Bras Cancerol. 2014;60:207–14.

[30] Pignati W, Oliveira NP, da Silva AMC. Surveillance on pesticides: quantification of use and prediction of impact on health, work and the environment for Brazilian municipalities. Cien Saude Colet. 2014;19:4669–78.

[31] Miranda-Filho AL, Monteiro GTR, Meyer A. Brain cancer mortality among farm workers of the State of Rio de Janeiro, Brazil: a population-based case–control study, 1996–2005. Int J Hyg Environ Health. 2012;215:496–501.

[32] Algranti E, Buschinelli JTP, Capitani EM. Occupational lung cancer. J Bras Pneumol. 2010;36:784–94.

[33] World Health Organization (WHO). Self-Help Strategies-For Cutting Down Orstopping Substance Use-A Guide. Genève: WHO; 2010.

[34] Leite RB, Marinho ACO, Costa BL, et al. The influence of tobacco and alcohol in oral cancer: literature review. J Bras Patol Med Lab. 2021;57:1–5.

[35] Brasil. Ministério da Saúde. Instituto Nacional de Câncer José Alencar Gomes da Silva. Coordenação de Prevenção e Vigilância. Secretaria Nacional de Assistência à Saúde. Registros hospitalares de câncer: planejamento e gestão. 2. ed. Rio de Janeiro: INCA; 2010. 538 p.

[36] Brasil. Ministério da Saúde. Instituto Nacional de Câncer José Alencar Gomes da Silva. Coordenação Geral de Prevenção e Vigilância. TNM: classificação de tumores malignos. 7. ed. Rio de Janeiro: INCA; 2012. 325 p.

[37] Floria-Santos M, Lopes-Júnior LC, Alvarenga LDM, et al. Self-reported cancer family history is a useful tool for identification of individuals at risk of hereditary cancer predisposition syndrome at primary care centers in middle-income settings: a longitudinal study. Genet Mol Biol. 2016;39:178–83.

[38] Pessanha RM, Schuab SIPC, Nunes KZ, et al. Use of Family history taking for hereditary neoplastic syndromes screening in primary healthcare: a systematic review protocol. PLoS One. 2022;17:e0271286.

[39] Silva TBDCE, Macdonald DJ, Ferraz VEDF, et al. Perception of cancer causes and risk, family history and preventive behaviors of users in oncogenetic counseling. Rev Esc Enferm USP. 2013;47:377–84.

[40] Lopes-Júnior LC, Carvalho Júnior PM, de Faria Ferraz VE, et al. Genetic education, knowledge and experiences between nurses and physicians in primary care in Brazil: a cross-sectional study. Nurs Health Sci. 2017;19:66–74.

[41] Lopes-Júnior LC, Bomfim E, Flória-Santos M. Genetics and genomics teaching in nursing programs in a latin American Country. J Pers Med. 2022;12:1128.

[42] Sundquist M, Brudin L, Tejler G. Improved survival in metastatic breastcancer 1985–2016. Breast. 2017;31:46–50.

[43] Bello MA, Menezes RF, Silva B, et al. Impact of treatment type on overall survival in elderly Brazilian women with breast cancer. Asian Pac J Cancer Prev. 2016;17:4769–74.

[44] Yu XQ, Luo Q, Kahn C, et al. Temporal trends show improved breast cancer survival in Australia but widening urban–rural differences. Breast. 2015;24:524–7.

[45] WHO report on cancer: setting priorities, investing wisely and providing carefor all. https://apps.who.int/iris/handle/10665/330745 Date: 2020.