Original Article

Hospital-Based Cancer Registry Audit of the Clínica Sagrada Esperança, Luanda, Angola

Auditoria ao Registo Oncológico de Base Hospitalar da Clínica Sagrada Esperança, Luanda, Angola

Lygia Vieira Lopes[1*], José Figueiredo[2*], Teresa Garcia[3,4,5], Creusa Costa[1], Sandra Costa[1], Lúcio Lara Santos[3,6,7]

1. Clínica Sagrada Esperança, Luanda, Angola
2. Unidade de Saúde Pública, Unidade Local de Saúde São João, Portugal
3. Centro de Especialização em Registo de Cancro dos Países Africanos de Língua Portuguesa (CEROPAL)
4. Grupo de Epidemiologia, Resultados, Economia e Gestão em Oncologia, CI-IPOP/RISE@CI-IPOP/Porto.CCC, Portugal
5. Serviço de Epidemiologia, Instituto Português de Oncologia do Porto (IPO-Porto), Portugal
6. Grupo de Patologia e Terapêutica Experimental, CI-IPOP, Instituto Português de Oncologia, Porto, Portugal
7. Escola de Medicina e Ciências Biomédicas, Universidade Fernando Pessoa, Porto, Portugal

* *These authors contributed equally*

Corresponding Author/Autor Correspondente:
Lygia Vieira Lopes [lygiavieiralopes@hotmail.com]
Clínica Sagrada Esperança, Luanda, Angola

https://doi.org/10.34635/rpc.1180

ABSTRACT

Introduction and Objectives: Luanda faces a growing cancer burden alongside a high prevalence of communicable diseases, yet lacks a population-based cancer registry. The Clínica Sagrada Esperança (CSE), a semi-private institution in Luanda, maintains a hospital-based cancer registry (CSE-HBCR) through a manual data curation process with recognized limitations. This study aimed to (1) perform a structured quality audit of the CSE-HBCR database and (2) establish a descriptive cancer profile of cases diagnosed and/or treated at the institution.

Methods: The study included 1,113 records registered at the CSE-HBCR between 2012 and 2025. The audit systematically assessed missing values, invalid entries, temporal, logical, and internal consistency, textual variability, and duplicate records. Following data curation, a descriptive analysis was performed to characterize the cancer profile.

Received/Recebido: 06/01/2026 **Accepted/Aceite:** 06/03/2026 **Published online/Publicado online:** 16/03/2026 **Published/Publicado:** 16/03/2026

Results: The audit identified 8,851 missing values (31.8% of all expected fields), with the highest proportions in clinical (57.9%) and follow-up (35.0%) variables. Additional quality issues included invalid entries, age discrepancies between recorded and calculated values (402 cases), sex–tumour incompatibilities (8 cases), high textual variability in key variables, and 50 duplicate records. After deduplication, 1,094 cancer cases were analysed. The most frequent cancers were prostate (29.4% of male cases) and breast (37.6% of female cases). Staging information was absent in 71.9% of cases.
Conclusions: This first comprehensive audit of the CSE-HBCR reveals critical data quality deficiencies that compromise the registry's usefulness for cancer surveillance and research. Implementing standardized protocols—including the adoption of ICD-O-3 coding, the CanReg5 software platform, and a structured quality improvement plan—is essential to strengthen the registry of hospitals that treat cancer in Luanda and support the future Luanda Population-Based Cancer Registry.

Keywords: Cancer registry; Data quality audit; Oncology; Epidemiology; Angola

RESUMO

Introdução e Objetivos: Luanda enfrenta um aumento progressivo da carga oncológica em contexto de elevada prevalência de doenças transmissíveis, mas carece de um registo de cancro de base populacional. A Clínica Sagrada Esperança (CSE) dispõe de um registo hospitalar (CSE-HBCR), sustentado por curadoria manual de dados com limitações reconhecidas. Este estudo teve como objetivos (1) realizar uma auditoria estruturada da qualidade da base de dados do CSE-HBCR e (2) estabelecer um perfil descritivo do cancro nos casos diagnosticados e/ou tratados na instituição.
Métodos: O estudo incluiu 1.113 registos do CSE-HBCR entre 2012 e 2025. A auditoria avaliou sistematicamente valores em falta, entradas inválidas, consistência temporal, lógica e interna, variabilidade textual e registos duplicados. Após a curadoria dos dados, realizou-se uma análise descritiva do perfil oncológico.
Resultados: A auditoria identificou 8.851 valores em falta (31,8% dos campos esperados), com maiores proporções nas variáveis clínicas (57,9%) e de seguimento (35,0%). Identificaram-se igualmente entradas inválidas, discrepâncias entre a idade registada e a calculada (402 casos), incompatibilidades sexo–tumor (8 casos), elevada variabilidade textual em variáveis-chave e 50 registos duplicados. Após remoção dos casos duplicados, foram analisados 1.094 casos. Os cancros mais frequentes foram o da próstata (29,4% nos homens) e o da mama (37,6% nas mulheres). A informação de estadiamento estava ausente em 71,9% dos casos.
Conclusões: Esta primeira auditoria abrangente do CSE-HBCR evidencia deficiências críticas na qualidade dos dados que comprometem a utilidade do registo para a vigilância e investigação oncológica. A implementação de protocolos normalizados — incluindo a codificação ICD-O-3, a plataforma CanReg5 e um plano estruturado de melhoria da qualidade — é essencial para fortalecer o registo de todos os hospitais que tratam o cancro em Luanda e apoiar o futuro Registo de Cancro de Base Populacional de Luanda.

Palavras-chave: Registo de cancro; Auditoria de qualidade de dados; Oncologia; Epidemiologia; Angola

INTRODUCTION

Luanda is the capital and largest city of Angola, serving as the country's main political, economic, and cultural hub. Located on the Atlantic coast, its population exceeded 8.8 million inhabitants according to the 2024 National Census.[1] Nationally, life expectancy at birth stands at 65 years, while the median age in Luanda is 19.4 years,[1,2] reflecting a predominantly young demographic structure.

This demographic profile places Angola in a dual disease burden scenario: a persistently high prevalence of communicable diseases coexisting with a rapidly growing impact of non-communicable diseases, particularly cardiovascular diseases and cancer.[3,4]

Luanda currently lacks a population-based cancer registry, which limits the production of reliable national cancer estimates. Until recently, GLOBOCAN projections for Angola were derived exclusively from data from neighbouring countries. The 2022 GLOBOCAN release represented an important advance by incorporating data from a population-based registry in Lubango and a hospital-based registry from the Instituto Angolano de Controlo do Cancro (IACC) in Luanda.[5]

The Clínica Sagrada Esperança (CSE), a semi-private institution in Luanda, contributes to the future Luanda population-based cancer registry through its hospital-based cancer registry (CSE-HBCR). Cancer data are currently retrieved and curated manually from clinical records of cases diagnosed and/or treated at the institution.[6] However, this manual process entails recognized limitations in data completeness, timeliness, and quality that must be systematically assessed to improve registry utility.

This study has two primary aims: (1) to perform a structured quality audit of the CSE-HBCR database, quantifying data quality deficiencies and proposing targeted corrective measures; and (2) to establish a descriptive cancer profile of cases diagnosed and/or treated at the Clínica Sagrada Esperança.

METHODS

1. STUDY POPULATION AND DATA SOURCE

The CSE-HBCR, developed in Microsoft Excel, contains records of cancer patients diagnosed and/or treated at the Clínica Sagrada Esperança. Active collection from the Oncology Unit has been ongoing since 2012, although a small number of cases dating back to 1998 are also included. The dataset, extracted in 2025, comprised 1,113 records and 25 variables covering: patient identification (name, date of birth, hospital ID, registry ID, date of registration); demographics (sex, place of birth, age); tumour characteristics (diagnosis date, topography and morphology, hormone receptor status, tumour grade, stage, lymph node involvement, metastasis location, tumour site); treatment (attending physician, surgery date and type, treatment modality); and follow-up (transfer, date of death, date of last consultation).

2. DATA QUALITY AUDIT

Variables were classified into four categories: (i) patient identification variables; (ii) core variables; (iii) clinical variables; and (iv) follow-up and treatment variables. For each variable, missing values were defined as blank fields with no recorded entry, and invalid values as completed entries that, after minimal standardization (trimming whitespace, normalizing formatting), did not conform to the expected data type or were uninterpretable for analysis.

Three dimensions of consistency were evaluated. Temporal consistency was assessed by identifying records where the date of birth preceded the dates of diagnosis, registration, or death, or where any of these dates fell after the dataset submission date. Internal consistency was examined on the age variable by comparing recorded age against the age calculated from date of birth and date of diagnosis (or date of registration, given the registry's ambiguity on the reference point). Logical consistency was assessed by flagging records with sex–tumour incompatibilities (male genital tract cancers in females and vice versa).

Textual variability was quantified by counting distinct categories in text variables expected to take a limited set of values, after standardizing capitalization, diacritics, punctuation, and spacing. Variables with expected free-text entries (e.g., patient name) or purely numeric/date fields were excluded. Distinct categories were calculated on valid entries only, and proportions were computed accordingly.
Duplicate records were defined as repeated registrations referring to the same patient and the same tumour type/location, or entries suggestive of progression of the same tumour episode. Identifier inconsistency was assessed by verifying whether the same medical record number had been assigned to different patients.

3. CANCER PROFILE

Following the quality audit, data were manually curated to generate the cancer profile. The analysis included absolute and relative frequencies, focused on variables identified in the audit as having the highest data quality. Given the identified data quality issues, particularly the high proportion of missing staging information, all proportions were calculated using the total number of cases as the denominator, unless otherwise stated.

RESULTS

1. CANCER REGISTRY AUDIT RESULTS

The CSE-HBCR dataset comprised 1,113 records in wide format with 25 variables. Table 1 summarizes missing values, invalid values, and the number of distinct categories for text variables.

A total of 8,851 missing values were observed, corresponding to 31.8% of all expected fields for the cancer registry. Clinical variables presented the highest proportion of missing data (57.9%), followed by follow-up and treatment variables (35.0%), core variables (4.9%), and identification variables (2.2%). Missing staging information and lymph node data were particularly pronounced, absent in 71.9% and 95.3% of cases, respectively.

Regarding temporal consistency, three records had a date of birth later than the date of diagnosis, one had a date of birth later than the date of death, and one had a date of birth later than the dataset submission date.

Table 1. *Data quality audit of the CSE-HBCR database: missing values, invalid values, and textual variability.*

Variables	Missing values		Invalid values		Distinct categories	
	n	%	n	%	n	%
Identification variables	**198**	**2.24**	**1**	**0.99**	**—**	**—**
Hospital ID	13	0.15	0	—	—	—
Registry ID	28	0.32	0	—	—	—
Name	0	—	0	—	—	—
Sex	0	—	0	—	—	—
Place of birth	157	1.77	1	0.99	77	4.12
Core variables	**430**	**4.86**	**45**	**44.55**	**—**	**—**
Date of birth	90	1.02	8	7.92	—	—
Age	46	0.52	0	—	—	—
Date of diagnosis	2	0.02	20	19.80	—	—
Type of cancer	21	0.24	7	6.93	473	25.32
Date of registration	271	3.06	10	9.90	—	—
Clinical variables	**5125**	**57.90**	**11**	**10.89**	**—**	**—**
Diagnosis situation	523	5.91	3	2.97	450	24.09
Tumour receptors	992	11.21	0	—	98	5.25
Grade of differentiation	769	8.69	0	—	84	4.50
Number/location of lymph nodes	1036	11.71	2	1.98	52	2.78
Stage	772	8.72	0	—	226	12.10
Location of metastasis	898	10.15	1	0.99	43	2.30
Location of tumour	2	0.02	5	4.95	237	12.69
Attending physician	133	1.50	0	—	44	2.36
Follow-up and treatment variables	**3098**	**35.00**	**44**	**43.56**	**—**	**—**
Transfer/Death follow-up	921	10.41	1	0.99	—	—
Date of death (n=217)	6	0.07	7	6.93	—	—
Date of surgery (n=110)	19	0.21	21	20.79	—	—
Type of surgery (n=110)	46	0.52	4	3.96	41	2.20
Treatment	1038	11.73	0	—	43	2.30
Last medical appointment	1068	12.07	11[1]	10.89	—	—
Total	**8851**	**100**	**101**	**100**	**1868**	**100**

[1] Considering that a value was expected to be entered in date format.

Regarding logical consistency, eight records presented incompatibilities between the recorded sex and the tumour type or anatomical site, all attributable to apparent sex miscoding upon cross-checking against the patient›s name.

Regarding internal consistency, recorded age differed from the age calculated using the date of diagnosis in 402 cases (36.3% of the total records) and from the age calculated using the date of registration in 37 cases (3.3%). Ages calculated from diagnosis versus registration dates differed in 302 patients (43.6% of records with valid values for both dates). This discrepancy is significant as international guidelines recommend recording age at diagnosis in cancer registries.[7,8]

High textual variability was observed across multiple variables. The type of cancer variable presented 473 distinct categories (25.3% of all distinct categories identified), reflecting inconsistent use of nomenclature, abbreviations, and free-text descriptions. Similarly, diagnosis situation (24.1%) and stage (12.1%) showed marked heterogeneity, complicating any systematic analysis.

A total of 50 duplicate records were identified. We found that in this group 6 patients had a second tumor. Additionally, four medical record numbers were assigned to more than one patient, representing a critical identifier inconsistency that compromises data integrity.

2. CANCER PROFILE

After removing duplicates, 1,094 cancer cases were included in the cancer profile analysis. Among these, 492 (45.0%) were female, and 602 (55.0%) were male (Table 2). The majority were born in Angola (81.4%), followed by Portugal (1.5%). The year of diagnosis was missing in 42 cases (3.8%).

Of the registered cases, diagnosis occurred between 1998 and 2025. The highest number of diagnoses was recorded in 2016 (n=150, 13.7%), followed by 2023 (8.3%), 2021 (8.2%), and 2024 (8.2%). Overall, 1,028 cases (93.9%) were recorded between 2010 and 2025, reflecting the active registration period since 2012.

Staging information was unavailable for 787 cases (71.9%), lymph node involvement data were absent in 1,043 cases (95.3%), and 199 cases (18.2%) had documented metastasis at presentation. Tumour location was recorded for all 1,094 cases.

Table 3 presents the distribution of cancer by sex. Among females, breast cancer was the most frequently diagnosed malignancy (185 cases; 37.6%), followed by cervical cancer (63 cases; 12.8%), corpus uteri (27 cases; 5.5%), ovarian cancer (19 cases; 3.9%), colon cancer (19 cases; 3.9%), and lung cancer (20 cases; 4.1%). Among males, prostate cancer was the predominant cancer type (177 cases; 29.4%), followed by stomach (52 cases; 8.6%), liver (39 cases; 6.5%), and lung cancer (38 cases; 6.3%). Kaposi's sarcoma was notably more frequent in males (26 cases) than females (5 cases), as were head and neck (20 vs. 7 cases) and oesophageal cancers (22 vs. 2 cases).

Table 2. *Characteristics of cancer cases at the Clínica Sagrada Esperança (n=1,094).*

Variables	n	%
Sex		
Female	492	45.0
Male	602	55.0
Country of birth		
Angola	891	81.4
Cabo Verde	5	0.5
Congo	1	0.1
England	1	0.1
Mozambique	1	0.1
Portugal	16	1.5
Democratic Republic of the Congo	1	0.1
Russia	1	0.1
São Tomé and Príncipe	5	0.5
Viet Nam	1	0.1
Unknown	171	15.6
Year of diagnosis		
1998–2009 (pre-registry period)	16	1.5
2010	8	0.7
2011	11	1.0
2012	53	4.8
2013	65	5.9
2014	53	4.8
2015	73	6.7
2016	150	13.7
2017	65	5.9
2018	23	2.1
2019	60	5.5
2020	66	6.0
2021	90	8.2
2022	75	6.9
2023	91	8.3
2024	90	8.2
2025	63	5.8
Missing	42	3.8
Stage information available		
No	787	71.9
Yes	307	28.1
Lymph node information available		
No	1043	95.3
Yes	51	4.7
Metastasis at presentation		
No	895	81.8
Yes	199	18.2

Table 3. *Distribution of cancer by sex (n=1,094).*

Cancer	Female (n=492)		Male (n=602)		Total (n=1094)	
	n	%	n	%	n	%
Head and neck	7	1.4	20	3.3	27	2.5
Oesophagus	2	0.4	22	3.7	24	2.2
Stomach	13	2.6	52	8.6	65	5.9
Small intestine	1	0.2	0	0.0	1	0.1
Colon	19	3.9	34	5.6	53	4.8
Rectum	18	3.7	20	3.3	38	3.5
Anus and anal canal	0	0.0	2	0.3	2	0.2
Liver	12	2.4	39	6.5	51	4.7
Gallbladder and biliary tract	6	1.2	5	0.8	11	1.0
Pancreas	11	2.2	10	1.7	21	1.9
Larynx	1	0.2	10	1.7	11	1.0
Lung	20	4.1	38	6.3	58	5.3
Other thoracic organs	0	0.0	1	0.2	1	0.1
Bones and cartilage	2	0.4	5	0.8	7	0.6
Malignant melanoma	3	0.6	5	0.8	8	0.7
Other skin cancers	6	1.2	11	1.8	17	1.6
Kaposi's sarcoma	5	1.0	26	4.3	31	2.8
Connective and soft tissue	1	0.2	12	2.0	13	1.2
Breast	185	37.6	5	0.8	190	17.4
Vulva	2	0.4	—	—	2	0.2
Cervix uteri	63	12.8	—	—	63	5.8
Corpus uteri	27	5.5	—	—	27	2.5
Ovary	19	3.9	—	—	19	1.7
Placenta	2	0.4	—	—	2	0.2
Prostate	—	—	177	29.4	177	16.2
Testis	—	—	4	0.7	4	0.4
Kidney	4	0.8	10	1.7	14	1.3
Bladder	9	1.8	12	2.0	21	1.9
Eye and adnexa	0	0.0	4	0.7	4	0.4
Brain and CNS	7	1.4	10	1.7	17	1.6
Thyroid gland	4	0.8	7	1.2	11	1.0
Adrenal gland	1	0.2	0	0.0	1	0.1
Endocrine glands	0	0.0	1	0.2	1	0.1
Hodgkin lymphoma	5	1.0	16	2.7	21	1.9
Non-Hodgkin lymphoma	5	1.0	15	2.5	20	1.8
Multiple myeloma	8	1.6	11	1.8	19	1.7
Lymphoid leukaemia	4	0.8	7	1.2	11	1.0
Myeloid leukaemia	0	0.0	2	0.3	2	0.2
Primary unknown origin	6	1.2	13	2.2	19	1.7
Other ill-defined sites	7	1.4	3	0.5	10	0.9

Table 4 presents follow-up and treatment data. Of 1,094 cases, surgery was recorded in only 90 (8.2%) and systemic or palliative therapy in 52 (4.8%), reflecting major gaps in treatment data capture. Death was recorded in 217 cases. Recorded mortality increased notably from 2020 onwards, with 121 deaths (55.8% of total recorded deaths) occurring between 2020 and 2025.

Table 4. *Follow-up and treatment information (n=1,094).*

Variables	n	%
Surgery (n=1094)		
No	1004	91.8
Yes	90	8.2
Chemo-, radio- or palliative therapy (n=1094)		
No	1042	95.2
Yes	52	4.8
Year of death (n=1090 patients)		
2012	8	0.7
2013	14	1.3
2014	15	1.4
2015	15	1.4
2016	11	1.0
2017	11	1.0
2018	4	0.4
2019	11	1.0
2020	18	1.6
2021	32	2.9
2022	19	1.7
2023	17	1.6
2024	19	1.7
2025	16	1.5

DISCUSSION

This study presents the first comprehensive quality audit and cancer profile of the CSE-HBCR, a hospital-based cancer registry in Luanda, Angola. The audit—presented here as the primary contribution—identified substantial and multidimensional data quality deficiencies, while the resulting cancer profile provides a preliminary characterization of the oncological case-mix at this institution.

1. DATA QUALITY AUDIT

The overall proportion of cancer registry missing values (31.8%) and the particularly high rates in clinical (57.9%) and follow-up variables (35.0%) signal that, while the registry performs adequately in recording patient identification and core demographic data, its robustness deteriorates when tracking treatment pathways and clinical progression. This pattern is, in part, attributable to the receipt of treatment outside the institution, resulting in loss to clinical follow-up. Furthermore, some missing values may reflect clinical inapplicability—hormone receptors are not relevant for all tumour types, and not all patients will present with metastatic disease requiring site documentation, rather than data entry failure per se.

However, the absence of staging information in 71.9% of cases and lymph node data in 95.3% cannot be attributed solely to clinical inapplicability. These represent critical data gaps that fundamentally limit the registry's utility for prognosis, treatment planning, and international comparability. Staging is a core variable in any cancer registry, and its systematic absence constitutes the most pressing quality challenge identified in this audit.

The high textual variability observed across key variables—473 distinct categories for tumour type alone—renders systematic descriptive analysis unreliable without extensive manual curation. This heterogeneity likely reflects the absence of a controlled vocabulary, a data dictionary, and restricted data entry fields, all of which are standard components of quality-assured registry systems.

The discrepancy between recorded and calculated age is a particularly relevant finding, with 402 cases (36.3%) showing inconsistencies when age was compared against date of birth and date of diagnosis. International guidelines for cancer registries specify that age at diagnosis—not age at registration—should be the reference variable, yet the CSE-HBCR does not unambiguously define the age reference point. Given that diagnosis and registration dates may differ by months or years, this ambiguity propagates systematic errors in age-based analyses[7,8].

The identification of four medical record numbers assigned to multiple patients is a serious data integrity concern, as it undermines the reliability of patient tracking and the deduplication process. The 50 duplicate records (4.5% of the initial dataset) further highlight the need for systematic identifier management.

Based on these findings, we propose corrective measures across four domains. First, on patient identification: each cancer case should be assigned a unique registry identifier,

Table 5. *Action plan for the reorganization and quality improvement of the CSE-HBCR.*

Activity	Description	Responsibility
Training in cancer registration	Participation in training activities organized by the Luanda Cancer Registry and CEROPAL, covering data collection standards, ICD-O-3 coding, staging, and use of CanReg5.	Luanda Cancer Registry & CEROPAL
Internal organization	– Formal recognition of the hospital-based cancer registry at Clínica Sagrada Esperança; – Formal appointment of the registry team and delegation of responsibilities; – Formal designation of the headquarters, location and resources.	Luanda Cancer Registry and Board of Clínica Sagrada Esperança
Registry quality improvement	– Creation of a CSE-HBCR data dictionary with strict variable-entry rules; – Restriction of cell inputs in the database to control data entry; – Cross-validation of hospital records against registry data to prevent duplicates; – Retrospective review of records for accuracy of sex, age, histological and morphological diagnosis per ICD-O-3; – Implementation of staging using the UICC/AJCC manual for all applicable cases; – Automated consistency checks (e.g., age at diagnosis vs. calculated age from date of birth); – Annotation, review, and correction of inconsistent records; – Data validation using CanReg5 quality-control tools (IARC/AFCRN) to eliminate duplicates and illogical age–sex–site–morphology combinations; – Production of annual activity reports.	Director of CSE-HBCR
External audit	Periodic follow-up external audit of registry quality.	AFCRN / CEROPAL

distinct from the hospital medical record number, with both recorded consistently in the database. Second, on database structure: date fields should be enforced in DDMMYYYY format; all text fields should be standardized for capitalization and diacritics; and restricted-input cells should prevent free-text entry where categorical responses are expected. Third, on variable nomenclature: a comprehensive data dictionary should be developed to define each variable, its acceptable values, and entry rules. Conflated variables—such as 'surgery date' and 'surgery type', which were frequently confused in the current dataset—should be clearly differentiated. Fourth, on cancer classification: the International Classification of Diseases for Oncology, third edition (ICD-O-3),[9] should be adopted as the mandatory coding system for tumour topography and morphology, enabling standardized analysis and international comparability. Staging should follow the American Joint Committee on Cancer (UICC/AJCC) manual,[10] applied systematically to all applicable cases.

These recommendations can be operationalized through the adoption of CanReg5, an open-source cancer registry software developed by the International Agency for Research on Cancer (IARC), which has been implemented by numerous registries worldwide. CanReg5 incorporates standardized data entry, built-in validation rules, and quality control tools aligned with IARC/AFCRN guidelines. A structured action plan for registry reorganization, encompassing training, internal organization, quality improvement activities, and external audit cycles, is presented in Table 5.

2. CANCER PROFILE

The cancer profile findings should be interpreted within the constraints imposed by the identified data quality deficiencies, particularly the high proportions of missing clinical variables.

The male predominance in the overall case distribution (55.0%) differs from previous cancer profiles in Angola and may partly reflect the high burden of prostate cancer in this registry. [6,11] The substantial increase in registered cases since 2010 likely reflects improved registry coverage, enhanced diagnostic capacity, and greater healthcare access rather than a true increase in incidence; temporal trends should therefore be interpreted with caution, given varying registration completeness over time.

Among females, the predominance of breast cancer (37.6%) is consistent with epidemiological trends across sub-Saharan Africa and globally.[11] Cervical cancer ranked second (12.8%), though the observed gap between these two malignancies may partly reflect referral bias towards this semi-private

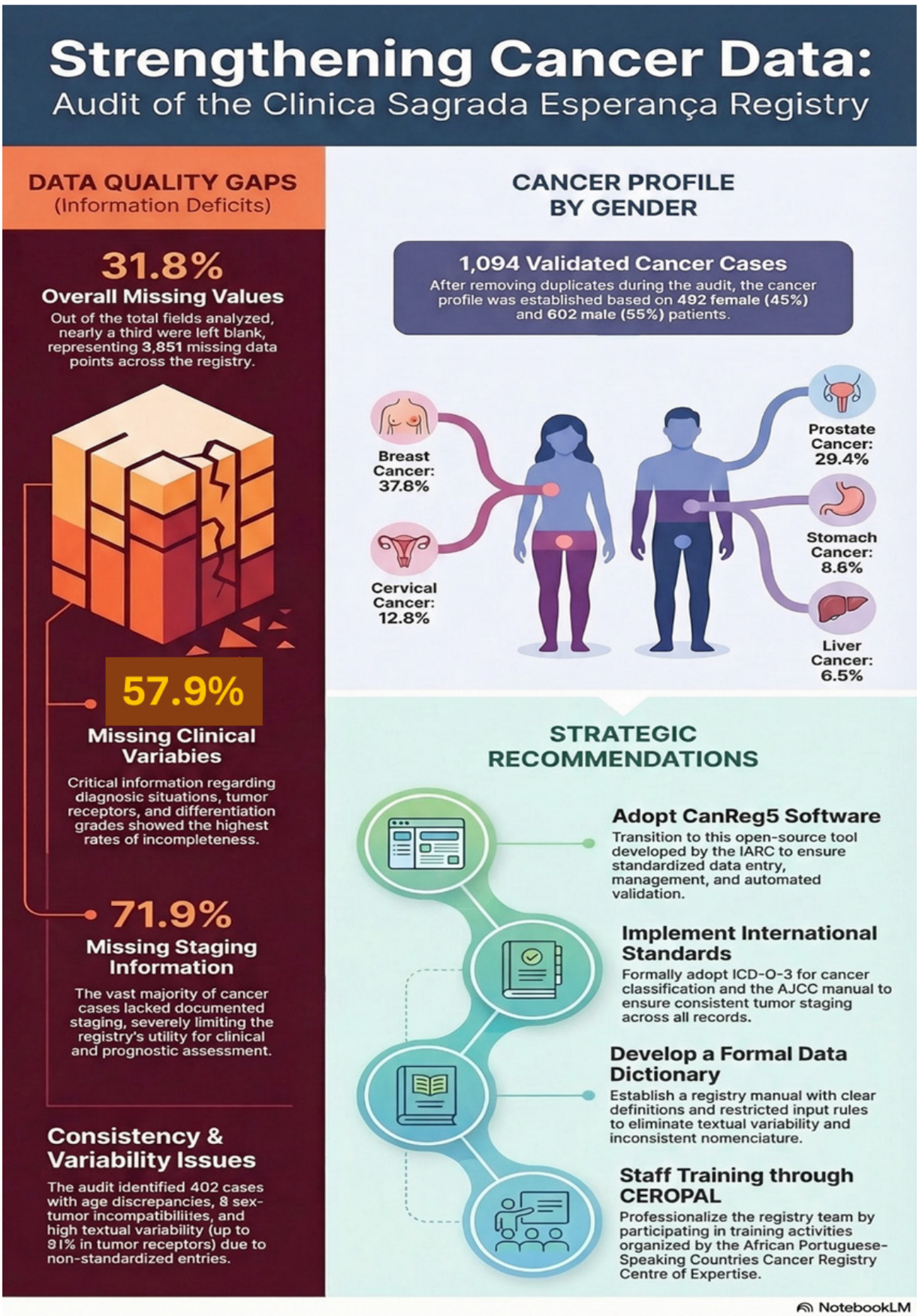

Figure 1. *A comprehensive audit and clinical profile of the hospital-based cancer registry at Clinica Sagrada Esperança in Luanda, Angola.*

institution, where access to colposcopy and oncological surgery may attract disproportionately more breast cancer cases. These findings nonetheless underscore the need for sustained investment in cervical cancer screening and HPV vaccination programmes.

In males, prostate cancer accounted for nearly one-third of all cases (29.4%), consistent with its status as the leading cancer among men in Africa.[11] The substantial proportions of stomach (8.6%), liver (6.5%), and lung cancer (6.3%) in males reflect established risk factor profiles, including tobacco and alcohol consumption, occupational exposures, and infectious aetiologies such as *H. pylori* and viral hepatitis, prevalent in many sub-Saharan African settings.[12] Similar risk factor profiles explain the male predominance in head and neck and oesophageal cancers.[13]

Kaposi's sarcoma comprised 2.8% of all registered cancers, with a marked male predominance (26 vs. 5 female cases). This pattern may reflect both HIV-associated and classic endemic Kaposi's sarcoma, warranting further investigation into aetiological factors in this population.[14]

The low proportion of cases with documented treatment data (8.2% for surgery; 4.8% for other therapies) likely reflects treatment delivery outside the registry's catchment institution as much as data capture failure. Recorded mortality data must also be interpreted cautiously, given incomplete follow-up: the 217 recorded deaths represent only a fraction of the true case fatality, and site-specific mortality proportions should not be equated with case-fatality rates given the severe follow-up limitations.

The increase in recorded deaths from 2020 onwards may reflect improved death registration, the impact of the COVID-19 pandemic on healthcare delivery and cancer outcomes, or both.

3. LIMITATIONS

This study has several limitations. The high proportion of missing data for key clinical variables—staging, treatment, and outcomes—substantially limits the interpretability of the cancer profile. The absence of population-based denominators precludes calculation of incidence rates and limits comparability with other populations. Temporal trends must be interpreted cautiously, given variations in registry completeness over time. Finally, the reliance on manual data curation means that the audit itself may not have captured all inconsistencies present in the original records.

CONCLUSIONS

This study provides the first comprehensive quality audit and cancer profile of the CSE-HBCR. The audit reveals multidimensional data quality deficiencies that critically limit the registry's value for cancer surveillance, research, and policy development. Addressing these deficiencies through standardized data collection protocols, adoption of ICD-O-3 and the UICC/AJCC staging manual, implementation of CanReg5, and investment in staff training is essential. The structured action plan proposed herein provides a practical roadmap for registry reorganization. Strengthening the CSE-HBCR is a strategic priority not only for the institution itself, but also for all hospitals treating cancer patients in Luanda, and for the future Luanda Population-Based Cancer Registry, which depends on high-quality data from contributing hospital registries to generate reliable national cancer estimates.

LEARNING POINTS / TAKE-HOME MESSAGES

- A structured quality audit of the CSE-HBCR identified critical data deficiencies: 31.8% missing values important for the Cancer registry, and 57.9% in clinical variables. Staging information was absent in 71.9% of cases.

- The cancer profile reveals prostate cancer (29.4% of male cases) and breast cancer (37.6% of female cases) as the predominant malignancies, consistent with sub-Saharan African epidemiological patterns.

- Adoption of ICD-O-3 coding, the CanReg5 software platform, and the structured action plan proposed herein are essential to transform the CSE-HBCR into a high-quality data source for the future Luanda Population-Based Cancer Registry.

ETHICAL DISCLOSURES

Conflicts of Interest: The authors have no conflicts of interest to declare.
Financing Support: This work has not received any external contribution, grant, or scholarship.
Confidentiality of Data: The authors declare that they followed their institution's protocols regarding the publication of patient data.
Protection of Human and Animal Subjects: The authors declare that the procedures followed were in accordance with the regulations of the relevant clinical research ethics committee and with the Code of Ethics of the World Medical Association (Declaration of Helsinki, as revised in 2024).
Provenance and Peer Review: Not commissioned; externally peer-reviewed.

RESPONSABILIDADES ÉTICAS

Conflitos de Interesse: Os autores declaram a inexistência de conflitos de interesse na realização do presente trabalho.
Fontes de Financiamento: Não existiram fontes externas de financiamento para a realização deste artigo.
Confidencialidade dos Dados: Os autores declaram ter seguido os protocolos da sua instituição acerca da publicação dos dados de doentes.
Proteção de Pessoas e Animais: Os autores declaram que os procedimentos seguidos estavam de acordo com os regulamentos estabelecidos pela Comissão de Ética responsável e de acordo com a Declaração de Helsínquia revista em 2024 e da Associação Médica Mundial.
Proveniência e Revisão por Pares: Não comissionado; revisão externa por pares.

CONTRIBUTORSHIP STATEMENT

LVL, LLS: Conceptualization/study design.
LVL, CC, SC: Data collection.
LVL, JF, TG, LLS: Statistical analysis, interpretation of results, drafting of the manuscript.
All authors: Critical revision of the article, final approval of the version to be published.

DECLARAÇÃO DE CONTRIBUIÇÃO

LVL, LLS: Conceptualização/desenho do estudo.
LVL, CC, SC: Recolha de dados.
LVL, JF, TG, LLS: Análise estatística, interpretação dos resultados, redação do manuscrito.
Todos os autores: Revisão crítica do artigo, aprovação final da versão a publicar.

REFERENCES

1. Instituto Nacional de Estatística (2025). Resultados Definitivos do Recenseamento Geral da População e Habitação 2024. INE. Luanda, Angola.
2. Statbase (2025). Life expectancy at birth | 2024. Available from: https://statbase.org/datasets/demographics/life-expectancy/, accessed [19 February 2026].
3. World Health Organization (2025). WHO Annual Report – Angola 2025. Available from: https://www.afro.who.int/pt/countries/angola/publication/relatorio-anual-de-oms-angola-2025, accessed [19 February 2026].
4. GBD 2023 Disease and Injury and Risk Factor Collaborators (2025). Burden of 375 diseases and injuries, risk-attributable burden of 88 risk factors, and healthy life expectancy in 204 countries and territories, 1990–2023: a systematic analysis for the Global Burden of Disease Study 2023. Lancet, 406(10513), 1873–1922. https://doi.org/10.1016/S0140-6736(25)01637-X
5. Ferlay J, Ervik M, Lam F, et al. (2024). Global Cancer Observatory: Cancer Today. Lyon: IARC. Available from: https://gco.iarc.who.int/today, accessed [19 February 2026].
6. Miguel F, Bento MJ, de Lacerda GF, Weiderpass E, Santos LL (2019). A hospital-based cancer registry in Luanda, Angola: the Instituto Angolano de Controlo do Cancer (IACC) Cancer Registry. Infect Agents Cancer, 14:35. https://doi.org/10.1186/s13027-019-0249-2

7. Bray F, Parkin DM (2009). Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. Eur J Cancer, 45(5):747–755. https://doi.org/10.1016/j.ejca.2008.11.032

8. Parkin DM, Bray F (2009). Evaluation of data quality in the cancer registry: principles and methods. Part II: completeness. Eur J Cancer, 45(5):756–764. https://doi.org/10.1016/j.ejca.2008.11.033

9. Fritz A, Percy C, Jack A, et al. (Eds.) International Classification of Diseases for Oncology (ICD-O), 3rd edition, 2nd revision. Geneva: World Health Organization; 2019.

10. Amin MB, Greene FL, Edge SB, et al. (2017). The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more 'personalized' approach to cancer staging. CA Cancer J Clin, 67(2):93–99. https://doi.org/10.3322/caac.21388

11. Bray F, Laversanne M, Sung H, et al. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin, 74(3):229–263. https://doi.org/10.3322/caac.21834

12. Parkin DM, Hämmerl L, Ferlay J, Kantelhardt EJ (2020). Cancer in Africa 2018: the role of infections. Int J Cancer, 146(8):2089–2103. https://doi.org/10.1002/ijc.32538

13. Hashibe M, Brennan P, Chuang SC, et al. (2009). Interaction between tobacco and alcohol use and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. Cancer Epidemiol Biomarkers Prev, 18(2):541–550. https://doi.org/10.1158/1055-9965.EPI-08-0347

14. Mesri EA, Cesarman E, Boshoff C (2010). Kaposi's sarcoma and its associated herpesvirus. Nat Rev Cancer, 10(10):707–719. https://doi.org/10.1038/nrc2888

15. Rawla P, Sunkara T, Gaduputi V (2019). Epidemiology of pancreatic cancer: global trends, etiology and risk factors. World J Oncol, 10(1):10–27. https://doi.org/10.14740/wjon1166