SJ Gynecology and Obstetrics Africa
Vol. 3 No. 3 (2026): August 2026 Issue
https://doi.org/10.51168/x488nx46
Original Article

Analysis of missing data and completeness modeling of cancer clinical records: The case of the Ngaliema Clinic in Kinshasa, Democratic Republic of Congo.

Pascal Atuba Mamenepi [1*], Héritier Kalubi Belangana[2], Jean-Pierre Omokende Olonge[3], Crispin Kakwaka Katombe [1], Jacques Lofandjola Masumbuku[1], Félicien Ilunga Ilunga[1], Félicien Tshimungu Kandolo[1]

[1]Teacher in the Nursing Sciences Department, Higher Institute of Medical Techniques of Kinshasa, Kinshasa, Democratic Republic of Congo

[2]Student in Nursing Sciences, Higher Institute of Medical Techniques of Kinshasa, Kinshasa, Democratic Republic of Congo

[3]Teacher at the Higher Institute of Medical Techniques of Kisantu, Kongo Central, Democratic Republic of Congo

Summary

Introduction

The present study aims to examine the completeness and model the exhaustiveness of clinical records of cancer patients followed at the Ngaliema Clinic in Kinshasa.

Materials and methods

A retrospective study analyzed the records of 81 patients with cancer followed at the Ngaliema clinic between 2023 and 2025. Using JASP 0.95.4.0 software, descriptive and inferential analyses based on the Little, Friedman, Durbin -Conover test and a multiple linear regression model were performed at the usual threshold of 0.05.

Results

There was a predominance of women and patients under 50 years of age, mostly residing in the Lukunga district, with cancer diagnosed at an invasive stage (stage II) among the patients whose records were analyzed. The MCAR hypothesis was verified ($p = 0.767$). A significant difference was observed between the five evaluated recording domains ($\chi^2 = 135$; $p < 0.001$). Furthermore, the overall completeness score decreases significantly according to age ($\beta = -0.174$; $p = 0.016$); bilateral organ involvement ($\beta = -7.349$; $p = 0.029$); year of detection 2024 vs 2025 ($\beta = -10.732$; $p < 0.001$); non-invasive nature of the disease ($\beta = -5.893$; $p = 0.047$); family history of cancer ($\beta = -4.491$; $p = 0.028$); of residence in the Funa district vs. Lukunga ($\beta = -4.547$; $p = 0.017$) and of initial traditional treatment/prayer ($\beta = -5.822$; $p = 0.025$). However, this score increases with the time elapsed before consultation ($\beta = 0.476$; $p < 0.001$) and then with origin from the Tshangu district ($\beta = 9.236$; $p = 0.009$).

Conclusion

The completeness of cancer records at the Ngaliema Clinic depends markedly on clinical severity, patient age, and therapeutic pathway.

Recommendation

Implement a standardized oncology form, strengthen digital skills, support digitization and develop a sustainable provincial cancer registry.

Keywords: *Cancer; Medical records; Missing data; Ngaliema Clinic; Kinshasa, Democratic Republic of Congo*
Submitted: March 23, 2026 **Accepted:** July 1, 2026 **Published:** August 1, 2026

Corresponding author: Pascal Atuba Mamenepi*
Email: pascalatuba@gmail.com
Teacher in the Nursing Sciences Department, Higher Institute of Medical Techniques of Kinshasa, Kinshasa, Democratic Republic of Congo ; Student in Biostatistics and Health Informatics at the Doctoral School of Health Sciences in Kinshasa.

Introduction

Cancer is increasingly becoming a central public health issue worldwide. Estimates for 2022 projected approximately 20 million new diagnoses and 9.7 million deaths, with a particularly pronounced increase in low- and middle-income countries. The development, implementation, and evaluation of prevention, diagnosis, treatment, and surveillance policies therefore require robust and reliable data.

However, cancer registries, along with hospital clinical records, constitute the main sources of information on patients, tumor characteristics, and treatment modalities (Parkin, 2006). Their scientific utility remains, however, closely contingent on the quality of the data collected (Bray

et al., 2014) . Insufficient completeness can impair the estimation of the burden of disease, reduce comparability between populations, and undermine data-informed decision-making (Jena et al., 2023) .

Furthermore, missing values represent a major methodological threat to the validity of epidemiological analyses. They degrade the accuracy of estimates and can introduce biases since their occurrence depends on patient characteristics or treatment. Their explicit handling is therefore essential to guarantee the robustness of results derived from clinical databases (Cro et al., 2020; Jakobsen et al., 2017) .

From this, the methodological literature classically distinguishes three mechanisms of missing data: completely random (MCAR), random conditional on observed information (MAR), and non-random (MNAR). Identifying the predominant mechanism guides the choice of analytical approaches and conditions the validity of statistical inferences (Bhaskaran & Smeeth, 2014) . This approach constitutes an essential step in any rigorous evaluation of the quality of health data.

Across sub-Saharan Africa, oncology information systems remain hampered by structural constraints that degrade the reliability and completeness of data (Ayubi et al., 2024) . Numerous analyses have highlighted coverage gaps, marked heterogeneity in completeness levels, and organizational obstacles that compromise the production of robust indicators necessary for epidemiological surveillance and the planning of cancer control interventions (Bray et al., 2022).

In the Democratic Republic of Congo, available knowledge on the quality of oncological information remains fragmented. Data originate primarily from a few specialized centers, and the determinants of incomplete clinical records have received virtually no rigorous scientific investigation. This deficiency hinders the development of information systems capable of effectively and sustainably supporting cancer surveillance.

This study examines the completeness and models the exhaustiveness of clinical records of cancer patients followed at the Ngaliema Clinic in Kinshasa. It aims to estimate the completeness of the main recording domains, to characterize the mechanisms of missing data, and to identify the determinants of the documentary quality of the records.

Methodology

Study setting

This study was conducted at the Ngaliema Clinic, a referral and tertiary level hospital in the Gombe health zone, one of 35 health zones in the Kinshasa Provincial Health Division, in the Democratic Republic of Congo, and the main zone where there has been consistency in the number of cancer cases reported over the last decade after an analysis of SNIS data at the DPS level.

This structure also provides diagnostic and therapeutic care for multiple neoplastic entities and has structured medical archives facilitating the retrospective exploitation of clinical data.

Study design

This is a retrospective observational study covering the period from 2023 to 2025.

Population and sampling

The population of interest consisted of all patients for whom cancer had been formally diagnosed and for whom a clinical file was kept in the archives of the Ngaliema Clinic. Only files containing, at a minimum, the information essential for patient identification and validation of the oncological diagnosis were included. Any illegible file or one that did not allow for a definitive diagnosis was excluded from the analysis. The primary statistical unit corresponded to each clinical file, and thus, 81 files were ultimately analyzed.

Data collection technique and source

Document analysis technique (document review) of the records provided by the oncology department of the Ngaliema Clinic was relied on. The data were recorded from the records into the KoboCollect application, which served as our data collection tool. This had the advantage of reducing the bias of multiple transcriptions that could occur if the information were initially recorded on printed paper for subsequent encoding.

Variables of interest

The analyzed information was divided into five levels of registration according to the terminology used in most cancer registries (Parkin , 2006; Bray et al., 2014; WHO, 2019; Jensen et al., 1991) . These levels are, for example:

(i) Patient identification: the name coded here, sex, age at diagnosis, level of education, occupation, and marital status.

(ii) Geographic location: province, municipality, district, avenue, number.

(iii) History, risk factors/behavior, and comorbidities: family history of cancer, obesity, smoking, alcoholism, diet, HIV status, Hepatitis B or C, HPV.

(iv) Tumor and diagnostic information: date of diagnosis, topography, laterality, morphology, evolutionary stage, TNM stage (AJCC - 8th edition), Biomarker status (Site-Specific variables) and histological type.

(v) Therapeutic management and follow-up: consultation time, initial therapeutic recourse, therapeutic regimen in hospital, patient outcome, method of monitoring vital status.

Each of these variables was coded "1" if information was marked in the patient's file, and "0" if it was not recorded.

Creation of completeness indicators

Completeness was assessed using a multi-level approach, successively considering variables, recording domains, and individual records. For each record, an overall completeness index was derived, corresponding to the proportion of variables actually completed divided by the total number of expected variables, then weighted as a percentage.

$$\text{Complétude (\%)} = \frac{\text{Number of variables entered}}{\text{Total number of variables expected}} \text{x } 100$$

After its application, five intermediate scores corresponding to the five levels or domains of registration were obtained, and then the overall score was also created.

Data processing and analysis

The data were processed using the modern version of Excel integrated with Microsoft 365 and JASP 0.95.4.0. The results were presented in figures and tables after descriptive and inferential analyses.

Initially, a descriptive characterization of the incompleteness was carried out in order to estimate, for each variable and for each of the recording fields, the proportions of missing values.

In a second step, Little's test was used to verify the hypothesis of completely random missing data (MCAR). In accordance with the usual non-rejection criteria, a p-value strictly greater than 0.05 was considered compatible with the MCAR hypothesis and the entire inferential part.

Thirdly, the distributions of completeness scores were compared between recording domains using the Friedman test, and then by post hoc comparisons according to the Durbin-Conover method.

Fourth, the identification of the determinants of completeness was based on a multiple linear regression model. The adjusted β coefficients, accompanied by their 95% confidence intervals, were presented.

However, the chosen regression model was systematically evaluated using the coefficient of determination, the Durbin -Watson test for autocorrelation, the Shapiro- Wilk test for residual normality, and the Cook and Mahalanobis distances for outliers and influential values. Collinearity was also verified based on VIF and tolerance, the values of which were all close to 1.

Ethical considerations

The overall protocol for this study was submitted to the ethics committee of ISTM/Kinshasa and received authorization No. 0036/CBE/ISTM/KIN/RDC/PMBBL/2025 on February 12, 2026. Additionally, the data underwent an anonymization protocol and were then processed according to standardized operating procedures to preserve confidentiality, maintain source integrity, and ensure the reproducibility of the analyses. The analytical basis was established without any alteration of the original clinical information.

Results

Sample description

Patient identification and location

Table 1 shows that the majority of patients whose records were analyzed were under 50 years old (43.2%), female (71.6%), had a higher education level (28.4%), were married (44.4%), and lived in the Lukunga district (32.1%).

Of the quality of the data, the level of education (66.7%), marital status (42.0), and province of origin (96.3%) were the main elements of patient identification and location with the most missing data.

Table 1: Proportion of missing data related to patient identification and geographic location

Variable	Data entered		Missing data	
	Effective	%	Effective	%
ID (patient name)	81	100	0	0.0
Age (in completed years)				
< 50	35	43.2	3	3.7
50-64	26	32.1		
≥ 65	17	21.0		
Sex				
Female	58	71.6	0	0.0
Male	23	28.4		
Level of education				
Secondary	4	4.9	54	66.7
Superior	23	28.4		
Marital status				
Bachelor	4	4.9	34	42.0
Divorce	1	1.2		
Bride)	36	44.4		
Widower/ widow	6	7.4		
Province of origin				
Kinshasa	3	3.7	78	96.3
Address/District				
Funa	21	25.9	5	6.2
Lukunga	26	32.1		
Mount Amba	21	25.9		
Tshangu	8	9.9		

https://doi.org/10.51168/x488nx46

Medical history, behavior, and comorbidities

Table 2 shows that many of the patients whose records were used had no family history of cancer (48.1%) or obesity (75.3%), had never received hormone therapy (38.3%), had a normal body weight (21.0%), reported never having smoked (53.1%) or consumed alcohol (28.4%), with a negative serological status for HIV/AIDS (11.1%), Hepatitis B or C (12.3%) and then HPV (12.3%).

Regarding data recording, a significant proportion of missing information was found concerning diet, HIV status, Hepatitis B/C, HPV infection, BMI, and prior hormone therapy, which were undocumented for more than half of the patients. Similarly, more than a third of the records lacked information regarding family history of cancer, smoking, and alcohol use.

Table 2: Proportion of missing data related to history, behavior, and comorbidities

Variables	Data entered Effective	%	Missing data Effective	%
Family history of cancer				
No	39	48.1	28	34.6
Yes	14	17.3		
Cases of obesity in families				
No	61	75.3	18	22.2
Yes	2	2.5		
Previous hormone therapy				
No	31	38.3	42	51.9
Yes	8	9.9		
BMI (in kg/m²)				
< 18.5	3	3.7	45	55.6
18.5 – 24.9	17	21.0		
25.0 – 29.9	7	8.6		
≥ 30.0	9	11.1		
Smoking				
Active smoker	1	1.2	36	44.4
Former smoker	1	1.2		
Never	43	53.1		
Alcohol consumption				
None	23	28.4	40	49.4
Casual	17	21.0		
Regular	1	1.2		
Diet				
Rich in salt/smoky	1	1.2	80	98.8
HIV/AIDS Status				
Negative	9	11.1	71	87.7
Positive	1	1.2		
Hepatitis (B or C)				
Negative	10	12.3	71	87.7
HPV infection (cervix)				
Negative	10	12.3	71	87.7

https://doi.org/10.51168/x488nx46

Disease characteristics (tumor variables)

The results in Table 3 describe the characteristics of cancer among the patients whose records were analyzed. They show that the majority were diagnosed in 2025 (46.9%), based on clinical findings combined with radiography (51.9%). Isolated breast cancer accounted for 34.6% of all diagnoses. Unilateral and single-organ involvement were predominant (82.7%). Carcinoma was the most common type of cancer, both in terms of morphology (80.2%) and histological type (58.0%). Finally, many of the diagnosed cancers were invasive (66.7%) and stage II (30.9%).

Regarding the quality of documentation, the evolutionary stage at the time of diagnosis shows the highest proportion of missing data (59.3%), followed by the degree of impairment (28.3%).

Table 3: Proportion of missing data related to disease characteristics

Variables	Data entered Effective	%	Missing data Effective	%
Date of diagnosis/Period				
2023	8	9.9	0	0.0
2024	35	43.2		
2025	38	46.9		
Diagnostic basis				
Clinical/X-ray	42	51.9	2	2.5
Cyto -histology	37	45.7		
Topography				
Other organs	38	46.9	0	0.0
Prostate	2	2.5		
Breast	28	34.6		
Gynecological sphere	13	16.0		
Laterality				
Bilateral	3	3.7	11	13.6
Unilateral/single organ	67	82.7		
Morphology				
Carcinoma	65	80.2	7	8.6
Leukemia	1	1.2		
Sarcomas	8	9.9		
Degree of impact				
Invasive	54	66.7	23	28.4
In situ	4	4.9		
Evolutionary stage				
Stage I	3	3.7	48	59.3
Stage II	25	30.9		
Stage III	5	6.2		
histological type				
Other types	14	17.3	20	24.7
Carcinoma	47	58.0		

Treatment and follow-up

Regarding patient management, Table 4 shows that more than half of the cancers were diagnosed more than three months after the onset of symptoms (55.6%), and the initial treatment was non-medical (54.3%). In the hospital, the predominant treatment regimen was chemotherapy alone (27.2%) or combined with radiotherapy (24.7%). More than half of these patients were lost to follow-up.

Regarding the quality of information, the monitoring method and the treatment regimen were very poorly documented, with 46.9% and 98.8% of data missing, respectively.

Table 4: Proportion of missing data related to Treatment and follow-up

Variables	Data entered Effective	%	Missing data Effective	%
Consultation timeframe (in months)				
≤ 1	11	13.6	1	1.2
2 - 3	24	29.6		
≥ 4 months	45	55.6		
Initial therapeutic approach				
Others	44	54.3	12	14.8
Modern medicine	25	30.9		
Treatment plan				
Chemotherapy alone	22	27.2	38	46.9
Chemotherapy & Surgery	1	1.2		
Chemotherapy & UV	20	24.7		
Patient's Outcome				
Continue treatment	9	11.1	1	1.2
Deceased	29	35.8		
Lost sight of	42	51.9		
Monitoring method used				
Phone	1	1.2	80	98.8

Missing value analysis

The results of the Little test (MCAR) presented in Table 5 indicate that the missing data mechanism is completely random ($\chi^2 = 1.830$; $p = 0.767$). This gives mathematical validity to the final analysis based on complete records in the event of the completeness modeling presented in point 3.4.

Table 5: Analysis of missing data

continuous variables	Data entered	Missing data	MCAR/χ^2	ddl	p-value
Consultation period	80	1 (1,2)	1,830	4	0.767
Age at diagnosis	78	3 (3.7)			
BMI	36	45 (55.6)			

Comparative analysis of the completeness score of cancer patients' medical records

Figure 1 presents the results of the comparison of the fill/completeness scores of individual information of cancer patients (Friedman Test) and shows a significant difference between the five record domains studied ($p<0.001$).

Furthermore, comparative pairwise analyses (Durbin - Conover) revealed seven notable disparities. The level/domain "Medical History/Behaviors/Comorbidities (c)" had a lower completeness score compared to the other categories (a,x; b,y; b,y,z) examined ($p < 0.001$). Moreover, the "Geographic Location" dimension showed significantly higher completeness compared to "Patient Identification" ($p < 0.001$) and "Management and Follow-up" ($p = 0.037$). Finally, "Patient Identification" was recorded less frequently than "Disease Information" ($p < 0.001$).

https://doi.org/10.51168/x488nx46
Original Article

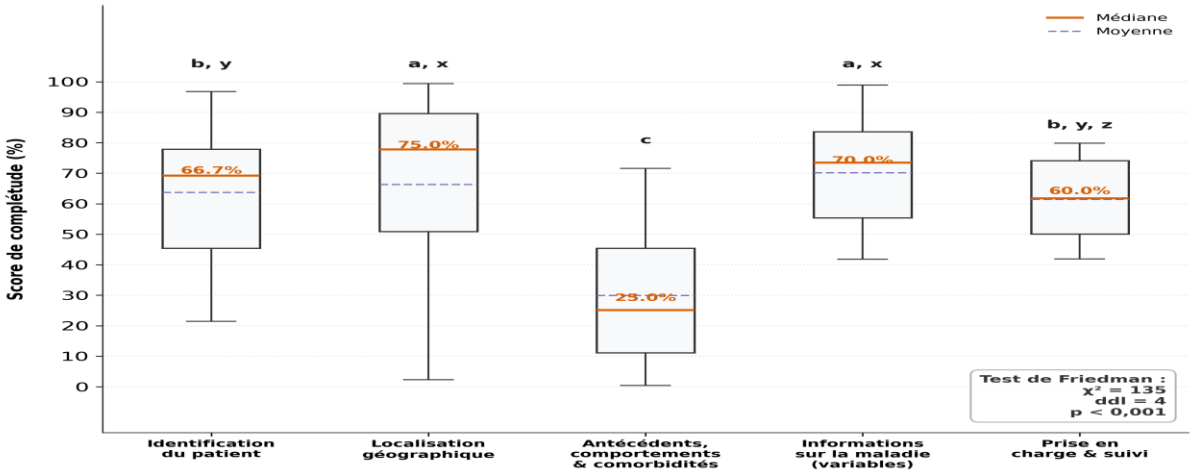

Figure 1: Comparative analysis of medical record completeness scores

Modeling the overall file completeness score

Identification of predictors of medical record completeness

The results in Table 6 indicate that the overall completeness score of the assessed medical records decreases significantly according to age at cancer diagnosis (β = -0.174; p = 0.016); bilateral organ involvement (β = -7.349; p = 0.029); year of detection 2024 vs 2025 (β = -10.732; p<0.001); non-invasive nature of the disease (β = -5.893; p = 0.047); family history of cancer (β = -4.491; p = 0.028); of residence in the Funa district versus Lukunga (β = -4.547; p = 0.017) and of initial traditional treatment/prayer (β = -5.822 ; p = 0.025). However, this score increases with the time elapsed before consultation (β = 0.476 ; p < 0.001) and then with residence in the Tshangu district compared to Lukunga (β = 9.236 ; p = 0.009).

Table 6: Modeling of the overall file completeness score

Predictor	Estimate (β)	95% confidence		p-value
		Terminal inf	Superior	
Originally ordered[a]	64.439	56,573	72.3046	<.001***
Age at diagnosis (in years)	-0.174	-0.311	-0.0368	0.016*
Consultation period	0.476	0.320	0.6323	<.001***
Laterality:				
Bilateral vs Unilateral	-7,349	-13,830	-0.8678	0.029*
Year:				
2023 vs 2025	-3.737	-8,784	1.3110	0.136
2024 vs 2025	-10.732	-14,581	-6.8825	<.001***
Degree of impact:				
In situ vs Invasive	-5.893	-11,686	-0.0994	0.047*
Family history of cancer:				
Yes vs No	-4.491	-8.424	-0.5594	0.028*
Municipality of residence:				
Mount Amba vs Lukunga	0.784	-3.073	4.6398	0.672
Tshangu – Lukunga	9.236	2.685	15.7867	0.009**
Funa vs Lukunga	-4.573	-8.223	-0.9223	0.017*
Initial therapeutic approach:				
Traditional/prayer vs Modern	-5.822	-10.796	-0.8486	0.025*
Integrated PEC vs Modern	-0.536	-4.479	3.4065	0.777

[a] Represents the overall average
** Significant difference at the 0.05 level; ** Significant difference at the 0.01 level; Significant difference at the 0.001 level.*

Quality of fit and model validation diagnostics

Table 7a describes the goodness of fit and validation diagnostics of the model. The results demonstrate a well-fitted model explaining 83.9% of the variance in the overall score of completeness of medical records of cancer patients (R^2 = 0.839); the independence or absence of residual autocorrelation (DW = 1.980; p = 0.908) and the normality (Shapiro- Wilk = 0.977; p = 0.766) of the residuals, and the absence of outliers (D-Cook = 0.449) and influential values (D- Mahalanobis = 11.40).

Table 7a: Fit quality and model validation diagnostics

Validation parameters	Value	p-value
Overall adjustment		
Correlation coefficient (R)	0.916	
Coefficient of determination (R^2)	0.839	
Residue analysis		
Durbin -Watson independence test	1,980	0.908
Wilk normality test	0.977	0.766
Detection of outliers and influential values		
Maximum Cook Distance	0.449	
Maximum Mahalanobis distance	11.40	

Furthermore, the results presented in Table 7b indicate the absence of collinearity between the variables. The VIFs range from 1.19 to 1.46, and the tolerances range from 0.69 to 0.84.

Table 7b: Collinearity Statistics

Factors	VIF	Tolerance
Age at diagnosis (in years)	1.19	0.838
Consultation period	1.42	0.704
Laterality	1.28	0.782
Year	1.24	0.805
Cancer behavior/degree of involvement	1.37	0.728
Family history of cancer	1.46	0.687
Municipality of residence	1.21	0.828
Initial therapeutic approach	1.27	0.790

Discussion

This section discusses the key findings of the present study, focusing on (i) the sociodemographic and clinical profile of the patients whose records were reviewed, (ii) the mechanisms of missing data and the validity of the analyses, (iii) the variability of completeness scores by domain, and (iv) the systemic and clinical determinants of the overall completeness score. The section concludes with the implications of the study, followed by its strengths and limitations.

Sociodemographic and clinical profile of patients

The results of this study, indicating a predominance of women and patients under 50 years of age, mostly living in the Lukunga district with cancer diagnosed at an invasive stage (stage II), illustrate the weight of this pathology in the Oncology department of the Ngaliema clinic.

This clinical picture, both young and severe, agrees with the epidemiological results of Dusingize et al (2026) in DRC (Dusingize et al., 2026) and those of Anyigba et al. (2021) in PMC, which describe the evolution of the incidence as well as the hormonal and reproductive specificities of patients in sub-Saharan Africa (Anyigba et al., 2021) .

Internationally, this late and invasive detention contrasts sharply with the situation in high-income countries, where structured screening and awareness campaigns allow for the identification of the disease at localized or in situ stages (Benitez Fuentes et al., 2024) . This marked disparity, largely due to the absence of systematic mammography programs, is corroborated by Kamalakanta 's meta-analysis. Gahan in 2024, which highlights major gaps in the completeness of imaging and staging data between registries (Kamalakanta Gahan, 2024).

The gaps observed in the present study concerning the level of education and marital status at the time of identification would be explained by systemic failures as presented in the various studies in Sub-Saharan Africa, recalling that the fragility of administrative infrastructures and the shortage of committed staff strongly limit the completeness of hospital oncology databases (Ngwa et al., 2022; WHO, 2023) .

The presence of unrecorded data in the patient records mentioned above can hinder advanced statistical analyses if

it is not accidental. Therefore, the MCAR hypothesis is verified before addressing disparities in data completeness by domain or recording level.

Mechanisms of missing data and validity of analyses

Little's test, demonstrating a completely random missing data mechanism, statistically justifies the exclusion of incomplete records from the final modeling of the completeness score. This methodological requirement eliminates selection bias and strengthens the legitimacy of decision-making analyses.

These results are promising and differ from those reported by other authors who found that incomplete medical records in oncology are most often not random, resulting from recurrent clinical negligence or structural losses of follow-up related to the cost of examinations (Habib Ullah) . Riaz et al., 2025; Gachau et al., 2020) .

According to international recommendations, rigorous statistical validation and transparent management of missing clinical data are essential to harmonize survival indicators, a major methodological challenge supported by recommendations for breast cancer registries (Dahm et al., 2023; “Correcting bias due to missing stage data in the non-parametric estimation of stage-specific net survival for colorectal cancer using multiple imputation,” 2017) , particularly the Global Stage multicenter study. Distributi o n de 2023 (Jama , 2023) .

That being said, in the following paragraphs, we proceed to analyze the variability of the completeness scores of medical records of patients with cancer according to the different levels or areas of information recording.

Variability of completeness scores by data recording field

The highly significant difference observed between the five recording fields ($p<0.001$) indicates inconsistent completeness of records for the same patient. The subscore related to history, behavior, and comorbidities (HIV, HPV, BMI) is the least detailed. This could be explained by the fact that, in practice, clinical urgency takes precedence over systematic behavioral questioning.

The absence of a structured collection of risk factors, including smoking, alcohol consumption (Jeandria) Nkodia et al., n.d.), and obesity, constitutes a recurring weakness in Central Africa, as indicated by the World Health Organization. Recent studies confirm this observation, showing that workload overload and a shortage of well-trained oncologists lead to lifestyle variables deemed non-immediate being relegated to the background (Habib Ullah) . Riaz et al., 2025) . Conversely, better completeness of the variable "Geographic location" would decouple administrative constraints inherent in hospital admission.

In view of this variability, it is essential to analyze the determinants of the overall completeness score of these patients' records.

Systemic and clinical determinants of the overall completeness score

The decrease in the overall completeness score in elderly patients, those with bilateral involvement, or those using traditional therapies illustrates the direct effect of clinical and cultural barriers on traceability. Faced with invasive bilateral pathology, medical teams would accelerate the initiation of treatment at the expense of exhaustive data entry, a phenomenon of document exhaustion reported at the regional level by Ngwa et al . (2022) and then Gbenonsi et al . (2024) .

Furthermore, the massive initial recourse to alternative medicines in the sample (54.3%) disrupts the temporal continuity of the care pathway. This therapeutic nomadism, both spiritual and traditional, results in late institutional records, lacking pre-diagnostic history, a problem recently analyzed and published, where the authors show that the use of herbal remedies and certain spiritual beliefs induce major diagnostic delays and the absence of initial clinical data (Asiimwe et al., 2023; Henke et al., 2022) .

The paradoxical increase in score with longer consultation delays could be explained by the gradual, often delayed, consolidation of assessments carried out outside the institution. To mitigate these temporal effects and smooth inter-annual variations (2024 vs. 2025), Gichoya et al . (2025), in Lancet Digital Health, recommend the adoption of suitable computerized hospital registries to establish a sustainable, robust governance of oncological data (Gichoya et al., 2025) .

Conclusion

The results indicate that the completeness of cancer records at the Ngaliema Clinic is dependent on clinical severity, patient age, and treatment pathway. The MCAR hypothesis is consistent with the observations, thus strengthening the validity of our modeling framework. This work provides the baseline diagnosis necessary for the development of the forthcoming thesis dedicated to the first pilot hospital cancer registry in the DRC in general and in Kinshasa in particular.

Recommendations

In light of the results, it is imperative to implement, upon patient admission, a standardized minimum oncological form to systematically record a comprehensive medical history, including comorbidities, past medical history, and BMI. Hospital administrations in Kinshasa should prioritize ongoing digital skills training for staff and the ergonomic optimization of data collection tools. Such measures will mitigate the impact of workload overload on document quality, particularly during clinical emergencies.

Furthermore, the Ministry of Health would benefit from integrating spiritual leaders and traditional healers into early referral networks to document symptom history and reduce waiting times for consultations. In addition, the government should provide financial support for the transition to electronic health records. Robust digital governance is

essential to ensure the sustainability of the future pilot cancer registry in the face of changing times.

Strengths and limitations of the study

The main strength of this work lies in the robustness of its statistical validation. The application of Little's test confirms that the pattern of missing data stems from a completely random mechanism. This characteristic establishes, for the final regression modeling, conditions of mathematical validity that are rarely met. Furthermore, the integration of observations within the operational context of Kinshasa, by visiting a tertiary-level healthcare facility, provides a unique empirical diagnostic reference for the region.

The major limitation stems from the retrospective and monocentric nature of the data collection, which restricts the external scope of the results to the provincial and national levels. Furthermore, the exclusion of incomplete records mechanically reduced the statistical power of the sample. Finally, the single data collection could increase the risk of transcription errors and the possibility of comparison with other transcriptions.

Implications of the study

Immediate clinical and managerial implications

Clinically, the completeness and accuracy of patient records are closely linked to tumor aggressiveness and the clinical workload of medical teams. Therefore, it is essential to emphasize the systematic inclusion of family history and major comorbidities (HIV, HPV, body mass index) by clinicians—dimensions that are too often omitted today, even though they are crucial for tailoring treatment plans and risk stratification.

Implications for public health policies

The high proportion of people resorting to traditional or spiritual practices (54.3%) is a warning sign for the health authorities in Kinshasa. It is imperative to design, without delay, community awareness campaigns that involve religious authorities and traditional practitioners in a structured way. The aim is to formalize early referral pathways to modern facilities, reduce waiting times for care, and simultaneously document patients' pre-hospital medical histories in a standardized manner, while also maintaining the interdisciplinary approach that remains essential for chronically ill patients.

Doctoral perspectives: towards a pilot registry in Kinshasa

This thesis establishes an essential baseline for a doctoral program. The completely random missing data hypothesis (MCAR) confirms the statistical viability of exploiting current hospital series. This work provides a detailed map of the variables and admission pathways that require a strict quality assurance system during subsequent collections.

The central objective of the thesis will be the design and implementation of the first pilot hospital cancer registry in Kinshasa. The predictive models already developed highlight high-risk profiles (elderly subjects, bilateral involvement, residents of Funa) for which active case finding (active case-finding) will be essential.

This register will constitute the first structuring step of a digital governance of oncology in the Democratic Republic of Congo.

Acknowledgments

The authors thank all the administrative staff of the Ngaliema Clinic in general and those of the oncology department in particular, for making the medical records available to them for analysis.

Abbreviations

BCZS	**:** Central Health Zone Office
CSR	**:** Referral Health Center
EMR	**:** Electronic Medical Record
DPS	**:** Provincial Health Division
FOSA	**:** Health Training
GLOBOCAN	**:** Global Cancer Observatory
HGR	**:** General Referral Hospital
IARC	**:** International Agency for Research on Cancer
IC:	: Confidence Interval
ISTM	**:** Higher Institute of Medical Techniques
MAR	**:** Missing At Random
MCAR	: Missing Completely At Random
MNAR	**:** Missing Not At Random
WHO	**:** World Health Organization
NGO	**:** Non-Governmental Organization
p-value	**:** probability value
R^2	**:** determination coefficient
RC	**:** Cancer registry
DRC	**:** Democratic Republic of Congo
RPCK	**:** Kinshasa piloted cancer register
SNIS	**:** National Health Information System
WHO	**:** World Health Organization
ZS	**:** Health Zone
β	**:** regression coefficient
χ^2	**:** chi-square statistic

Conflicts of interest

The authors state that they have no competing interests.

Funding statement

This study did not receive any specific funding from any organization or third party.

Author contributions

AMP designed the study, conducted the formal analyses, and drafted the initial manuscript. KBH and OOJP participated in data collection and contributed to manuscript writing. KKC participated in data collection and analysis. LMJ proposes the methods to be used. IIF and TKF were responsible for supervision and substantial revision of the study manuscript.

All authors critically reviewed and approved the final manuscript and accepted full responsibility for the integrity of the work.

Data availability

Data is available upon request.

Author biographies

Atuba Mamenepi Pascal is a trainer of advanced-level nurses (BAC+5). Since 2020, he has been teaching at the Faculty of Nursing at the Higher Institute of Medical Techniques in Kinshasa. He is currently pursuing research in Biostatistics and Health Informatics at the Doctoral School of Health Sciences in Kinshasa. He reviews a few articles in the Asian Journal of Research in Infectious Diseases and BP International.

Kalubi Belanga Heir is a registered nurse (Bachelor's degree + 3 years). He is currently pursuing a Master's degree in Nursing and is a member of the NGO Pallia Familli, which provides palliative care in the homes of patients in need in the city of Kinshasa.

Omokende Olonge Jean-Pierre is a nurse specializing in anesthesiology and resuscitation. He practices as an anesthesiologist at the Makala General Reference Hospital in Kinshasa and teaches full-time at the Higher Institute of Medical Techniques of Kisantu in the Kongo Central province of the Democratic Republic of Congo.

Kakwaka Katombe Crispin is a demographer by training. He regularly managed the Hospital Statistics department in Kinshasa and other provinces of the DRC. Currently, he teaches courses related to his experience at the Higher Institute of Medical Techniques in Kinshasa and at other medical training institutions in the city.

Lofandjola Masumbuku Jacques holds a PhD in Public Health. He teaches at the Higher Institute of Medical Techniques in Kinshasa and the University of the DRC. He is currently the research officer for the National Association for Palliative Care of the Democratic Republic of Congo. Editorial member of the African Annals of Medicine, journal of the Faculty of Medicine, University of Kinshasa.

Ilunga Ilunga Félicien holds a PhD in Public Health, specializing in Epidemiology, Biostatistics, Clinical Research, and Applied Computer Science. He teaches at the Higher Institute of Medical Techniques in Kinshasa and other higher education institutions and universities in the DRC. From 2005 to 2014, he was responsible for public health projects and served as the academic coordinator for university programs associated with the Higher Institute of Medical Techniques in Kinshasa and the Free University of Brussels. He was an editorial member of the African Annals of Medicine, a journal of the Faculty of Medicine at the University of Kinshasa, in 2018.

Tshimungu Kandolo Félicien holds a Ph.D. in Public Health. He teaches at the Higher Institute of Medical Techniques in Kinshasa and other universities in the DRC. He was Dean of the Faculty of Nursing, where he significantly contributed to the reform of the education system, transitioning to the LMD (Bachelor's, Master's, Doctorate) system. He was a member of the thesis monitoring and support committee and a member of the thesis examination board at the Erasme Doctoral School, Research Chair in Nursing Sciences, University of Paris 13, Bobigny-Université, Paris, France, from 2015 to 2019.

References

1. Anyigba, CA, Awandare, GA, & Paemka, L. (2021). Breast cancer in sub-Saharan Africa: The current state and uncertain future. Experimental Biology and Medicine, 246 (12), 1377-1387. https://doi.org/10.1177/15353702211006047 PMid:33926257 PMCid:PMC8243219
2. Asiimwe , JB, Nagendrappa , PB, Atukunda , EC, Nambozi , G., Tolo , CU, Ogwang , PE, & Kamatenesi , MM (2023). The meaning of caring for patients with cancer among traditional medicine practitioners in Uganda: A grounded theory approach. PLOS Global Public Health, 3 (7), e0001764. https://doi.org/10.1371/journal.pgph.0001764 PMid:37459297 PMCid:PMC10351711
3. Ayubi, E., Lyus, R., Brhlikova, P., & Pollock, A.M. (2024). A critical appraisal of the quality of data submitted by sub-Saharan African cancer registries to GLOBOCAN 2020. JRSM Open, 15 (1), 20542704231217888. https://doi.org/10.1177/20542704231217888 PMid:38223747 PMCid:PMC10785726
4. Benitez Fuentes, J.D., Morgan, E., de Luna Aguilar, A., Mafra , A., Shah, R., Giusti , F., Vignat , J., Znaor , A., Musetti , C., Yip, C.-H., Van Eycken , L., Jedy-Agba , E., Piñeros , M., & Soerjomataram , I. (2024). Global Stage Distribution of Breast Cancer at Diagnosis. JAMA Oncology, 10 (1), 71-78. https://doi.org/10.1001/jamaoncol.2023.4837 PMid:37943547 PMCid:PMC10636649
5. Bhaskaran, K., & Smeeth, L. (2014). What is the difference between missing completely at random and missing at random? International Journal of Epidemiology, 43 (4), 1336-1339. https://doi.org/10.1093/ije/dyu080 PMid:24706730 PMCid:PMC4121561
6. Bray, F., Laversanne , M., Sung, H., Ferlay , J., Siegel, R. L., Soerjomataram , I., & Jemal , A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 74 (3), 229-263.

https://doi.org/10.3322/caac.21834
PMid:38572751

7. Bray, F., Parkin, DM, Gnangnon , F., Tshisimogo , G., Peko, J.-F., Adoubi , I., Assefa , M., Bojang , L., Awuah , B., Koulibaly , M., Buziba , N., Korir , A., Dzamalala , C., Kamate , B., Manraj , S., Ferro, J., Lorenzoni , C., Hansen, R., Nouhou , H., … Chingonzoh , T. (2022). Cancer in sub-Saharan Africa in 2020: A review of current estimates of the national burden, data gaps, and future needs. The Lancet Oncology, 23 (6), 719-728. https://doi.org/10.1016/S1470-2045(22)00270-4 PMid:35550275
8. Bray, F., Znaor , A., Cueva, P., Korir , A., Swaminathan , R., Ullrich , A., Wang, SA, & Parkin, DM (2014). Planning and Developing Population-Based Cancer Registration in Low- or Middle-Income Settings. International Agency for Research on Cancer. http://www.ncbi.nlm.nih.gov/books/NBK566957/
9. Correcting bias due to missing stage data in the non-parametric estimation of stage-specific net survival for colorectal cancer using multiple imputation. (2017). Cancer Epidemiology, 48, 16-21. https://doi.org/10.1016/j.canep.2017.02.005 PMid:28315607
10. Cro, S., Morris, T.P., Kenward, M.G., & Carpenter, J.R. (2020). Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: A practical guide. Statistics in Medicine, 39 (21), 2815-2842. https://doi.org/10.1002/sim.8569 PMid:32419182
11. Dahm, S., Barnes, B., & Kraywinkel, K. (2023). Detection of missed deaths in cancer registry data to reduce bias in long-term survival estimation. Frontiers in Oncology, 13, 1088657. https://doi.org/10.3389/fonc.2023.1088657 PMid:36969013 PMCid:PMC10034313
12. Dusingize , JC, Zotova , N., Kabarriti , R., Sehrawat , K., Babakazo , P., Alisho , AS, Kasindi , FL, Yessoufou , I., & Yotebieng , M. (2026). Determinants of cancer care delays in Kinshasa, Democratic Republic of the Congo (DRC). Oncology. https://doi.org/10.64898/2026.05.19.26353550
13. Gachau , S., Quartagno , M., Njagi , E.N., Owuor , N., English, M., & Ayieko , P. (2020). Handling missing data in modeling quality of clinician-prescribed routine care: Sensitivity analysis of departure from the Missing at Random (MAR) assumption. Statistical methods in medical research, 29 (10), 3076-3092. https://doi.org/10.1177/0962280220918279
14. Gbenonsi , G.Y., Martini, J., & Mahieu , C. (2024). An analytical framework for breast cancer public policies in Sub-Saharan Africa: Results from a comprehensive literature review and an adapted policy Delphi. BMC Public Health, 24 (1), 1535. https://doi.org/10.1186/s12889-024-18937-5 https://doi.org/10.21203/rs.3.rs-3142436/v1
15. Gichoya , J. W., Mwavu , R., Minja , F., Kaonga , N., Purkayastha , S., & Newsome, J. (2025). Leveraging digital technologies to reduce cancer disparities in low- and middle-income countries. The Lancet Digital Health, 7 (10). https://doi.org/10.1016/j.landig.2025.100937 PMid:41241583
16. Habib Ullah Riaz, Faiza Munir, Nimra Khalid, Tasleem Bibi, Anza Saleem, Afaq Khizer, & Muhammad Kamran. (2025). Delays in Breast Cancer Care: The Roles of Nurse Navigation in Facilitating Timely Treatment at a Tertiary Care Hospital, Lahore. Journal of Health, Wellness and Community Research, e839. https://doi.org/10.61919/j2dn4b60
17. Henke, O., Bruchhausen , W., & Massawe , A. (2022). Use of Herbal Medicine Is Associated With Late-Stage Presentation in Tanzanian Patients With Cancer: A Survey to Assess the Utilization of and Reasons for the Use of Herbal Medicine. JCO Global Oncology, 8, e2200069. https://doi.org/10.1200/GO.22.00069 PMid:35820080 PMCid:PMC9296184
18. Jakobsen , JC, Gluud , C., Wetterslev , J., & Winkel , P. (2017). When and how should multiple imputation be used for handling missing data in randomized clinical trials - a practical guide with flowcharts. BMC Medical Research Methodology, 17 (1), 162. https://doi.org/10.1186/s12874-017-0442-1 PMid:29207961 PMCid:PMC5717805
19. Jama. (2023). Global Stage Distribution of Breast Cancer at Diagnosis. https://jamanetwork.com/journals/jamaoncology/fullarticle/2811796
20. Jeandria Nkodia, A., Ewuinh Tsiobinda, B., Jesse, A., & Missongo, SR (n.d.). (PDF) Prevalence and factors associated with smoking among adults in the Moungali district of Brazzaville in 2021. Medical Journal Pan-African, 43 (6). https://doi.org/10.11604/pamj.2022.43.6.34180
21. Jena, S., Sahoo, K. C., Samantaray, K., Satpathy, N., & Epari, V. (2023). Operational Feasibility of Hospital-Based Cancer Registries in Low- and Middle-Income Countries: A Systematic Review. Cureus , 15 (7), e42126. https://doi.org/10.7759/cureus.42126
22. Jensen, O., Parkin, D., MacLennan, R., Muir, C., & Skeet, R. (1991). Cancer Registration: Principles and Methods. https://publications.iarc.who.int/Book-And-Report-Series/Iarc-Scientific-Publications/Cancer-Registration-Principles-And-Methods-1991

23. Kamalakanta Gahan. (2024). Disparities in Breast Cancer Screening, Diagnosis, and Outcomes in Low-and Middle-Income Countries: A Systematic Review and Meta-Analysis. https://doi.org/10.13140/RG.2.2.22853.46565
24. Ngwa, W., Addai, BW, Adewole, I., Ainsworth, V., Alaro, J., Alatise, OI, Ali, Z., Anderson, BO, Anorlu, R., Avery, S., Barango, P., Bih, N., Booth, CM, Brawley, OW, Dangou, J.-M., Denny, L., Dent, J., Elmore, SNC, Elzawawy , A., … Kerr, D. (2022). Cancer in sub-Saharan Africa: A Lancet Oncology Commission. The Lancet Oncology, 23 (6), e251-e312. https://doi.org/10.1016/S1470-2045(21)00720-8 PMid:35550267 PMCid:PMC9393090
25. Parkin, D. M. (2006). The evolution of the population-based cancer registry. Nature Reviews Cancer, 6 (8), 603-612. https://doi.org/10.1038/nrc1948 PMid:16862191
26. Sung, H., Ferlay , J., Siegel, R.L., Laversanne , M., Soerjomataram , I., Jemal , A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians, 71 (3), 209-249. https://doi.org/10.3322/caac.21660 PMid:33538338
27. WHO. (2019). International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3). https://www.who.int/standards/classifications/other-classifications/international-classification-of-diseases-for-oncology
28. WHO. (2023). The global initiative for cancer registry development. The Global Initiative for Cancer Registry Development. https://gicr.iarc.fr/about-the-gicr/

Publisher details

SJC PUBLISHERS COMPANY LIMITED

SJC Publishers Company Limited
TODAY AND TOMORROW

Catergory: Non Government & Non profit Organisation
Contact: +256 775 434 261 (WhatsApp)
Email: info@sjpublisher.org or studentsjournal2020@gmail.com
Website: https://sjpublisher.org
Location: Scholar's Summit Nakigalala, P. O. Box 701432,Entebbe Uganda, East Africa