Check for updates

RESEARCH ARTICLE

REVISED   **Cancer in Kenya: types and infection-attributable. Data from the adult population of two National referral hospitals (2008-2012) [version 5; peer review: 2 approved, 2 not approved]**

Previously titled: Cancer in Kenya: types and infection-attributable. Data from two National referral hospitals

Lucy Wanjiku Macharia [iD] [1], Marianne Wanjiru Mureithi [iD] [1,2], Omu Anzala [1,2]

[1]Department of medical Microbiology, Faculty of Medicine, University of Nairobi, Nairobi, Kenya
[2]KAVI-Institute of Clinical Research (KAVI-ICR), College of Health Sciences, University of Nairobi, Nairobi, Kenya

Abstract

Background: Cancer in Africa is an emerging health problem. In Kenya it ranks third as a cause of death after infectious and cardiovascular diseases. Nearly 31% of the total cancer burden in sub-Saharan Africa is attributable to infectious agents. Information on cancer burden is scanty in Kenya and this study aimed to provide comprehensive hospital based data to inform policies.

Method: A cross-sectional retrospective survey was conducted at Kenyatta National Hospital (KNH) and Moi Teaching and Referral Hospital (MTRH) from January 2008 to December 2012. Data was obtained from the patients files and the study was approved by the KNH/University of Nairobi and MTRH Ethics and Research Committees.

Results: In KNH, the top five cancers were: cervical (62, 12.4%), breast (59, 11.8%), colorectal (31, 6.2%), chronic leukemia (27, 5.4%) and stomach cancer (26, 5.2%). Some 154 (30.8%) of these cancers were associated with infectious agents, while an estimated 138 (27.6%) were attributable to infections. Cancers of the cervix (62, 12.4%), stomach (26, 5.2%) and nasopharynx (17, 3.4%) were the commonest infection-associated cancers. In MTRH, the five common types of cancers were Kaposi's sarcoma (93, 18.6%), breast (77, 15.4%), cervical (41, 8.2%), non-Hodgkin's lymphoma (37, 7.4%) and colorectal, chronic leukemia and esophageal cancer all with 27 (5.4%). Some 241 (48.2%) of these cancers were associated with infectious agents, while an estimated 222 (44.4%) were attributable to infections. Kaposi's sarcoma (93, 18.6%), cancer of the cervix (41, 8.2%) and non-Hodgkin's lymphoma (37, 7.4%) were the commonest infection-associated cancers.

Conclusion: Our results suggest that 30.8% and 48.2% of the total cancer

Open Peer Review

Reviewer Status ✗ ✗ ✓ ✓

	Invited Reviewers			
	1	2	3	4
REVISED **version 5** published 14 Nov 2019			✓ report	
REVISED **version 4** published 03 Oct 2019			? report	
REVISED **version 3** published 04 Mar 2019			? report	✓ report
REVISED **version 2** published 19 Nov 2018		✗ report		
version 1 published 25 Sep 2018	✗ report			

1   **Clement A Adebamowo** [iD] , University of Maryland School of Medicine, Baltimore, USA

Institute of Human Virology Nigeria, Garki, Nigeria

Center for Bioethics and Research, Ibadan,

cases sampled in KNH and MTRH respectively were associated with infectious agents, while 27.6% and 44.4% were attributable to infections in the two hospitals respectively. Reducing the burden of infection-attributable cancers can translate to a reduction of the overall cancer burden.

Keywords
Cancer burden, infectious agents, sub-Saharan Africa, Kenya

Nigeria

Michael Odutola, Institute of Human Virology Nigeria, Garki, Nigeria

2   **Maria Paula Curado**, Federal University of Goiá´s, Goiânia, Brazil

3   **Delphine Casabonne**, Catalan Institute of Oncology, Barcelona, Spain Epidemiology and Public Health Networking Biomedical Research Centre (CIBERESP), Madrid, Spain

4   **Kalina Duncan** [iD], U.S. National Cancer Institute, Rockville, USA

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Lucy Wanjiku Macharia (macharialw@gmail.com)

Author roles: Macharia LW: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Mureithi MW**: Conceptualization, Funding Acquisition, Methodology, Project Administration, Resources, Supervision, Validation, Visualization, Writing – Review & Editing; **Anzala O**: Conceptualization, Funding Acquisition, Methodology, Resources, Supervision, Validation, Visualization, Writing – Review & Editing

Background

Cancer in Africa is an emerging health problem where about 847,000 new cancer cases and 591,000 deaths occurred in 2012, with about three quarters of these occurring in the sub-Saharan region[1]. In Kenya, cancer ranks third as a cause of death, after infectious and cardiovascular diseases, and in 2012 there was an estimated 37,000 new cancer cases, and 28,500 cancer deaths reported[2]. Infectious agents are an important cause of cancer, particularly in less developed countries. According to the International Agency for Research on Cancer (IARC), 11 infectious agents have been classified and established as carcinogenic agents in humans namely: *Helicobacter pylori*, hepatitis B virus (HBV), hepatitis C virus (HCV), human immunodeficiency virus type 1 (HIV-1), human papillomavirus (HPV), Epstein-Barr virus (EBV), human herpes virus type 8 (HHV-8; also known as Kaposi's sarcoma herpes virus), human T-cell lymphotropic virus type 1 (HTLV-1), *Opisthorchis viverrini*, *Clonorchis sinensis*, and *Schistosoma haematobium*[3].

Nearly 31% of the total cancer burden in sub-Saharan Africa is attributable to infections[4]. Specifically, *H. pylori*, HPV, HBV, and HCV are the leading infectious agents contributing to the global cancer burden. When summed together they account for 92% of all infection-attributable cancers worldwide with 35.4%, 29.5%, 19.2%, and 7.8% respectively[4]. The rise of the HIV epidemic concentrated in low and middle-income countries has resulted to an increase in HIV-associated malignancies[5]. In Kenya, a HIV prevalence of 5.6% (95% CI: 4.9 to 6.3), and HIV incidence of 0.5% (95% CI: 0.2 to 0.9), corresponding to an annual HIV transmission rate of 8.9 per 100 HIV-infected persons has been reported[6]. This population is at greater risk of acquiring HIV associated cancers such as Kaposi's sarcoma (KS), non-Hodgkin lymphoma (NHL) and invasive cancer of the cervix (ICC)[7]. According to a recent study, KS was the second largest contributor to the cancer burden in sub-Saharan Africa[4], while NHL is the second most common malignant disorder associated with HIV infection worldwide[7]. Previous studies done in Africa have been discussed in details in the methodology section.

Information on the burden of cancer and especially the burden attributable to infections is sparse in Kenya. In this study, we highlight the results from the adult population of two National referral hospitals in Kenya for five-year period between January 2008 to December 2012.

Methods

This was a retrospective cross-sectional study conducted at Kenyatta National Hospital (KNH) and Moi Teaching and Referral Hospital (MTRH) in Kenya. Initially the study targeted four teaching and referral hospitals located in the former Nairobi, Rift Valley, Coast and Nyanza Provinces, but the authorization to access medical records was only granted by the above mentioned hospitals. Kenya was divided into eight provinces (see map from the Kenya bureau of statistics[8]) before the new constitution of Kenya that came into force in 2013. KNH is located in the former Nairobi Province which is the capital and the largest city of Kenya and according to the last official census taken in 2009 it had a population of 3,138,369 whose number has since grown to approximately 3.5 million people. KNH has a capacity of 2,000 beds and attends to an annual average of 70,000 inpatients and 600,000 outpatients. The Hospital has 50 wards and 24 operating theatres as well as 24 consultant clinics with over 6000 staff members. As a referral hospital, KNH offers specialized quality health care to patients from all over the Nation, East and Central Africa Region. MTRH is located in the former Rift Valley Province (Kenya's largest Province) with a population of 10,006,805[8]. MTRH has a bed capacity of about 991 patients, an average number of 1200 patients at any time and about 1500 out patients per day. The Hospital serves residents of Western Kenya Region (representing at least 22 Counties), parts of Eastern Uganda and Southern Sudan. AMPATH-Oncology centre is located at MTRH which evolved from a HIV program to a pediatric cancer program and later to cancer care program. It is integrated in the MTRH department of Hematology and Oncology[9]. At AMPATH oncology centre more than 1000 cancer patients are treated each month with about 45 oncology specialists. According to the Kenyan network of cancer organizations KNH and MTRH are amongst the oldest and largest public referral hospitals with cancer treatment services. Both hospitals offer screening, diagnosis and treatment services. There are few radiation machines and according to Strother *et al.*, 2013, there were two cobalt-60 radiation oncology machines, both housed in KNH[9]. The two national hospitals are also the largest source of data for the two main cancer registries in Kenya. KNH provides data to the Nairobi cancer registry located in Nairobi while MTRH provides data to the Eldoret cancer registry located in Rift Valley. Recently, with an aim to decongest the national hospitals more (private, mission and public) health facilities have been equipped with the cancer services. However, because of the affordability of the services many patients opt for the public health facilities.

Data source

Data for this study was obtained from hospital records of patients as this was the most convenient data source for all the information targeted.

Inclusion criteria
- Hospital records of patients diagnosed with cancer during the period January 2008 to December 2012.

- Records of patients above the age of 18 at the time of diagnosis and with a confirmed diagnosis either by histology, radiology or haematology.

Exclusion criteria
- Hospital records with incomplete data or not meeting the above criteria.

Sample size calculation

The sample size (n) was calculated according to the guidelines outlined for calculating sample sizes for cross-sectional studies (qualitative variable) as explained by 10–12 and as shown below;

$$n = \frac{z^2 \hat{p}(1 - \hat{p})}{d^2}$$

Where:

p = Prevalence of condition or health state or the expected prevalence or proportion or estimated proportion of a disease

d = degree of precision of the estimate or the absolute error

z = Z statistic for a level of confidence or is the normal distribution critical value for a probability of α/2 in each tail. For a 95% CI, z=1.96. A 95% level of confidence and a ±5% (0.05) degree of precision were considered.

The prevalence of cancer in Kenya was not known at the time of the study and therefore a prevalence of 50% (0.5) was used in calculating the sample size. Elsewhere, it has been highlighted that when d=0.05 and a z=1.96, using a p of 0.5 (50%) yields the highest estimates for n (sample size)[13].

A sample size of 384 was estimated as the minimal necessary to achieve the required power of the study. As the data

from this study was collected at a single point in time, a 30% (116) non-response allowance was factored in resulting to a final sample size of 500. This allowance would help us in case one of the 384 files would be found incomplete after the completion of the data collection period. In KNH an estimated 17,584 (inpatient and outpatient) cancer files were reported while in MTRH 4304 (inpatient) cancer cases were reported during the five year period. Due to cost and time constraints, it was only feasible to study 500 files as calculated from the sample size. To obtain the final number of 500 files from the totals in each hospital, a proportional stratified sampling method was used as described by 11. (see Table 1)

To randomly select the calculated proportions of files for each year obtained in the previous step, a systematic random sampling method was used as described by 11. At KNH, the files were all available at the health information department and in the databases where systematic random sampling was an automated process. At the time of data collection, MTRH was in the process of updating the database and only 2012 files were available at the health information department. The files for 2008 to 2011 were obtained from the oncology centre at the Academic Model for the Prevention and Treatment of HIV/AIDS (AMPATH), and convenient sampling was used to achieve the required number of files. The reason for choosing convenient sampling was because we could not establish the total sampling frame to do the systematic randomization. However, an indirect systematic randomization was applied to select the files needed from the total files accessed per year and not the sampling frame. By the end of the data collection period, all the targeted 500 files were available from both the hospitals. In case a selected file was found incomplete during the data collection period, it was replaced by the next file following it after randomization.

Data collection

A pre-designed questionnaire was used to abstract the information (Supplementary File 1). The information abstracted

Table 1. Sample calculations for Kenyatta National Hospital (KNH) and Moi Teaching and Referral Hospital (MTRH).

	KNH			MTRH		
Year	Estimated Files (N)	Sampling fraction	Final sample size (Proportion)	Estimated Files (N)	Sampling fraction	Final sample size (Proportion)
2008	3168	18%	90	947	22%	110
2009	2834	16%	80	1325	31%	154
2010	3048	17%	85	489	11%	57
2011	4161	24%	120	732	17%	85
2012	4373	25%	125	811	19%	94
Total	**17,584**	**100%**	**500**	**4,304**	**100%**	**500**

from the files included patients age, sex, origin, type of cancer, method of cancer diagnosis, year of diagnosis and whether the patient was referred from another hospital.

Ethical consideration

The study was approved by the Kenyatta National Hospital/ University of Nairobi (KNH/UoN-ERC) and the Moi Teaching and Referral Hospital (IREC/MTRH) Ethics and Research Committees with approval numbers P24/01/2013 and FAN: IREC 1027 respectively. Endorsement for the study was obtained from the Director of Medical Services while the permission to access data from the hospital databases and patients files was obtained from the director of the Health Information Department. Although, the study aimed at conducting the research at four referral hospitals in Kenya, only two hospitals granted permission to access the patient files while the other two refused and the reasons for refusal were unknown. The study was a minimal risk study and patient consent was not sought since there was no direct patient involvement but a retrospective review of patients' files. However, the patient identifying information was not included in the data collection forms.

Data handling and analysis

Data was entered into statistical package for social sciences programme (IBM-SPSS) version 23 that was labeled using the exact fields as the questionnaires and the excel files. Quality control checks were performed to prevent double entry and to ensure accurate entry of the data. The proportions of cancer cases were analyzed with reference to the study site, sex and the age group. GraphPad Prism 6 (GraphPad Software Inc., San Diego, CA, USA) was used to draw the figure images. The cancers were listed according to the third edition of the International Classification of Diseases for Oncology (ICD-O). For cancers like BL (9687/3), HL (9650/3), NHL (9596/3), adult T-cell leukemia (9827/3), KS (9140/3), acute leukemia (9835/3 & 9861/3), chronic leukemia (9863/3 & 9823/3) and multiple myeloma (9732/3) where ICD-0-3 topographical code was not clear, we converted the IDC-0-3 morphological code to ICD10 topographical code for clarity.

Although many studies have shown a variety of cancers associated with infectious agents, the attributable fraction (AF) standard formula was applied to a group of cancers classified as carcinogenic in the IARC monograph 100b[3] namely; cervix (C53), liver (C22.0), stomach (C16), Kaposi's sarcoma (C46), non-Hodgkin's lymphoma (C82-85, C96), Hodgkin's lymphoma (C81), nasopharynx (C11), oropharynx (C10), bladder (C67), vulva (C51), vagina (C52), penis (C60), anus (C21) and bile duct (C22.1). Calculation of the attributable fraction (AF) relies upon the standard formula for population attributable risk as shown by 14, 15:

$$AF = p(r-1) / [p(r-1)+1]$$

Where:

p=prevalence of exposure to the infectious agents to the population

r=relative risk of exposure

However, the use of this formula requires prior knowledge on *p* and *r* as defined above and since this information is limited in Kenya, we used AF derived from other studies done elsewhere either in developing countries, sub-Saharan Africa, or world estimates that used data from developing countries. The formula results in a proportion that is applied to the total number of cases in the target population to obtain the number of cases that can theoretically be attributed to the factor in that population[14].

In determining AF estimates for cervical cancer, we identified other African studies that reported the prevalence of HPV in cervical cancer to be between 76% to 91%[16–18]. However, an AF of 100% was used for this study since the causality of HPV in cervical cancer is generally known[4,19,20]. For vulvar cancer, studies from Botswana reported a HPV prevalence of 56.8 to 100%[21,22], but an AF of 40% derived from developing countries estimates was used[14]. We did not find any clear study showing the prevalence of HPV in vaginal cancers in sub-Saharan Africa, but an AF of 78% derived from world estimates was used[4,20]. Studies from Central African Republic and Botswana identified a prevalence of HPV in anal cancer to be 69.1%[21,23], but a world estimate AF of 88% was used[4,20]. The prevalence of HPV in penile cancer has been shown to be 41.9% to 68% from Kenya and Botswana studies[21,24] but a world estimate AF of 50% was used[4,20]. A study from South Africa showed the prevalence of HPV in oropharyngeal cancer to be 5.6% among men (N=125)[25], while another from Ghana found a HPV prevalence of 13.8% (4 of 29) from the oral cavity, and 18.2% (2 of 11) from the pharynx[26], but an AF of 30.8% derived from world estimate was used[4,20] (Table 2).

In determining AF estimates for KS, a study from Western Kenya found a Kaposi´s sarcoma-associated herpes virus (KSHV) positivity of 50.1 to 63.5% using 228 surgical cases[27] but a world estimate AF of 100% was used since HHV8 is recognized as a necessary cause of KS in HIV infections[4,7]. For EBV in NHL, A study from Lusaka, Zambia, detected EBV in 54.5% of the cases[28]. We used an estimated AF of 100% based on meta-analysis studies from developing countries[14,15]. Elsewhere in Malawi, EBV in HL was present in 18 of 24 (75%) tumor specimens, including 14 of 20 (70%) HIV- and 4 of 4 (100%) HIV+[29]. Similarly a study from Kenya found an EBV in HL positivity of 100% present in pediatric cases[30] while Zambia, detected EBV in HL of 40.9%[28]. However, an AF of 74% derived from African estimates was used[4]. No studies were encountered showing the prevalence of EBV in nasopharyngeal cancer in sub-Saharan Africa, therefore an AF of 95.5% derived from world estimates was used[4].

A study aiming to study the prevalence of HBV/HCV in hepatocellular carcinoma (HCC) patients in Gambia found that HBV carriage was present in 61% (129/211) of HCC while HCV present in 19% (36/191) of HCC patients[32]. An AF of 50% for HBV and 21% for HCV obtained from sub-Saharan estimates was used[33]. A study aiming to study the prevalence of H. pylori in Kano, Nigeria found a prevalence of 81.7%[34] while in Uganda 18 of the 21 cases of stomach cancer had

Table 2. Previous prevalence studies on infectious agents and cancers.

Infectious agent	Cancer Site	ICD-0 code	Studies done in sub Saharan Africa	Prevalence of the infectious agent (%)	AF estimates used for this study (%)
HPV	Cervix uteri	C53	South Africa, Gabon[16–18]	76–91	100[4,19,20]
	Vulva	C51	Botswana[21,22]	57–100	40[14]
	Vagina	C52	-	-	78[4,20]
	Anus	C21	Central African Republic, Botswana[21,23]	69	88[4,20]
	Oropharynx	C10	South Africa, Ghana[25,26]	6–18	31[4,20]
	Penis	C60	Kenya, Botswana[21,24]	42–68	50[4,20]
HHV8/HIV	KS	**C46**	Kenya[27]	50–64	100[4,7,20]
EBV	HL	**C81**	Kenya, Zambia, Malawi[28–30]	41–100	74[4]
	Nasopharynx	C11	-	-	96[4]
	BL	C83.7	Africa[31]	-	-
EBV/HIV	NHL	**C82-C85, C96**	Zambia[28]	55	100[14,15]
HBV/HCV	Liver	C22.0	Gambia[32]	70	71[33]
H. pylori	Stomach	C16	Nigeria, Uganda[34,35]	82–86	89[4]
Schistosoma	Bladder	C67	South Africa[36]	10–85	41[4]
O. viverrini C. sinensis	Bile duct	C22.1	-		-
HTLV-1	ATLL	C91.5	-	-	-

Helicobacter pylori (H. pylori), hepatitis B virus (HBV), hepatitis C virus (HCV), human immunodeficiency virus type 1 (HIV-1), human papillomavirus (HPV), Epstein-Barr virus (EBV), human herpes virus type 8 (HHV-8), human T-cell lymphotropic virus type 1 (HTLV-1), Adult T-cell Leukaemia/lymphoma (ATLL), *Opisthorchis viverrini, Clonorchis sinensis, Schistosoma haematobium*, non-Hodgkin's lymphoma (NHL), Kaposi's sarcoma (KS), Hodgkin's lymphoma (HL) Burkitt's lymphoma (BL), AF=attributable fraction, - = not available or not obtainable

H.pylori[35]. However, an AF of 89% derived from world estimates was used[4] (Table 2).

In South Africa, ova of *Schistosoma haematobium* were seen in microscopic sections of bladder tumours in 85% of the patients with squamous cell carcinoma, in 50% of those with undifferentiated tumours and adenocarcinoma, in 17% of those with mixed tumours or sarcoma, and in only 10% of the patients with transitional cell carcinoma (all classifications of the bladder tumours)[36]. However, an AF of 41% was used derived from endemic areas in Africa[4].

We did not come across any studies in Africa showing the prevalence of *Opisthorchis viverrini* and *Clonorchis sinensis* in cancer of the bile duct. Similarly the AF could not be obtained. Burkitt's lymphoma (BL) was first described in Eastern Africa where the highest incidence and mortality rates are seen. It has been associated with EBV and affects mainly children, where boys are more susceptible than girls[31]. However, we did not come across any cases of BL or Adult T-cell leukaemia/lymphoma (ATLL) from our study.

Results
Demographics
An estimated total of 17,584 and 4304 cancer files were recorded in the KNH and MTRH respectively from 2008 to 2012. For

the purpose of this study data was only obtained from 500 randomly selected files per hospital. In KNH, 60% of these were females while 40% were males with a mean age of 50.57 (18 to 95 years), giving a male to female (M:F) ratio of 1:1.5. In MTRH, 56% were females while 44% were males with a mean age of 48 years (18 to 90 years) giving a M:F ratio of 1:1.2. More than 70% of patients were within the age group of 25 to 64 years in both hospitals, largest age group being 45 to 64 years (43.4%) in KNH, and 25 to 44 years (39.4%) in MTRH. By Province, most KNH patients were from Central, Eastern and Western, (44%, 24.8% and 10.2% respectively), with specific regions being Muranga, Kiambu, Machakos, Nyeri and Kisii (11.2%, 10.2%, 8%, 8% and 5.4% respectively) with 66% of these patients being referred from other hospitals. On the other hand, most MTRH cancer patients were from the Rift valley, Western and Nyanza Provinces with 66.8%, 24.8% and 6.4% respectively with specific highest regions being Uasin Gishu, Nandi, Lugari, Trans Nzoia and Bungoma (22.8%, 8.4%, 8.4%, 7.4% and 6% respectively). 41% of these cancer cases were referrals from other hospitals (Table 3).

Types of cancers
In KNH (n=500), the top five types of cancers were: cervical (62, 12.4%), breast (59, 11.8%), colorectal (31, 6.2%), chronic leukemia (27, 5.4%) and stomach cancers with 26 (5.2%). In females (n=300) the five most common cancers were cervical

Table 3. Cancer cases by age groups, gender, method of cancer diagnosis, region and referral information in KNH/MTRH: 2008–2012.

Site	KNH						MTRH					
Year	08	09	10	11	12	All	08	09	10	11	12	All
Age groups												
≤24	5	4	6	9	7	31	8	11	4	2	3	28
%	5.6	5	7.1	7.5	5.6	6.2	7.3	7.1	7	2.4	3.2	5.6
25–44	30	22	24	34	39	149	42	70	20	35	30	197
%	33.3	27.5	28.2	28.3	31.2	29.8	38.2	45.5	35.1	41.2	31.9	39.4
45–64	35	35	36	52	59	217	45	50	21	34	39	189
%	33.9	43.8	42.4	43.3	47.2	43.4	40.9	32.5	36.8	40	41.5	37.8
65–85	19	18	18	25	18	98	15	21	10	14	17	77
%	21.1	22.5	21.2	20.8	14.4	19.6	13.6	13.6	17.5	16.5	18.1	15.4
≥85	1	1	1	0	2	5	0	2	2	0	5	9
%	1.1	1.3	1.2	0	1.6	1	0	1.3	3.5	0	5.3	1.8
Total	90	80	85	120	125	500	110	154	57	85	94	500
%	100	100	100	100	100	100	100	100	100	100	100	100
Gender												
Male	39	37	33	51	40	200	50	74	17	25	52	218
%	43	46.3	38.8	42.5	32	40	45.5	48.1	29.8	29.4	55.3	43.6
Female	51	43	52	69	85	300	60	80	40	60	42	282
%	56.7	53.7	61.2	57.5	68	60	54.5	51.9	70.2	70.6	44.7	56.4
Total	90	80	85	120	125	500	110	154	57	85	94	500
%	100	100	100	100	100	100	100	100	100	100	100	100
Method of Cancer Diagnosis												
Radiology	61	68	63	95	107	394	7	12	20	28	84	151
%	67.8	85	74.1	79.2	85.6	78.8	6.4	7.8	35.1	32.9	89.4	30.2
Biopsy	90	80	85	120	125	500	110	154	57	85	94	500
%	100	100	100	100	100	100	100	100	100	100	100	100
Cancer Cases by Province												
Central	40	30	37	57	56	220	1	1	1	3	1	7
%	44.4	37.5	43.5	47.5	44.8	44	0.9	0.6	1.8	3.5	1.1	1.4
Coast	2	1	1	0	2	6	0	0	0	0	0	0
%	2.2	1.3	1.2	0	1.6	1.2	0	0	0	0	0	0
Eastern	24	22	22	30	26	124	0	0	0	0	1	1
%	26.7	27.5	25.9	25	20.8	24.8	0	0	0	0	1.1	0.2
Nairobi	2	1	3	6	8	20	0	0	0	0	0	0
%	2.2	1.3	3.5	5	6.4	4	0	0	0	0	0	0
Nyanza	4	8	12	7	9	40	4	15	4	5	4	32
%	4.4	10	14.1	5.8	7.2	8	3.6	9.7	7.0	5.9	4.3	6.4
Rift Valley	7	7	4	6	13	37	77	95	33	60	69	334
%	7.8	8.8	4.7	5	10.4	7.4	70	61.7	57.9	70.6	73.4	66.8
Western	11	11	5	13	11	51	28	41	19	17	19	124
%	12.2	13.8	5.9	10.8	8.8	10.2	25.5	26.6	33.3	20	20.2	24.8
Others	0	0	1	1	0	2	0	2	0	0	0	2
%	0	0	1.2	0.8	0	0.4	0	1.3	0	0	0	0.4
Total	90	80	85	120	125	500	110	154	57	85	94	500
%	100	100	100	100	100	100	100	100	100	100	100	100
Referral												
Yes	57	48	57	85	83	330	44	64	19	39	41	207
%	63.3	60	67.1	70.8	66.4	66	40	41.6	33.3	45.9	43.6	41.4
No	33	32	28	35	42	170	66	90	38	46	53	293
%	36.7	40	32.9	29.2	33.6	34	60	58.4	66.7	54.1	56.4	58.6
Total	90	80	85	120	125	500	110	154	57	85	94	500
%	100	100	100	100	100	100	100	100	100	100	100	100

% = percentage

(62, 20.7%), breast (59, 19.7%), ovarian (22, 7.3%), chronic leukemia (16, 5.3%), and endometrial and stomach cancers both with 15 (5%). In males (n=200) the five most common types of cancers were prostate (23, 11.5%), laryngeal (19, 9.5%), colorectal (17, 8.5%), esophageal (14, 7.0%) and nasopharyngeal cancers with 12 (6.0%). In MTRH (n=500), the five most common types of cancer were Kaposi's sarcoma (93, 18.6%), breast (77, 15.4%), cervical (41, 8.2%), non-Hodgkin's lymphoma (37, 7.4%) and colorectal, chronic leukemia and esophagus cancer all with 27 (5.4%). In females (n=282) the five most common types of cancers were breast cancer (74, 26.2%), cervical (41, 14.5%), Kaposi's sarcoma (38, 13.5%), non-Hodgkin's lymphoma (15, 5.3%) and ovarian cancer with 14 (5%). In males (n=218) the five most common types of cancers were Kaposi's sarcoma 55(25.2%), non-Hodgkin's lymphoma 22(10.1%), chronic leukemia 17(7.8%), colorectal and esophageal cancer both with 16 (17.3%) (Table 4 and Figure 1).

Cancer cases by age groups

From the results generated from both hospitals, it was suggestive that some cancers were predominant in specific age groups. Acute leukemia, NHL, cancer of the bone, genitalia,

Table 4. Cancer cases by gender in Kenyatta National Hospital (KNH) and Moi Teaching and Referral Hospital (MTRH) from 2008–2012.

Cancer site	ICD-0 code	KNH No. of cases	P	Males		Females		MTRH No. of cases	P	Males		Females	
		N	%	n	%	n	%	N	%	n	%	N	%
Cervix uteri	C53	62	12.4	0	0	62	20.7	41	8.2	0	0	41	14.5
Breast	C50	59	11.8	0	0	59	19.7	77	15.4	3	1.4	74	26.2
Colon & Rectum	C18,C20	31	6.2	17	8.5	14	4.7	27	5.4	16	7.3	11	3.9
Chronic leukemia	C91.1,C92.1	27	5.4	11	5.5	16	5.3	27	5.4	17	7.8	10	3.5
Stomach	C16	26	5.2	11	5.5	15	5	15	3	5	2.3	10	3.5
Esophageal	C15	25	5	14	7	11	3.7	27	5.4	16	7.3	11	3.9
Prostate	C61	23	4.6	23	11.5	0	0	7	1.4	7	3.2	0	0
Ovary	C56	22	4.4	0	0	22	7.3	14	2.8	0	0	14	5
Larynx	C32	21	4.2	19	9.5	2	0.7	3	0.6	3	1.4	0	0
Skin	C44	19	3.8	11	5.5	8	2.7	11	2.2	7	3.2	4	1.4
Nasopharynx	C11	17	3.4	12	6	5	1.7	11	2.2	8	3.7	3	1.1
Endometrium	C54.1	15	3	0	0	15	5	5	1	0	0	5	1.8
Pancrease	C25	14	2.8	6	3	8	2.7	11	2.2	3	1.4	8	2.8
Acute leukemia	C91.0,C92.0	13	2.6	5	2.5	8	2.7	8	1.6	4	1.8	4	1.4
Non Hodgkin´s lymphoma	C82-C85, C96	13	2.6	7	3.5	6	2	37	7.4	22	10.1	15	5.3
Lung and bronchus	C34	11	2.2	8	4	3	1	3	0.6	1	0.5	2	0.7
Lip and oral cavity	C00-C14	11	2.2	6	3	5	1.7	9	1.8	4	1.8	5	1.8
Hypopharynx	C13	11	2.2	4	2	7	2.3	0	0	0	0	0	0
Liver	C22	10	2	5	2.5	5	1.7	16	3.2	13	5.9	3	1.1
Bone	C40-C41	8	1.6	6	3	2	0.7	10	2	7	3.2	3	1.1
Bladder	C67	7	1.4	5	2.5	2	0.7	2	0.4	1	0.5	1	0.4
Thyroid	C73	7	1.4	0	0	7	2.3	0	0	0	0	0	0
Genitalia	*	6	1.2	3	1.5	3	1	8	1.6	3	1.4	5	1.8
Eye	C69	6	1.2	3	1.5	3	1	0	0	0	0	0	0
Multiple myeloma	C90.0	6	1.2	4	2	2	0.7	7	1.4	5	2.3	2	0.7
Hodgkin´s lymphoma	C81	5	1	2	1	3	1	14	2.8	7	3.2	7	2.5
Parotid gland	C07	4	0.8	3	1.5	1	0.3	0	0	0	0	0	0
Bile duct	C22.1	4	0.8	3	1.5	1	0.3	3	0.6	1	0.5	2	0.7
Kaposi's sarcoma	C46	2	0.4	1	0.5	1	0.3	93	18.6	55	25.2	38	13.5
Others	**	15	3	11	5.5	4	1.3	14	2.8	10	4.6	4	1.4
Adult T cell Leukemia	C91.5	0	0	0	0	0	0	0	0	0	0	0	0
Burkitt´s Lymphoma	C83.7	0	0	0	0	0	0	0	0	0	0	0	0
Total		**500**	**100**	**200**	**100**	**300**	**100**	**500**	**100**	**218**	**100**	**282**	**100**

*Genitalia included cancer of the penis (C60), vagina (C52), vulva (C51), testis (C62) and pelvis (C63) while **others included cancers of renal (C65), head (C76), brain (C71), anus (C21), ear (C30), rhabdomyosarcoma (C49) and meninges (C70). N = number of cases P = percentage.
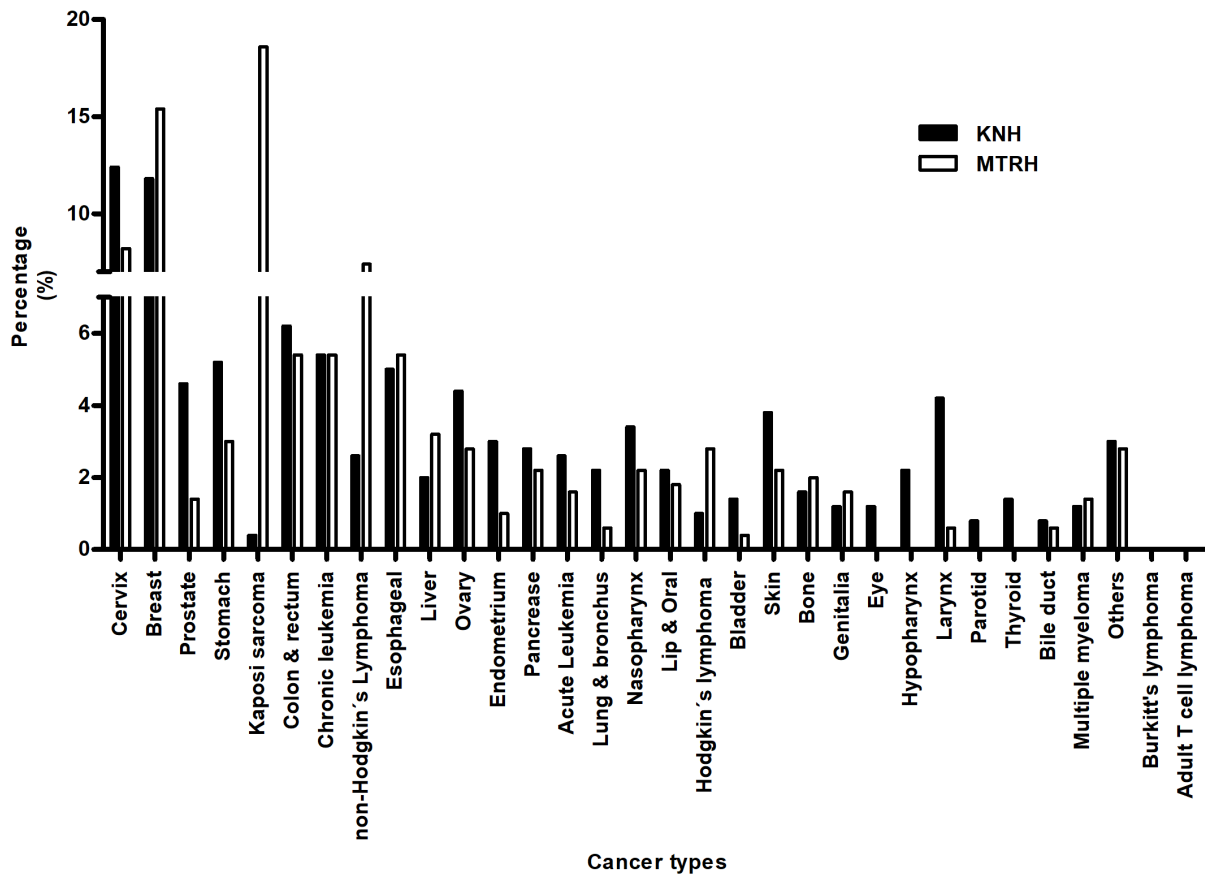
Figure 1. Types of cancers at Kenyatta National Hospital (KNH) and Moi Teaching and Referral Hospital (MTRH) from 2008–2012, Sample (500/500).

HL and nasopharyngeal were predominant in the age group of 24 years and below. Cancer of the cervix was predominant in the age group of 22 to 44 years while breast and bile duct cancers were predominant between 45 to 64 years. The age-group of 65 to 84 years was predominated by esophagus, prostrate, larynx, endometrium and lip and oral cavity cancers. A larger age group ranging from 45 to 65 years was predominated by colorectal, stomach, ovary, lung and bronchus, liver, bladder and multiple myeloma cancers (Table 5 and Table 6). This information is particularly important for estimating the age to go for the cancer checkups.

Infection-attributable cancers

The cancers listed here are cancers associated with the 11 infectious agents as classified and established by the International Agency for Research on Cancer (IARC). In KNH, 154 (30.8%) of the total cancers sampled were associated with infectious agents, while an estimated 138 (27.6%) were attributable to infections. Cancers of the cervix (62, 12.4%), stomach (26, 5.2%), nasopharynx (17, 3.4%), non-Hodgkin's lymphoma (13, 2.6%) and liver (10, 2%) were the commonest infection-associated cancers. In MTRH, 241 (48.2%) of the 500 cancers cases sampled were associated with infectious agents, while an estimated

222 (44.4%) were attributable to infections. Kaposi's sarcoma (93, 18.6%), cancer of the cervix (41, 8.2%) and non-Hodgkin's lymphoma (37, 7.4%), liver (16, 3.2%) and stomach (15, 3%) were the commonest infection-associated cancers (Table 7).

Discussion

In females (n=300) the five most common cancers in KNH were cervical, breast, ovarian, chronic leukemia, endometrial and stomach while in MTRH (n=282), they were breast, cervical, Kaposi's sarcoma, non-Hodgkin's lymphoma and cancer of the ovary. These results are comparable with the data obtained from a previous study conducted retrospectively in Tenwek Hospital, in Bomet District, western Kenya in the period of 1999 to 2007 that showed that the common types of cancer in women were cervical, breast, stomach, uterus and esophageal[37]. Similarly an incidence rates study done using the Nairobi cancer registry data found breast, cervical, esophageal, large bowel stomach and ovarian cancers to be the most incident[38]. The high number of cervical cancer could reflect a potential higher prevalence of HPV infection, low screening rates or late detection of the disease. Most patients with cervical cancer were between in the age groups of 24 to 44 and 45 to 54

Table 5. Cancer cases by age groups in Kenyatta National Hospital (KNH) from 2008–2012.

KNH										
Cancer site	**ICD-0 code**	**Age groups**								
		≤24		**24–44**		**45–64**		**65–84**		**≥85**	
		cases	**%**	**cases**	**%**	**Cases**	**%**	**Cases**	**%**	**Cases**	**%**
Cervix uteri	C53	3	9.7	26	**17.4**	29	13.4	4	4.1	0	0
Breast	C50	2	6.5	19	12.8	30	**13.8**	8	8.2	0	0
Colon & Rectum	C18,C20	1	3.2	4	2.7	19	**8.8**	7	7.1	0	0
Chronic leukemia	C91.1,C92.1	1	3.2	12	**8.1**	10	4.6	4	4.1	0	0
Stomach	C16	1	3.2	3	2	14	**6.5**	6	6.1	2	40
Esophageal	C15	0	0	2	1.3	14	6.5	9	**9.2**	0	0
Prostate	C61	0	0	0	0	10	4.6	12	**12.2**	1	20
Ovary	C56	1	3.2	7	4.7	11	**5.1**	3	3.1	0	0
Larynx	C32	0	0	1	0.7	13	6	6	**6.1**	1	20
Skin	C44	2	**6.5**	9	6	4	1.8	4	4.1	0	0
Nasopharynx	C11	2	**6.5**	5	3.4	7	3.2	3	3.1	0	0
Endometrium	C54.1	0	0	4	2.7	6	2.8	5	**5.1**	0	0
Pancrease	C25	0	0	4	2.7	5	2.3	5	**5.1**	0	0
Acute leukemia	C91.0,C92.0	2	**6.5**	9	6	1	0.5	1	1	0	0
Non Hodgkin´s lymphoma	C82-C85, C96	3	**9.7**	5	3.4	4	1.8	1	1	0	0
Lung and bronchus	C34	0	0	4	2.7	6	**2.8**	1	1	0	0
Lip and oral cavity	C00-C14	1	3.2	4	2.7	1	0.5	5	**5.1**	0	0
Hypopharynx	C13	0	0	2	1.3	6	2.8	3	3.1	0	0
Liver	C22	0	0	3	2	6	**2.8**	1	1	0	0
Bone	C40-C41	4	**12.9**	3	2	1	0.5	0	0	0	0
Bladder	C67	0	0	3	2	1	0.5	3	**3.1**	0	0
Thyroid	C73	0	0	3	2	4	1.8	0	0	0	0
Genitalia	*	1	**3.2**	4	2.7	1	0.5	0	0	0	0
Eye	C69	1	3.2	3	2	1	0.5	1	1	0	0
Multiple myeloma	C90.0	0	0	1	0.7	4	**1.8**	0	0	1	20
Hodgkin´s lymphoma	C81	2	**6.5**	3	2	0	0	0	0	0	0
Parotid gland	C07	0	0	2	1.3	1	0.5	1	1	0	0
Bile duct	C22.1	0	0	0	0	2	**0.9**	2	2	0	0
Kaposi's sarcoma	C46	0	0	0	0	0	0	2	1	0	0
Others	**	4	12.9	4	2.7	6	2.8	1	1	0	0
Adult T cell Leukemia	C91.5	0	0	0	0	0	0	0	0	0	0
Burkitt´s Lymphoma	C83.7	0	0	0	0	0	0	0	0	0	0
Total		**31**	**100**	**149**	**100**	**217**	**100**	**98**	**100**	**5**	**100**

*Genitalia included cancer of the penis (C60), vagina (C52), vulva (C51), testis (C62) and pelvis (C63) while **others included cancers of renal (C65), head (C76), brain (C71), anus (C21), ear (C30), rhabdomyosarcoma (C49) and meninges (C70). N = number of cases P = percentage

Table 6. Cancer cases by age groups in Moi Teaching and Referral Hospital (MTRH) from 2008–2012.

MTRH											
Cancer site	**ICD-0 code**	**Age groups**									
		≤24		24–44		45–64		65–84		≥85	
		Cases	**%**	**Cases**	**%**	**Cases**	**%**	**Cases**	**%**	**Cases**	**%**
Cervix uteri	C53	0	0	20	**10.2**	16	8.5	4	5.2	1	11.1
Breast	C50	3	10.7	28	14.2	37	**19.6**	8	10.4	1	11.1
Colon & Rectum	C18,C20	0	0	9	4.6	10	5.3	7	**9.1**	1	11.1
Chronic leukemia	C91.1,C92.1	1	3.6	10	5.1	10	5.3	6	**7.8**	0	0
Stomach	C16	0	0	4	2	6	3.2	4	**5.2**	1	11.1
Esophageal	C15	0	0	3	1.5	15	7.9	8	**10.4**	1	11.1
Prostate	C61	0	0	0	0	2	1.1	4	**5.2**	1	11.1
Ovary	C56	0	0	6	3	4	2.1	4	**5.2**	0	0
Larynx	C32	0	0	1	0.5	1	0.5	1	**1.3**	0	0
Skin	C44	1	3.6	2	1	3	1.6	4	**5.2**	1	11.1
Nasopharynx	C11	2	**7.1**	5	2.5	3	1.6	1	1.3	0	0
Endometrium	C54.1	0	0	0	0	2	1.1	3	**3.9**	0	0
Pancrease	C25	1	**3.6**	2	1.0	6	3.2	0	0	2	**22.2**
Acute leukemia	C91.0,C92.0	2	**7.1**	4	2.0	1	0.5	1	1.3	0	0
Non Hodgkin´s lymphoma	C82-C85, C96	6	**21.4**	9	4.6	17	9	5	6.5	0	0
Lung and bronchus	C34	0	0	0	0	1	0.5	2	**2.6**	0	0
Lip and oral cavity	C00-C14	1	3.6	3	1.5	1	0.5	4	**5.2**	0	0
Hypopharynx	C13	0	0	0	0	0	0	0	0	0	0
Liver	C22	1	3.6	6	3	5	2.6	4	**5.2**	0	0
Bone	C40-C41	2	**7.1**	5	2.5	2	1.1	1	1.3	0	0
Bladder	C67	0	0	0	0	2	**1.1**	0	0	0	0
Thyroid	C73	0	0	0	0	0	0	0	0	0	0
Genitalia	*	1	**3.6**	6	3	1	0.5	0	0	0	0
Eye	C69	0	0	0	0	0	0	0	0	0	0
Multiple myeloma	C90.0	0	0	0	0	5	**2.6**	2	**2.6**	0	0
Hodgkin´s lymphoma	C81	3	**10.7**	7	3.6	3	1.6	1	1.3	0	0
Parotid gland	C07	0	0	0	0	0	0	0	0	0	0
Bile duct	C22.1	0	0	1	0.5	2	**1.1**	0	0	0	0
Kaposi's sarcoma	C46	4	14.3	61	31	27	14.3	1	1.3	0	0
Others	**	0	0	5	2.5	7	3.7	2	2.6	0	0
Adult T cell Leukemia	C91.5	0	0	0	0	0	0	0	0	0	0
Burkitt´s Lymphoma	C83.7	0	0	0	0	0	0	0	0	0	0
Total		**28**	**100**	**197**	**100**	**189**	**100**	**77**	**100**	**9**	**100**

*Genitalia included cancer of the penis (C60), vagina (C52), vulva (C51), testis (C62) and pelvis (C63) while **others included cancers of renal (C65), head (C76), brain (C71), anus (C21), ear (C30), rhabdomyosarcoma (C49) and meninges (C70). N = number of cases
P = percentage

Table 7. Attributable fraction (AF) and estimated number of cancers attributable to infectious agents in KNH/MTRH from 2008–2012.

Infectious agent	Cancer site	ICD-0 code	AF%	No. of cancer cases KNH	No. of cancer cases attributable to infections KNH	No. of cancer cases MTRH	No. of cancer cases attributable to infections MTRH	No. of cancer cases attributable to infections KNH&MTRH
Female								
HPV	Cervix	C53	100	62	62	41	41	103
	Vulva	C51	40	2	1	3	1	2
	Vagina	C52	78	1	1	0	0	1
	Anus	C21	88	2	2	0	0	2
	Oropharynx	C10	31	0	0	0	0	0
Schistosoma	Bladder	C67	41	2	1	1	0	1
HBV/HCV	Liver	C22.0	71	5	4	3	2	6
EBV	Nasopharynx	C11	96	5	5	3	3	8
	HL	**C81**	74	3	2	7	5	7
H. pylori	Stomach	C16	89	15	13	10	9	22
EBV/HIV	NHL	**C82-85,C96**	100	6	6	15	15	21
HIV/HHV8	KS	**C46**	100	1	1	38	38	39
O. viverrini C. sinensis	Bile duct	C22.1	*	1	**	2	**	
Total (% female cancers)				105(35.0%)	98 (32.7%)	123(43.6%)	114 (40.4%)	212 (36.4%)
Male								
HPV	Penis	C60	50	0	0	2	1	1
	Anus	C21	88	3	3	4	4	7
EBV	Oropharynx	C10	31	0	0	0	0	0
Schistosoma	Bladder	C67	41	5	2	1	0	2
HBV/HCV	Liver	C22.0	71	5	4	13	9	13
EBV	Nasopharynx	C11	96	12	12	8	8	20
	HL	**C81**	74	2	1	7	5	6
H. pylori	Stomach	C16	89	11	10	5	4	14
EBV/HIV	NHL	**C82-85,C96**	100	7	7	22	22	29
HIV/HHV8	KS	**C46**	100	1	1	55	55	56
O. viverrini C. sinensis	Bile duct	C22.1	*	3	**	1	**	0
Total (% of male cancers)				49 24.5%	40 20%	118 54.1%	108 49.5%	148 (35.41%)
Total (% of cancers in both sexes)				154 30.8%	138 27.6%	241 48.2%	222 44.4%	360 36%

Over the 5-year period the total number of cases sampled in each hospital were 500. KNH had 300 female and 200 male cases while MTRH had 282 female and 218 male cases. *AF not available, ** Not possible to estimate without AF,

years while for breast cancer, they were in the age group of 45 to 64 years followed by 24 to 44 years with 35.5% and 29% respectively. The risk factors associated with breast cancer include early menarche, late childbearing, having fewer children, obesity, lack of awareness and early detection[39].

In males, (n=200) the five most common types of cancers in KNH were prostate, laryngeal, colorectal, esophageal and nasopharyngeal carcinoma while in MTRH, (n=218) they were: Kaposi's sarcoma, non-Hodgkin's lymphoma, chronic leukemia, colorectal and esophageal cancers. The high number of

prostate and esophageal cancer is comparable with the data from the Kenya National cancer control strategy that showed that Kaposi's sarcoma, prostate and esophageal cancer to be the most common cancers in men[40]. Similarly the retrospective study done in Tenwek Hospital in Western Kenya from 1999 to 2007 showed that the most common cancers in men were esophagus, stomach, prostate and colorectal and non-Hodgkin's lymphoma (NHL)[37]. Similarly an incidence rates study using the Nairobi cancer registry data found prostate, esophageal, large bowel, stomach, oral and liver cancers to be the most incident[38]. Other studies show that esophageal is the leading cause of death among both men and women in East Africa[37,39,41]. The majority of the esophageal cancers were in patients aged 65 to 84 years, followed by 45 to 64 years and was more common in males than women. Some of the risk factors independently associated with esophageal cancer ($P < 0.05$) identified from a study conducted at MTRH were low socio-economic status, smoking, alcohol consumption, tooth loss, cooking with charcoal and firewood, consumption of hot beverage and use of a traditional fermented milk referred to as *mursik*[41].

Elsewhere, a study aiming to determine the burden and pattern of cancer in Western Kenya by use of data from the Eldoret cancer registry, from 1999 to 2006, found out that about 21% of the patients had haematological malignancies where lymphomas were the most common (11.9%) followed by acute and chronic leukemia with 4.0% and 3.2% respectively. Esophageal (10.5%), breast (6.2%) and Kaposi's sarcoma (5.9%) were the top most non-haematological cancers. From our study, KS, NHL and chronic leukemia were high especially in MTRH[42]. Chronic leukemia was specifically the 4th most common type of cancer in KNH and 5th most common in MTRH while acute leukemia was also high and highest in the age-group of 24 years and below. It was suggestive that prostate cancer commonly affected males in the age-group of 65 to 84 years in both hospitals and with notable higher occurrence also in the age-group of 45 to 64 years. These differences could possibly be attributed to lifestyle choices and family history of the disease which are among the risk factors associated elsewhere with the cancer[43].

Infection-attributable cancers contributed up to a total of 36% of the total cancer burden in both hospitals. Specifically this was 27.6% for KNH and 44.4% for MTRH. This is closer to estimates of 31.1% for sub-Saharan Africa; higher than 9.2% estimated for developed countries and expectedly higher than global average of 15.4% reported by Plummer *et al.* in 2016[4]. Fortunately, the majority of cancers associated with infectious agents can be controlled by controlling the infectious agent[14,44]. Cancers of the cervix, stomach, nasopharynx, liver and NHL were the commonest infection-associated cancers in KNH; while in MTRH, KS, cancer of the cervix, NHL, liver and stomach cancer were the commonest infection-associated cancers. The high burden of cancers of the cervix, stomach and liver are comparable with other population based studies that show that among infection-related cancers, stomach, liver and cervical cancer, not only account for the vast majority of the total cancer burden associated with infections, but they have the highest incidences[4,14,45]. Kaposi´s sarcoma,

non-Hodgkin´s lymphoma and cervical cancer are HIV-associated malignancies[5,7]. The high numbers of Kaposi´s sarcoma, non-Hodgkin´s lymphoma and cervical cancer at MTRH could have been influenced by the source of data that was obtained from the AMPATH-Oncology centre that evolved from an existing HIV program[9]. The information on the HIV status of the patients from whose files were used, was not obtained to accurately link the associations. In Kenya, a HIV prevalence of 5.6% (95% CI: 4.9 to 6.3), and HIV incidence of 0.5% (95% CI: 0.2 to 0.9), corresponding to an annual HIV transmission rate of 8.9 per 100 HIV-infected persons has been reported[6]. This HIV pandemic could have influenced the high number of HIV-associated malignancies as documented by 4, 7, 46.

A high number of liver cancer cases were observed in males rather than females. This can be explained by higher prevalence of risk factors for liver cancer in men as compared to women such as higher alcohol consumption and infection with HBV or HCV[14]. The high number of gastric cancer in KNH suggested a higher prevalence of the risk factors associated with the cancer including infection with *H. pylori,* poor diet, poor hygiene, lack of awareness or the late stage of diagnosis[14,47].

More than 70% of patients were 25 to 64 years of age in both hospitals with the highest age group being 45 to 64 years in KNH (43.4%) and 25 to 44 years in MTRH (39.4%). According to the latest data published in WHO 2018, life expectancy in Kenya is 64.4 years for males and 64.8 for females the age frequently associated with cancer occurrence[48]. There were more female cases than males with 60% and 56.4% in KNH and MTRH respectively, this could be explained by the fact that probably males fear going to the hospitals, or women tend to have frequent contact with the health professionals and show up in even greater numbers than men during health campaigns. The findings of the study show a high number of referrals at KNH (66%) as compared to the 41.4% at MTRH. The possible reasons for the high referral rates at KNH could be explained by the establishment of a functional oncology centre which encourages the referral of patients for care while the decline at MTRH could be due to the establishment of an outreach satellite oncology service sites based at Eldoret where a significant number of patients are attended without the need of going to MTRH[42]. The distant referrals was an indication that majority of the patients have difficulty accessing cancer services.

Limitations
Our study had several limitations. First, causality could not be proven and we could not ascertain that a given cancer was actually caused by an infectious agent including the high number of HIV-associated malignancies at MTRH. Therefore, we used AF generated from other studies. Information on associations is sparse in Kenya which opens up new avenues for future research studies. Our study used a sampling faction of the total number of files available at the two hospitals as compared to studying all the files that could have influenced our results. For failure of not knowing the whole sampling frame, there was a probability of introducing selection bias by the use of con-

venient sampling method to select the files at MTRH. This could result in over or under-representation of the cases. Future studies should focus more on population based data or use proportions of individual cancers to calculate the sample size. Population based data would be more reflective of the overall population. Our choice of study population was influenced by the fact that cancer registration in Kenya was in its fairly early stages of development at the time the study was done, and the use of hospital based data seemed suitable. The percentage attributable to infections for certain cancers, such as BL, adult T-cell leukemia or bile duct cancers could not be calculated because we did not come across any cases of the first two cancers while the AF of bile duct was unobtainable.

Conclusion

Our study presented a picture of the burden of cancer and infection-attributable cancer from a hospital point of view. Despite the limitations, the role played by infectious agents in contributing to the overall cancer burden was highlighted. Controlling for the infectious agents could translate to a significant reduction in the cancer burden. Further research is warranted to prove causality between infection-attributable cancers and the infectious agents in Kenya as this may provide new avenues for effective cancer prevention.

Data availability

The data underlying this study is available from the Open Science Framework (OSF) Dataset 1: Burden of Cancer in Kenya; Types, Infection-Attributable and Trends: A National Referral Hospital Retrospective Survey https://doi.org/10.17605/OSF.IO/MD2PY[49]

Data is available under a CC0 1.0 universal license.

Supplementary material

Supplementary File 1: Data collection form

Click here to access the data.

References

1. Parkin DM, Bray F, Ferlay J, *et al.*: **Cancer in Africa 2012.** *Cancer Epidemiol Biomarkers Prev.* 2014; **23**(6): 953–966.
   PubMed Abstract | **Publisher Full Text**

2. Ministry of Health: **National Cancer Control Strategy 2017 – 2022.** 2017.
   Reference Source

3. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans: **Biological agents. Volume 100 B. A review of human carcinogens.** *IARC Monogr Eval Carcinog Risks Hum.* 2012; **100**(Pt B): 1–441.
   PubMed Abstract | **Free Full Text**

4. Plummer M, de Martel C, Vignat J, *et al.*: **Global burden of cancers attributable to infections in 2012: a synthetic analysis.** *Lancet Glob Health.* The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND license; 2016; **4**(9): e609–e616.
   PubMed Abstract | **Publisher Full Text**

5. Chinula L, Moses A, Gopal S: **HIV-associated malignancies in sub-Saharan Africa: progress, challenges, and opportunities.** *Curr Opin HIV AIDS.* 2017; **12**(1): 89–95.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

6. Kimanga DO, Ogola S, Umuro M, *et al.*: **Prevalence and incidence of HIV infection, trends, and risk factors among persons aged 15-64 years in Kenya: results from a nationally representative study.** *J Acquir Immune Defic Syndr.* 2014; **66 Suppl 1**: S13–26.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

7. Sasco AJ, Jaquet A, Boidin E, *et al.*: **The challenge of AIDS-related malignancies in sub-Saharan Africa.** Maartens G, editor. *PLoS One.* 2010; **5**(1): e8621.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

8. Kenya National Bureau of Statistics: **The 2009 Kenya Population & Housing. Counting our people for the Vision 2030.** *Population Distribution by Age Sex and Administrative Unit.* 2010; **IC**.
   Reference Source

9. Strother RM, Asirwa FC, Busakhala NB, *et al.*: **AMPATH-Oncology: A model for comprehensive cancer care in sub-Saharan Africa.** *J Cancer Policy.* 2013; **1**(3–4): e42–e48.
   Publisher Full Text

10. Charan J, Biswas T: **How to calculate sample size for different study designs in medical research?** *Indian J Psychol Med.* 2013; **35**(2): 121–6.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

11. Daniel WW: **Biostatistics: A Foundation for Analysis in the Health Sciences.** 7th ed. New York; 1999.
    Reference Source

12. Pourhoseingholi MA, Vahedi M, Rahimzadeh M: **Sample size calculation in medical studies.** *Gastroenterol Hepatol Bed Bench.* 2013; **6**(1): 14–17.
    PubMed Abstract | **Free Full Text**

13. Arya R, Antonisamy B, Kumar S: **Sample size estimation in prevalence studies.** *Indian J Pediatr.* 2012; **79**(11): 1482–8.
    PubMed Abstract | **Publisher Full Text**

14. Parkin DM: **The global health burden of infection-associated cancers in the year 2002.** *Int J Cancer.* 2006; **118**(12): 3030–3044.
    PubMed Abstract | **Publisher Full Text**

15. Odutola M, Jedy-Agba EE, Dareng EO, *et al.*: **Burden of Cancers Attributable to Infectious Agents in Nigeria: 2012-2014.** *Front Oncol.* 2016; **6**: 216.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

16. Moussavou PB, Koumakpayi IH, Nkili-Meyong AA, *et al.*: **Molecular analysis of human Papillomavirus detected among women positive for cervical lesions by visual inspection with acetic acid/Lugol's iodine (VIA/VILI) in Libreville, Gabon.** *Infect Agent Cancer.* 2016; **11**(1): 50.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

17. Ebrahim S, Mndende XK, Kharsany AB, *et al.*: **High Burden of Human Papillomavirus (HPV) Infection among Young Women in KwaZulu-Natal, South**

Africa. Meyers C, editor. *PLoS One.* 2016; **11**(1): e0146603.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

18. Denny L, Adewole I, Anorlu R, *et al.*: **Human papillomavirus prevalence and type distribution in invasive cervical cancer in sub-Saharan Africa.** *Int J Cancer.* 2014; **134**(6): 1389–98.
PubMed Abstract | **Publisher Full Text**

19. Walboomers JM, Jacobs MV, Manos MM, *et al.*: **Human papillomavirus is a necessary cause of invasive cervical cancer worldwide.** *J Pathol.* 1999; **189**(1): 12–9.
PubMed Abstract | **Publisher Full Text**

20. de Martel C, Plummer M, Vignat J, *et al.*: **Worldwide burden of cancer attributable to HPV by site, country and HPV type.** *Int J Cancer.* 2017; **141**(4): 664–70.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

21. Rantshabeng PS, Moyo S, Moraka NO, *et al.*: **Prevalence of oncogenic human papillomavirus genotypes in patients diagnosed with anogenital malignancies in Botswana.** *BMC Infect Dis.* 2017; **17**(1): 731.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

22. Tesfalul M, Simbiri K, Wheat CM, *et al.*: **Oncogenic viral prevalence in invasive vulvar cancer specimens from human immunodeficiency virus-positive and -negative women in Botswana.** *Int J Gynecol Cancer.* 2014; **24**(4): 758–65.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

23. Mboumba Bouassa RS, Mbeko Simaleko M, Camengo SP, *et al.*: **Unusual and unique distribution of anal high-risk human papillomavirus (HR-HPV) among men who have sex with men living in the Central African Republic.** Mossong J, editor. *PLoS One.* 2018; **13**(5): e0197845.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

24. Davidson CL, Richter KL, Van der Linde M, *et al.*: **Prevalence of oral and oropharyngeal human papillomavirus in a sample of South African men: a pilot study.** *S Afr Med J.* 2014; **104**(5): 358–61.
PubMed Abstract | **Publisher Full Text**

25. Kaba G, Dzudzor B, Gyasi RK, *et al.*: **Human papillomavirus genotypes in a subset of head and neck squamous cell carcinoma.** *West Afr J Med.* 2014; **33**(2): 121–4.
PubMed Abstract

26. Senba M, Buziba N, Mori N, *et al.*: **Human papillomavirus infection induces NF-kB activation in cervical cancer: A comparison with penile cancer.** *Oncol Lett.* 2011; **2**(1): 65–8.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

27. Senba M, Buziba N, Mori N, *et al.*: **Increased prevalence of Kaposi´s sarcoma-associated herpesvirus in the Kaposi´s sarcoma-endemic area of western Kenya in 1981-2000.** *Acta Virol.* 2011; **55**(2): 161–4.
PubMed Abstract | **Publisher Full Text**

28. Kafita D, Kaile T, Malyangu E, *et al.*: **Evidence of EBV infection in lymphomas diagnosed in Lusaka, Zambia.** *Pan Afr Med J.* 2018; **29**: 181.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

29. Westmoreland KD, Stanley CC, Montgomery ND, *et al.*: **Hodgkin lymphoma, HIV, and Epstein-Barr virus in Malawi: Longitudinal results from the Kamuzu Central Hospital Lymphoma study.** *Pediatr Blood Cancer.* 2017; **64**(5): e26302.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

30. Weinreb M, Day PJ, Niggli F, *et al.*: **The consistent association between Epstein-Barr virus and Hodgkin's disease in children in Kenya.** *Blood.* 1996; **87**(9): 3828–36.
PubMed Abstract | **Publisher Full Text**

31. Orem J, Mbidde EK, Lambert B, *et al.*: **Burkitt's lymphoma in Africa, a review of the epidemiology and etiology.** *Afr Health Sci.* 2007; **7**(3): 166–75.
PubMed Abstract | **Free Full Text**

32. Kirk GD, Lesi OA, Mendy M, *et al.*: **The Gambia Liver Cancer Study: Infection with hepatitis B and C and the risk of hepatocellular carcinoma in West Africa.**

33. Maucort-Boulch D, de Martel C, Franceschi S, *et al.*: **Fraction and incidence of liver cancer attributable to hepatitis B and C viruses worldwide.** *Int J Cancer.* 2018; **142**(12): 2471–7.
PubMed Abstract | **Publisher Full Text**

34. Bello A, Umar A, Borodo M: **Prevalence and risk factors for helicobacter pylori infection in gastroduodenal diseases in Kano, Nigeria.** *Afr J Med Health Sci.* 2018; **17**(1): 41–46.
Publisher Full Text

35. Newton R, Ziegler JL, Casabonne D, *et al.*: *Helicobacter pylori* **and cancer among adults in Uganda.** *Infect Agent Cancer.* 2006; **1**: 5.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

36. Groeneveld AE, Marszalek WW, Heyns CF: **Bladder cancer in various population groups in the greater Durban area of KwaZulu-Natal, South Africa.** *Br J Urol.* 1996; **78**(2): 205–8.
PubMed Abstract | **Publisher Full Text**

37. Parker RK, Dawsey SM, Abnet CC, *et al.*: **Frequent occurence of esophageal cancer in young people in Western Kenya.** *NIH.* 2012; **34**: 13–23.

38. Korir A, Okerosi N, Ronoh V, *et al.*: **Incidence of cancer in Nairobi, Kenya (2004-2008).** *Int J Cancer.* 2015; **137**(9): 2053–9.
PubMed Abstract | **Publisher Full Text**

39. Jemal A, Bray F, Ferlay J: **Global Cancer Statistics: 2011.** *CA Cancer J Clin.* 2011; **61**(2): 69–90.
PubMed Abstract | **Publisher Full Text**

40. MOPHS & MOMS: **Strategic Plan.** 2011-2016 National Cancer Control Strategies. 2010.
Reference Source

41. Patel K, Wakhisi J, Mining S, *et al.*: **Esophageal Cancer, the Topmost Cancer at MTRH in the Rift Valley, Kenya, and Its Potential Risk Factors.** *ISRN Oncol.* 2013; **2013**: 503249.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

42. Tenge CN, Kuremu RT, Buziba NG, *et al.*: **Burden and pattern of cancer in western Kenya.** *East Afr Med J.* 2009; **86**(1): 7–10.
PubMed Abstract | **Publisher Full Text**

43. Ferlay J, Shin HR, Bray F, *et al.*: **Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008.** *Int J Cancer.* 2010; **127**(12): 2893–2917.
PubMed Abstract | **Publisher Full Text**

44. Pagano JS, Blaser M, Buendia MA, *et al.*: **Infectious agents and cancer: criteria for a causal relation.** *Semin Cancer Biol.* 2004; **14**(6): 453–471.
PubMed Abstract | **Publisher Full Text**

45. Thun MJ, DeLancey JO, Center MM, *et al.*: **The global burden of cancer: priorities for prevention.** *Carcinogenesis.* 2009; **31**(1): 100–110.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

46. Rochford R, Korir A, Newton R: **Viral-associated malignancies in Africa: are viruses 'infectious traces' or 'dominant drivers'?** *Curr Opin Virol.* 2016; **20**: 28–33.
PubMed Abstract | **Publisher Full Text**

47. Plummer M, Franceschi S, Muñoz N: **Epidemiology of gastric cancer.** *IARC Sci Publ.* 2004; (157): 311–26.
PubMed Abstract

48. Torre LA, Bray F, Siegel RL, *et al.*: **Global cancer statistics, 2012.** *CA Cancer J Clin.* 2015; **65**(2): 87–108.
PubMed Abstract | **Publisher Full Text**

49. Macharia LW: **Burden of Cancer in Kenya; Types, Infection-Attributable and Trends: A National Referral Hospital Retrospective Survey.** 2018.
http://www.doi.org/10.17605/OSF.IO/MD2PY

Hepatology. 2004; **39**(1): 211–9.
PubMed Abstract | **Publisher Full Text**

Open Peer Review

Current Peer Review Status: ✗ ✗ ✓ ✓

Version 5

Reviewer Report 27 November 2019

https://doi.org/10.21956/aasopenres.14118.r27260

✓ **Delphine Casabonne**

[1] Unit of Infections and Cancer, Cancer Epidemiology Research Programme, Catalan Institute of Oncology, Barcelona, Spain
[2] Epidemiology and Public Health Networking Biomedical Research Centre (CIBERESP), Madrid, Spain

The issue on the data collection and availability of the HIV status of the patients should be also mentioned in the Methods.

The sentence below in the overall Discussion is not clear:
"The information on the HIV status of the patients from whose files were used, was not obtained to accurately link the associations."
Could you rephrase it? Here is a suggestion:
"Unfortunately, the information on the HIV status of the patients from whose files were used was not collected and therefore it was not possible to accurately link a cancer site to HIV infection."

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Cancer epidemiology and infections.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 4

Reviewer Report 25 October 2019

https://doi.org/10.21956/aasopenres.14096.r27194

? **Delphine Casabonne**

[1] Unit of Infections and Cancer, Cancer Epidemiology Research Programme, Catalan Institute of Oncology, Barcelona, Spain

[2] Epidemiology and Public Health Networking Biomedical Research Centre (CIBERESP), Madrid, Spain

The authors addressed most of the issues. However, further points should be corrected and/or specified:

Main comments:

1. The authors mentioned that "*The files for 2008 to 2011 were obtained from the oncology centre at the Academic Model for the Prevention and Treatment of HIV/ AIDS (AMPATH)*". The authors should clarify if all patients selected in this center for these years (81% of the sample for MTRH) were HIV+ or the authors should report if the HIV status of these patients was unknown. This information should be added to the manuscript to help interpreting the higher and different cancer proportions in MTRH and KNH.

2. In Table 2 (Previous prevalence studies on infectious agents and cancers): The virus HTLV should be associated with T-cell lymphoma and not Burkitt lymphoma. Burkitt lymphoma should be associated with EBV.

3. Table 6 (Cancer cases by age groups in Moi Teaching and Referral Hospital (MTRH) from 2008–2012) should be checked again. For instance the total number of KS across the age for MTRH is not correct (N=2) and it should be N=93. The authors should check the tables again.

Minor comments:

1. In the abstract: Replace "stomach cancer 26 (5.2%)" by "stomach cancer (26, 5.2%)".

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Cancer epidemiology and infections.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 02 Nov 2019

Lucy Wanjiku Macharia, University of Nairobi, Nairobi, Kenya

Dear Dr. Delphine Casabonne,
The authors are thankful to you for reviewing our manuscript and for the suggestions. We have modified the manuscript and double-checked the tables to incorporate the suggestions. Kindly find below the point-by-point response to the queries.

The authors mentioned that "*The files for 2008 to 2011 were obtained from the oncology

centre at the Academic Model for the Prevention and Treatment of HIV/ AIDS (AMPATH)". **The authors should clarify if all patients selected in this center for these years (81% of the sample for MTRH) were HIV+ or the authors should report if the HIV status of these patients was unknown. This information should be added to the manuscript to help interpreting the higher and different cancer proportions in MTRH and KNH.**
We thank the reviewer for the question and the suggestion. We do not know the HIV information of the patients whose files were obtained from AMPATH. We have highlighted the suggestion in the discussion and limitations section.

In Table 2 (Previous prevalence studies on infectious agents and cancers): The virus HTLV should be associated with T-cell lymphoma and not Burkitt lymphoma. Burkitt lymphoma should be associated with EBV.
We agree with the reviewer's observation. We have modified Table 2 and corrected the misclassification. We are thankful for the observation.

Table 6 (Cancer cases by age groups in Moi Teaching and Referral Hospital (MTRH) from 2008–2012) should be checked again. For instance the total number of KS across the age for MTRH is not correct (N=2) and it should be N=93. The authors should check the tables again.
We are grateful to the reviewer for the keen observation. We have checked the tables and made the necessary modifications to Tables 5 and 6. We are thankful for the suggestion.

Minor comments:
In the abstract: Replace "stomach cancer 26 (5.2%)" by "stomach cancer (26, 5.2%)".
We thank the reviewer for the observation. We have modified the statement highlighted above in the abstract section.

Competing Interests: The authors have no competing interests to declare.

Version 3

Reviewer Report 30 August 2019

https://doi.org/10.21956/aasopenres.14030.r27129

Kalina Duncan (iD)
Center for Global Health, U.S. National Cancer Institute, Rockville, MD, USA

In the methods section: You note there are "few" radiation machines but more may be coming. It would help the context here to report the number and perhaps remove the bit about more coming unless it can

be cited in some way? Or if you have provide more specificity about how and where those machines are intended to go, that will be helpful.

Also in Methods: you introduce AMPATH without clarifying what it is. I would recommend providing a sentance of background of their role at MTRH.

In Discussion: At KNH, there is a variability in breast cancer proportions that may be interesting to look at. From 2008-2011, the proportion was between 10-15%. In 2012, it drops to below 5%. What happened that year? Were there really fewer cases? Was there an issue with data quality (more incomplete records, etc.)?

In Discussion: You have this line: "The high number of Kaposi's sarcoma, non-Hodgkin's lymphoma and cervical cancer could have been influenced by the association of the cancers with HIV" regarding the high number of HIV-related cases at MTRH. I think this actually requires further discussion, as it pushes the infection-related rates way above average. The presence of the AMPATH program, and its specialization in treating patients with HIV should be noted in discussion, as explanation. Because these are hospital-based data, this is especially important to note as this hospital attracts many HIV positive patients, most likely.

I would consider an examination and a citation to the article listed below, "Viral-associated malignancies in Africa: are viruses 'infectious traces' or 'dominant drivers'?"[1].

References
1. Rochford R, Korir A, Newton R: Viral-associated malignancies in Africa: are viruses 'infectious traces' or 'dominant drivers'?. *Curr Opin Virol*. **20**: 28-33 PubMed Abstract | Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: My areas of expertise include evaluative research methods, cancer control, data analysis, and health systems. I have noted that I would prefer is a qualified statistician or epidemiologist confirmed the analysis and conclusions section of this paper. While it reads as very acceptable to me, I do not consider myself an expert in this area.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 03 Oct 2019

Lucy Wanjiku Macharia, University of Nairobi, Nairobi, Kenya

Dear Dr. Kalina Duncan,
Thank you for reading our manuscript, for the suggestions and the approval status. Kindly find below the response to the queries.

Comments

In the methods section: You note there are "few" radiation machines but more may be coming. It would help the context here to report the number and perhaps remove the bit about more coming unless it can be cited in some way? Or if you have provide more specificity about how and where those machines are intended to go, that will be helpful.
We thank the reviewer for the comment. We have edited the statement, included the number of radiation machines and removed "the more are coming" bit due to lack of a proper citation.

Also in Methods: you introduce AMPATH without clarifying what it is. I would recommend providing a sentence of background of their role at MTRH.
We are thankful for the suggestion. We have included a sentence to connect AMPATH and MTRH.

In Discussion: At KNH, there is variability in breast cancer proportions that may be interesting to look at. From 2008-2011, the proportion was between 10-15%. In 2012, it drops to below 5%. What happened that year? Were there really fewer cases? Was there an issue with data quality (more incomplete records, etc.)?
We agree with the reviewer's concern. The variability from the proportions of the breast cancer at KNH could have been influenced by various factors including the sample size used as compared to sampling the whole frame which was not feasible due to cost constraints. Similarly, maybe there were high numbers of patients with breast cancer that visited the hospital in certain years as compared to others. We have removed the proportion section from the results and discussion in line too with the suggestion from the previous reviewer. We are thankful for the keen observation.

In Discussion: You have this line: "The high number of Kaposi's sarcoma, non-Hodgkin's lymphoma and cervical cancer could have been influenced by the association of the cancers with HIV" regarding the high number of HIV-related cases at MTRH. I think this actually requires further discussion, as it pushes the infection-related rates way above average. The presence of the AMPATH program, and its specialization in treating patients with HIV should be noted in discussion, as explanation. Because these are hospital-based data, this is especially important to note as this hospital attracts many HIV positive patients, most likely. I would consider an examination and a citation to the article listed below, "Viral-associated malignancies in Africa: are viruses 'infectious traces' or 'dominant drivers'?".

References
1. Rochford R, Korir A, Newton R: Viral-associated malignancies in Africa: are viruses 'infectious traces' or 'dominant drivers'?. *Curr Opin Virol*. **20: 28-33 PubMed**

We thank the reviewer for the suggestion. We have modified the discussion section to incorporate the new suggestions and reference.

Competing Interests: The authors have no competing interests to disclose.

Reviewer Report 17 April 2019

https://doi.org/10.21956/aasopenres.14030.r26842

? **Delphine Casabonne**

[1] Unit of Infections and Cancer, Cancer Epidemiology Research Programme, Catalan Institute of Oncology, Barcelona, Spain
[2] Epidemiology and Public Health Networking Biomedical Research Centre (CIBERESP), Madrid, Spain

The authors selected 500 (partially randomly selected) files of cancer patients from two National referral hospitals in Kenya and reported the cancer prevalence. The authors then applied attributable fraction (AF) for infection-associated cancers from published literature (meta-analysis, prevalence data and relative risk estimates from previous studies and WHO report). The authors estimated that 28% and 44% of the observed cancers were attributable to infections in KNM and MTRH hospitals, respectively.

Title:
- Please add in the title the words "adult" and the years "2008-2012".

Abstract:
- Please indicate in the abstract and methods not only the selected years (2008-2012) but also the months.

Methods:
- Please indicate in the abstract and methods not only the selected years (2008-2012) but also the months.

- *Sample size calculation*: This is a case-study and the prevalence of cancers is not the prevalence of cancers in Kenya but the proportion of a specific cancer subtypes within the total number of cancer cases in the hospitals. Proportions estimated in Kenya in different published reports could have been used. For example, in https://gco.iarc.fr/today, the cancer prevalence by subtypes in Kenya for 2018 were estimated to be: breast: 13.6%, cervix: 11.5%, oesophagus: 8.4%, prostate: 5.8%, colorectum: 4.8%, etc. It might therefore be more informative to give the number of patients needed from different cancer subtype proportions (p=0.01, 0.05, 0.10, 0.15 and 0.20).

- The details of the calculations are not needed in the text. Please delete these two sections of the text (page 4):
  a) "Therefore, n=1.96^2..." to "n=384".
  b) In the text from "N1= 3168..." to "... ∑ni=500". This information is provided in Table 1.

- More details should be given in relation to the selection of the records. What is the "non-response allowance" in this cross-sectional design? Were all 500 selected records available for both hospitals? If a selected record had missing data (based on the pre-designed questionnaire) was it replaced by another randomly selected record? This part is unclear; please clarify.

- For MTRH, the origin of the 2008-2011 files is unclear. Which "convenient sampling" was used for the 390 files of MTRH? Could the authors explain better the procedure of selection for this hospital? How could biases be introduced for interpretability? How is representativeness kept?

- Please modify the methods section accordingly.

Results:
- As mentioned by previous reviewers, the higher prevalence of KS and NHL in MTRH might reflect the tight relationship of this hospital with AMPATH/HIV. Also, the convenience sampling method used for the selection of the records for this hospital might accentuate the number of KS and NHL cases observed in this hospital. Please add to the current Table 7 the estimates of the 2 hospitals altogether and mention the overall 36% of cancers attributable to infections agents.

- *Figures 2 & 3*: The titles and the figures are unclear and inconsistent with the text. From the results section, in KNH hospital, the first commonest cancers in males were prostate (11.5%) and laryngeal (9.5%) cancers. Why is colorectal cancer plotted in Figure 2? In page 9, why do the authors mention access to treatment in relation to the histograms in the result section? Also, the different units in the y-axis are misleading. Finally, due to the small sample size by year and the use of convenience sampling for MTRH, I would suggest not to show Figures 2 and 3. Please delete the text in the results section associated with these Figures (page 9, "Proportions of the top most cancer cases").

Minor comments:
- Since penile cancer is one of the cancer sites associated with infections please provide the number of cases of penile cancers that were observed in each hospital. If none were observed please provide this information in the text.

- The authors mentioned 11 infections in page 9 but there are less infections in Table 2. Please modify accordingly.

- Can the referral hospital for KNH be MTRH? And vice versa?

Discussion:
- As mentioned by previous reviewers, the authors gave too much emphasis on comparing the two hospital results. The presence of KS and NHL in the top four cancers for MTRH should be put in perspective with the hospital origin of the patients (HIV center). The authors should discuss this point in more detail. Given the method of selection of the patients, KNH might reflect better the overall prevalence.

- In the limitations, the representativeness of the selected patients should be addressed. How complete are the hospital registration of the patients? Could any biases be introduced at this stage? Are some cancer subtypes less likely to be recorded in these hospitals? Is misdiagnosis a potential issue?

Minor comments:

- In page 13, "Similarly an incidence using the Nairobi cancer registry..." , please correct this sentence. Do you mean incidence rates?

- Please delete all references to the result part (Tables X, figures Y, ...) in the Discussion section.

Summary:

- This article describes the cancer prevalence and cancers attributable to infections in two hospitals from Kenya. Due to the selection methods there are important issues to take into account when interpreting the results - in particular in relation to HIV-associated cancers - that have not been addressed adequately.

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Cancer epidemiology and infections.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 03 Oct 2019

Lucy Wanjiku Macharia, University of Nairobi, Nairobi, Kenya

Dear Dr. Delphine Casabonne,
We are thankful to you for reading our manuscript, for the positive criticism and the suggestions. Kindly find below the answers to the questions highlighted.

Title:
Please add in the title the words "adult" and the years "2008-2012".
We have added the words "adult" and years to the title. Thank you for the suggestion.

Abstract:
Please indicate in the abstract and methods not only the selected years (2008-2012) but also the months.
We have modified the abstract and method section to include the years and months. We are grateful for the suggestion.

Methods:
Please indicate in the abstract and methods not only the selected years (2008-2012) but also the months.
We have included the new suggestions to the abstract and methods section.
*Sample size calculation***: This is a case-study and the prevalence of cancers is not the prevalence of cancers in Kenya but the proportion of a specific cancer subtypes within the total number of cancer cases in the hospitals. Proportions estimated in Kenya in different published reports could have been used. For example, in https://gco.iarc.fr/today, the cancer prevalence by subtypes in Kenya for 2018 were estimated to be: breast: 13.6%, cervix: 11.5%, oesophagus: 8.4%, prostate: 5.8%, colorectum: 4.8%, etc. It might therefore be more informative to give the number of patients needed from different cancer subtype proportions (p=0.01, 0.05, 0.10, 0.15 and 0.20).**

We are thankful for the comment. Our objective was to sample 500 files from each hospital and identify how many cancer types were in the 500 selected files as a representative sample. This is because we could not afford to sample the whole population. For this, the *p* we used in the sample size calculation was *p* for expected prevalence or proportion or estimated proportion of a disease (Cancer) as explained by references (9 to 11) which we assumed a p of 50% for not knowing the prevalence at the time which is also acceptable as explained by Arya et. al., 2012 (Reference 12). This is the sample size that was also approved by the ethical committees. To apply the proportions of individual cancer type, we would need to know the information on the total numbers of the specific cancers in both the hospitals, information that was unavailable.
Lastly, as we cannot change the sample size calculation strategy at this stage, we can only highlight how we arrived at the sample size of 500 files and highlight the new suggestion and a potential limitation. We are thankful for the suggestion.

The details of the calculations are not needed in the text. Please delete these two sections of the text (page 4):
a) "Therefore, n=1.96^2..." to "n=384".
b) In the text from "N1= 3168..." to "... Σni=500". This information is provided in Table 1.
The calculation details highlighted above have been removed from the text. We are thankful for the suggestion

More details should be given in relation to the selection of the records. What is the "non-response allowance" in this cross-sectional design?
We have inserted a statement to clarify the meaning of the "non-response allowance" in the context of our study. We appreciate the comment.

Were all 500 selected records available for both hospitals? If a selected record had missing data (based on the pre-designed questionnaire) was it replaced by another randomly selected record? This part is unclear; please clarify.

Yes. All the targeted 500 files were available from both hospitals by the end of the data collection period. In case a selected file was found incomplete, it was replaced by the next file following it after randomization. We thank the reviewer for the question.

For MTRH, the origin of the 2008-2011 files is unclear. Which "convenient sampling" was used for the 390 files of MTRH? Could the authors explain better the procedure of selection for this hospital? How could biases be introduced for interpretability? How is representativeness kept? Please modify the methods section accordingly.

It was difficult to apply probability randomization as the total sampling frame could not be established. But from the total number of files that we could access from each year, the study files were chosen in a "semi-randomized" manner based on the total number of files that we could access each year and not the sampling frame. Being a non-probability sampling method, bias could be introduced. The bias will be mentioned in the limitations section. We are thankful for the comment.

Results:

As mentioned by previous reviewers, the higher prevalence of KS and NHL in MTRH might reflect the tight relationship of this hospital with AMPATH/HIV. Also, the convenience sampling method used for the selection of the records for this hospital might accentuate the number of KS and NHL cases observed in this hospital.

We agree with the reviewer's concern. We have highlighted this as a possible source of the study's limitation and modified the explanation in the discussion section as suggested also by reviewer 4. We appreciate the observation.

Please add to the current Table 7 the estimates of the 2 hospitals altogether and mention the overall 36% of cancers attributable to infectious agents.

We have added the estimates of the 2 hospitals to table 7. We appreciate the suggestion.

*Figures 2 & 3***: The titles and the figures are unclear and inconsistent with the text. From the results section, in KNH hospital, the first commonest cancers in males were prostate (11.5%) and laryngeal (9.5%) cancers. Why is colorectal cancer plotted in Figure 2? In page 9, why do the authors mention access to treatment in relation to the histograms in the result section? Also, the different units in the y-axis are misleading. Finally, due to the small sample size by year and the use of convenience sampling for MTRH, I would suggest not to show Figures 2 and 3. Please delete the text in the results section associated with these Figures (page 9, "Proportions of the top most cancer cases").**

We have deleted Figures 2 and 3 together with their associated texts in the manuscript document. We thank the reviewer for the suggestion.

*Minor comments***:**

Since penile cancer is one of the cancer sites associated with infections please provide the number of cases of penile cancers that were observed in each hospital. If none were observed please provide this information in the text.

There were two penile cases in MTRH. The numbers have been included in table 7 in the males' section. Thank you for the suggestion.

The authors mentioned 11 infections in page 9 but there are less infections in Table 2. Please modify accordingly.
We have included modified table 2 accordingly. Some infectious agents have been listed together under one cancer type. We appreciate the observation.

Can the referral hospital for KNH be MTRH? And vice versa?
I strongly believe that both hospitals can receive patients referred from the other. However, I lack the evidence to support my opinion.

Discussion:
As mentioned by previous reviewers, the authors gave too much emphasis on comparing the two hospital results. The presence of KS and NHL in the top four cancers for MTRH should be put in perspective with the hospital origin of the patients (HIV center). The authors should discuss this point in more detail. Given the method of selection of the patients, KNH might reflect better the overall prevalence.
We thank the reviewer for the comment. We have modified the infectious agents section where we have highlighted that the high numbers of KS, NHL and cervical cancer could also have been influenced by AMPATH-Oncology as the source of data for MRTH. Selection bias has been addressed in the limitation section.

In the limitations, the representativeness of the selected patients should be addressed. How complete are the hospital registration of the patients? Could any biases be introduced at this stage? Are some cancer subtypes less likely to be recorded in these hospitals? Is misdiagnosis a potential issue?
We thank the reviewer for the concern. We have addressed the selection bias in the limitation section. For the 500 files that we collected, they all had the information we were aiming to get. During the data collection period, if a file had missing information, it was immediately excluded and replaced with the one following if after randomization. Convenient sampling could have led to under-representation or over-representation which we have included in the limitations section. We come across files with incomplete information or missing or even unclear diagnosis but these were excluded from the study.

Minor comments:
In page 13, "Similarly an incidence using the Nairobi cancer registry...", please correct this sentence. Do you mean incidence rates?.
We are thankful for the observation. We meant incidence rates and have modified the text.
Please delete all references to the result part (Tables X, figures Y, ...) in the Discussion section.
Thank you for the suggestions. We have deleted the "Tables X/Figures Y" references in the discussion section.

Summary
This article describes the cancer prevalence and cancers attributable to infections in two hospitals from Kenya. Due to the selection methods there are important issues to take into account when interpreting the results - in particular in relation to HIV-associated cancers - that have not been addressed adequately.
We are thankful to the reviewer for the suggestions and we have modified the text in agreement with the suggestions. We have included a text to explain or justify the high numbers of HIV-

associated malignancies and a new reference. We have also highlighted the bias that could have resulted from convenient sampling or using data from AMPATH

Competing Interests: The authors have no competing interests to declare.

Version 2

Reviewer Report 31 December 2018

https://doi.org/10.21956/aasopenres.13993.r26698

❌ **Maria Paula Curado**

Postgraduate Program in Health Sciences, Federal University of Goia´s, Goiânia, Brazil

Title

The title is not clear with the aims ” Cancer in Kenya data from two National referral hospitals "

It is not possible to have trends as it is a cross sectional study in two hospitals in Kenya.

Abstract

No comments.

Introduction

No comments.

Methods

Why were patients aged less 18 was excluded from the study?
Why was clinical diagnosis not included?
This type of study has limitations it can give more information about the hospital activities. More description about the hospital facilities (number of beds, medical, nurses, etc, radiotherapy etc ..)

Results

Figure 1 please add the period 2008-2012 sample (500/500).
Figure 2 this is not trends it is a proportion of cases treated in the period, it means that more patients has access to treatment in 2008 to 2011 and less in 2012. I suggest to use a histogram instead of lines for this figure and change the title.
Why were only two cancer sites used?

Same observations for figure 3.

Discussion

I was expecting a comparison of your data with Eldoret and Nairobi cancer registry.
The discussion is long for the data, please make it shorter.
ASR are only possible to generate if you have a population based cancer registry database.

References
1. Jedy-Agba EE, Curado MP, Oga E, Samaila MO, Ezeome ER, Obiorah C, Erinomo OO, Ekanem IO, Uka C, Mayun A, Afolayan EA, Abiodun P, Olasode BJ, Omonisi A, Otu T, Osinubi P, Dakum P, Blattner W, Adebamowo CA: The role of hospital-based cancer registries in low and middle income countries-The Nigerian Case Study.*Cancer Epidemiol*. 2012; **36** (5): 430-5 PubMed Abstract | Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
No

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: epidemiology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 25 Feb 2019
Lucy Wanjiku Macharia, University of Nairobi, Nairobi, Kenya

 Authors' Response to Reviewers' Comments:V2

Journal: AAS Open Research
Manuscript #: 12910

Manuscript Title: Cancer in Kenya: types and infection-attributable. Data from two national referral hospitals

We thank the reviewer for having reviewed our manuscript, for the suggestions and for the positive criticism. We have revised the manuscript in agreement with the suggestions. The discussion section has been edited to shorten the section. We hope that this new version of the manuscript can be considered for approval. Kindly find below point-by-point response to the comments and queries.

Sincerely,

Lucy Wanjiku Macharia

Comments
1: Title
"The title is not clear with the aims" Cancer in Kenya data from two National referral hospitals"
It is not possible to have trends as it is a cross sectional study in two hospitals in Kenya."
Response:
We have modified the title to make the aim clear and have removed the "trends" from the title. We are thankful for the suggestions

2: Abstract
"No comments".
Response:
We are thankful for the positive feedback

3: Introduction
"No comments".
Response:
We are grateful for the positive feedback.

4: Methods
4.1 "Why were patients aged less than 18 excluded from the study."
4.2 "Why was clinical diagnosis not included."
4.3 "This type of study has limitations it cannot give more information about the hospital activities. More description about the hospital facilities (number of beds, medical, nurses, etc, radiotherapy etc."

Responses:
We are thankful for the questions. There was no objective exclusion of patients under the age of 18 years. It was mainly to circumvent possible ethical debates revolving the use of data from under age group.
All the cases had a clinical diagnosis but for them to be included in the study, a confirmatory diagnosis besides clinical, was prerequisite.
We agree with the reviewer's concern. An improved description of the hospital facilities has been provided in the methodology section of the new manuscript. However, to get the exact information additional permission would need to be sought from the Hospitals.

5: Results

5.1. " Figure 1 please add the period 2008-2012 sample (500/500).

5.2. "Figure 2 this is not trends it is a proportion of cases treated in the period, it means that more patients have access to treatment in 2008 to 2011 and less in 2012. I suggest to use a histogram instead of lines for this figure and change the title.

5.3. "Why were only two cancer sites used?"

5.4. "Same observations for figure 3"

Responses:

We are thankful for the suggestions.

The period 2008-2012 sample (500/500) has been added in Figure 1.

We agree with the reviewer's concern and have modified Fig. 2 and Fig. 3 into a histogram and changed their titles.

We used two cancer sites as they are the only ones who gave permission to access their data among all the four referral hospitals targeted at the time of the study. The reasons for refusal by the other hospitals are unknown.

6: Discussion

I was expecting a comparison of your data with Eldoret and Nairobi cancer registry. The discussion is long for the data, please make it shorter. ASR are only possible to generate if you have a population based cancer registry database.

We are grateful for the suggestion. The discussion section has been modified to shorten the section. There are limited published studies on "Cancer in Kenya" and the few mentioned were among the few obtainable.

Specifically, Reference 37 used the Nairobi Cancer Registry data

Reference 41 used the Eldoret Cancer Registry data

Reference 40 The study was done at MTRH

Reference 39 is a Kenya National cancer control strategy report

Competing Interests: No competing interests were disclosed.

Version 1

Reviewer Report 01 October 2018

https://doi.org/10.21956/aasopenres.13980.r26613

Clement A Adebamowo 🆔

[1] Greenebaum Comprehensive Cancer Center, Institute of Human Virology, University of Maryland School of Medicine, Baltimore, MD, USA

[2] Institute of Human Virology Nigeria, Garki, Abuja, Nigeria

[3] Center for Bioethics and Research, Ibadan, Nigeria

Michael Odutola

Institute of Human Virology Nigeria, Garki, Abuja, Nigeria

Overview

This paper provided information on the prevalence of cancers and the proportion of cancers that were attributable to infections in Kenyatta National Hospital (KNH) and Moi Teaching and Referral Hospital (MTRH), Kenya over a 5 year period from 2008 to 2012. These 2 centers are the oldest and largest national referral hospitals in Kenya. The authors randomly selected some cancer cases reported at these centers within the study period and reviewed the charts for cancer types and numbers of cancers attributable to infections. The study reported that 40.0% to 53.2% of cancers that were seen in these referral hospitals were attributable to infections.

Title

In this paper, the numbers and proportions of cancers associated with infections were presented and not the proportion of cancers attributable to infections. In order to obtain the proportion of cancers that were attributable to infections, the author would need to apply the formula for Population Attributable Fraction (PAF) which was not done in this paper.

Introduction

The introduction is repetitive and should be edited. Separate paragraphs should be devoted to those infections that are established as causes of cancer while those will weaker associations should be in a separate paragraph.

The role of HIV infection in cancer etiology, prevalence and incidence in Kenya deserves more discussion than just mentioning HIV in a list of pathogens.

Previous studies of infection-associated cancers in Kenya, East Africa or Sub-Saharan Africa should be discussed briefly to provide more introduction to the field and establish why this study was necessary and what it hopes to accomplish.

Methods

It would be more informative to describe the demographic characteristics of the catchment areas, the pattern of healthcare services including other hospitals, and referral systems for the two hospitals included in this study. While these are the biggest hospitals, they may not see the most cancer patients if specialized cancer hospitals are located in the same region or city.

The rationale, application, and choice of sample size calculation are unclear. This is a retrospective case series. The researchers should analyze all the case records that they can find irrespective of sample size calculation. Further, it is highly unlikely that the prevalence of cancer in adult Kenyan population is 50%.

The justification for randomly selecting case notes to review is not clear. Rather than statistically estimate the number of files, why not obtain the number of patients seen and the number who have a cancer diagnosis directly from the institutions? If this is not possible, the authors should say so.

The randomization procedure is not clear. The authors state that "the first record was selected randomly every year". What is the "first record"? Record of the first patient seen in the institution within a specific year? The first record selected from among all the patients' records for a specific year? If the latter how was the selection done? Were all the records assigned numbers and random number generators used to

select records?

Do the authors have an objective assessment of the completeness of the records of these two institutions?

What is the relationship of the two institutions that refused to grant permission for this study and their locations with the institutions that gave permission for the study? A map showing this information would be informative for international readers

The analytic methods used in the paper and described in the data analysis section should be described better.

Results

The authors should be precise in reporting results. "Around 4304 inpatient cancer cases were available in MTRH" is imprecise and unacceptable.

There is a clear difference in the cases of cancer at the KNH compared to MTRH, most probably because one site hosts AMPATH HIV treatment and prevention programs. It is therefore not justifiable to make too much of the differences in the proportion of cancers seen in these two institutions and attempt to extrapolate that to the general population.

Discussion

The interpretation of the result is not justifiable based on the data, methods, and analyses.

Proportions of cancers presenting in specific institutions may not reflect population level epidemiology of the cancers. See Chapter 1 of the Cancer in Africa (Ed) Max Parkin et al.

The authors suggest that their data shows the incidence of the cancers that they described but this is not correct.
The authors can describe the pattern of cancers presenting to each of these hospitals over a period of time, but this is not "incidence" of those cancers in the population. The pattern of presentation of cancers to specific hospitals is influenced by many factors. An example is illustrated in their paper where the pattern of cancers presenting at the AMPATH related hospital reflects the focus on HIV treatment and prevention at the institution.

General comments

1. There are several grammatical and typographical errors in the paper. For example in the Discussion section, Paragraph 3 "Elsewhere, a study aiming determine the burden and pattern of cancer in ……."
2. There is inconsistency in the format of data presentation. For example, age groups were written as 22 to 44 years, and in another 45-65 years and elsewhere 45-64 years.
3. Some sentences are too long. E.g. "Nearly 15% of the global cancer burden is attributable to infectious agents, with two-thirds of infection attributable cancers occurring in the less developed countries and in which infections accounts for nearly one in four cancers."

Summary

In this paper, the authors described the prevalence of cancers and the proportion of cancers that were associated with infections in two referral hospitals in Kenya.
The data was over-interpreted. The methods are inadequately described and the results are poorly presented. The conclusions drawn from the data are not justifiable.

Is the work clearly and accurately presented and does it cite the current literature?
No

Is the study design appropriate and is the work technically sound?
No

Are sufficient details of methods and analysis provided to allow replication by others?
No

If applicable, is the statistical analysis and its interpretation appropriate?
No

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Cancer epidemiology

We confirm that we have read this submission and believe that we have an appropriate level of expertise to state that we do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 28 Oct 2018

Lucy Wanjiku Macharia, University of Nairobi, Nairobi, Kenya

Authors' Response to Reviewers' Comments:

Journal: AAS Open Research
Manuscript #: 12910

Manuscript Title: Burden of cancer in Kenya: types, infection-attributable and trends. A national referral hospital retrospective survey

We thank the reviewers for having reviewed our manuscript, for the suggestions and for the positive criticism. We have performed a major revision in the manuscript in agreement with the suggestions. We have applied the PAF/AF standard formula in estimating the number of cancer cases attributable to infections. Secondly, we have conformed to the third edition of the International Classification of Diseases for Oncology (ICD-O) way of listing the cancers. For cancers like leukemias and lymphomas where the ICD-0-3 topographical code was not clear, we converted the IDC-0-3 morphological code to ICD10 topographical code for clarity.
We hope that this new version of the manuscript can be reconsidered to be of an acceptable scientific standard.
Kindly find below a point-by-point response to the comments and queries.

Sincerely,
L.W.M

Overall Comment:
"This paper provided information on the prevalence of cancers and the proportion of cancers that were attributable to infections in Kenyatta National Hospital (KNH) and Moi Teaching and Referral Hospital (MTRH), Kenya over a 5 year period from 2008 to 2012. These 2 centers are the oldest and largest national referral hospitals in Kenya. The authors randomly selected some cancer cases reported at these centers within the study period and reviewed the charts for cancer types and numbers of cancers attributable to infections. The study reported that 40.0% to 53.2% of cancers that were seen in these referral hospitals were attributable to infections."

Response:
We are thankful to the reviewers for revising our manuscript, for the positive criticism and for the suggestions. We have revised the manuscript text, tables, figures and reference list as described below and in the revised manuscript. We have also updated our results and discussions section together with additional analysis as suggested.

Main comments
1: Title
"In this paper, the numbers and proportions of cancers associated with infections were presented and not the proportion of cancers attributable to infections. In order to obtain the proportion of cancers that were attributable to infections, the author would need to apply the formula for the Population Attributable Fraction (PAF) which was not done in this paper."

Response:
We agree with the reviewers' concern. We have applied the PAF/AF standard formula in estimating the number of cancer cases attributable to infections. We are thankful for the suggestion.

2: Introduction
"The introduction is repetitive and should be edited. Separate paragraphs should be devoted to those infections that are established as causes of cancer while those will weaker associations should be in a separate paragraph.

The role of HIV infection in cancer etiology, prevalence and incidence in Kenya deserves more discussion than just mentioning HIV in a list of pathogens.

Previous studies of infection-associated cancers in Kenya, East Africa or Sub-Saharan Africa should be discussed briefly to provide more introductions to the field and establish why this study was necessary and what it hopes to accomplish."

Response:
We are grateful for the insightful suggestions. We have edited the introduction as recommended in the new manuscript.

We have included a new paragraph on HIV prevalence and incidence in Kenya together with its associated cancers.

Previous studies on infection-associated cancers in sub-Saharan Africa have been expounded in details in the data analysis section.

Comment 3: Methods

3.1 "It would be more informative to describe the demographic characteristics of the catchment areas, the pattern of healthcare services including other hospitals, and referral systems for the two hospitals included in this study. While these are the biggest hospitals, they may not see the most cancer patients if specialized cancer hospitals are located in the same region or city."

3.2 "The rationale, application, and choice of sample size calculation are unclear. This is a retrospective case series. The researchers should analyze all the case records that they can find irrespective of sample size calculation. Further, it is highly unlikely that the prevalence of cancer in adult Kenyan population is 50%."

3.3 "The justification for randomly selecting case notes to review is not clear. Rather than statistically estimate the number of files, why not obtain the number of patients seen and the number who have a cancer diagnosis directly from the institutions? If this is not possible, the authors should say so."

3.4 "The randomization procedure is not clear. The authors state that "the first record was selected randomly every year". What is the "first record"? Record of the first patient seen in the institution within a specific year? The first record selected from among all the patients' records for a specific year? If the latter how was the selection done? Were all the records assigned numbers and random number generators used to select records?"

3.5 "Do the authors have an objective assessment of the completeness of the records of these two institutions?"

3.6 "What is the relationship of the two institutions that refused to grant permission for this study and their locations with the institutions that gave permission for the study? A map showing this information would be informative for international readers"

3.7 "The analytic methods used in the paper and described in the data analysis section should be described better. ."

Response:

3.1. We are thankful to the reviewers for the suggestions. We have included the demographic characteristics of the catchment areas together with the pattern of healthcare services in the study site section of the new manuscript. To clarify the reviewers' concern, there are 12 health facilities in Kenya that offer cancer services; seven private hospitals, two mission hospitals and three public facilities (KNH, MTRH and Coast Province General Hospital). Because of the affordability of the cancer services, most patients opt for the public facilities resulting in congestion of the facilities and long waiting times of the patients. Patients with private insurance and the government-sponsored

scheme, National Health Insurance Fund, are more likely to undergo treatment than those without insurance.

3.2. We have edited the sample size calculation section and included detailed explanations with references in the new manuscript. We do agree with the reviewers' concern of analyzing all the chart reviews. However, due to cost and time constraints, we were only able to analyze 500 files in each hospital, as this was the minimum necessary sample size needed to achieve the required power of the study. We have also come across an article showing that sampling is also applicable in retrospective chart review studies [1] but to avoid any misunderstanding, we have included this as one of the limitations of the study. Regarding the prevalence, we stated in the sample size calculation section that the prevalence of cancer in Kenya was unknown at the time of conducting the study. For a better clarification, a 50% prevalence (p) was used in calculating the sample size since it is known that when d =0.05 and a z=1.96, using a p of 0.5 (50%) yields the highest sample size required for cross-sectional studies [2]. A detailed explanation has been included in the new manuscript.

3.3. We thank the reviewers for the positive criticism. For clarification, knowing the number of patients with cancer was not our only objective. We also aimed at knowing the of cancer types, the sex and age of the patients under study, their origin by birth, the method of cancer diagnosis used, year of diagnosis and whether the patient was referred from another health facility. At the time of the study, the hospital databases could not provide all the information needed and therefore we had to use the files. Randomization was done because of the cost and time constraints. In a facility like KNH that had 17,584 cancer files, abstracting all the information needed from the patient files to the data collection form would cost more and need longer time that the time allocated to us by the ethics committee.

3.4. We agree with the reviewer's concern. We have modified the randomization section. For clarification, the files in KNH were captured in a database and randomization was purely an automated process since all files have a hospital number (that cannot be disclosed to protect the privacy of the patients). However, in MTRH, the records department was in the middle of updating their databases when this study was being done and because of this we were only given an estimate of the number of files and to achieve the required sample size we used convenient sampling method.

3.5. Yes, the authors had an objective assessment of the completeness of the records at the two institutions. First, randomization depended on the total number given to us of the number of cancer files available for the five year period. Secondly, to obtain all the data for the questions we had, we needed to use a patient file with complete information.

3.6. We thank the reviewers for the question. The four hospitals initially selected for the study included; Kenyatta National Hospital (KNH), Moi Teaching and Referral Hospital (MTRH), Jaramogi Oginga Odinga Teaching and Referral Hospital (JOOTRH) and Coast General Hospital previously known as Coast Province General Hospital (CPGH). At the time, they were all teaching and referral hospitals. KNH and MTRH are national hospitals (level 6) while JOOTRH and CPGH are level 5 hospitals. KNH is located in the former Nairobi province, MTRH is located in the former Rift Valley province, JOOTRH is located in the former Nyanza province and CPGH is located in the former Coast Province. Although we had ethical approval together with a letter of authority from the Ministry of Health giving us the authorization to conduct research in the four facilities, permission to

access the files was only granted by the National hospitals. We have included a map in the new manuscript to guide international readers. We are very grateful for the suggestion.

3.7. We are thankful for the suggestion. We have edited the "data analysis" section in the new manuscript.

Comment 4: Results
"The authors should be precise in reporting results. "Around 4304 inpatient cancer cases were available in MTRH" is imprecise and unacceptable."

"There is a clear difference in the cases of cancer at the KNH compared to MTRH, most probably because one site hosts AMPATH HIV treatment and prevention programs. It is therefore not justifiable to make too much of the differences in the proportion of cancers seen in these two institutions and attempt to extrapolate that to the general population."

Response:
We appreciate the suggestion. We have improved the precision of reporting the results.

We thank the reviewers for the suggestion. We have edited the results section and minimized extrapolating the differences from the two facilities to the general population.

Comment 5: Discussion
5.1. "The interpretation of the result is not justifiable based on the data, methods, and analyses.

5.2. Proportions of cancers presenting in specific institutions may not reflect population level epidemiology of the cancers. See Chapter 1 of the Cancer in Africa (Ed) Max Parkin et al.

5.3. The authors suggest that their data shows the incidence of the cancers that they described but this is not correct.

5.4. The authors can describe the pattern of cancers presenting to each of these hospitals over a period of time, but this is not "incidence" of those cancers in the population. The pattern of presentation of cancers to specific hospitals is influenced by many factors. An example is illustrated in their paper where the pattern of cancers presenting at the AMPATH related hospital reflects the focus on HIV treatment and prevention at the institution."

Response:
5.1. We thank the reviewers for the positive criticism. We have modified the analysis and improved on the interpretation of the data. We have emphasized where necessary that we were comparing our hospital-based results to other results obtained from population-based studies.

5.2. We agree with the reviewer's concern that the proportions of cancers presenting in specific institutions may not be reflective at a population level. We have highlighted this as a limitation of this study.

5.3. The use of the term "Incidence" was an unintended mistake. We have edited the discussion section and thank the reviewers for the observation.

5.4. We agree with the reviewer's observation. We have edited our new manuscript accordingly and avoided the use of the term incidence.

Comment 6: General comments
1. *There are several grammatical and typographical errors in the paper. For example in the Discussion section, Paragraph 3 "Elsewhere, a study aiming determine the burden and pattern of cancer in ...."*
2. *There is inconsistency in the format of data presentation. For example, age groups were written as 22 to 44 years and in another 45-65 years and elsewhere 45-64 years.*
3. *Some sentences are too long. E.g. "Nearly 15% of the global cancer burden is attributable to infectious agents, with two-thirds of infection attributable cancers occurring in the less developed countries and in which infections accounts for nearly one in four cancers."*

Response:
6.1. We thank the reviewer for the observation. We have improved on the wording of the manuscript and corrected the typographical errors.
6.2. We thank the reviewer for the comment. We have ensured consistency in the data presented in the new manuscript.
6.3. We thank the reviewer for the suggestion. We have minimized the use of long sentences in the new manuscript.

Comment 7: Summary
"In this paper, the authors described the prevalence of cancers and the proportion of cancers that were associated with infections in two referral hospitals in Kenya.
The data was over-interpreted. The methods are inadequately described and the results are poorly presented. The conclusions drawn from the data are not justifiable."

Response:
We thank the reviewer for the positive criticism. We have performed major revision in the new manuscript taking all suggestions into account. We have applied the standard formula of calculating the cancers attributable to infections. We have updated the methodology section and we have ensured that the results or tables are listed according to the third edition of the International Classification of Diseases for Oncology (ICD-O). We have only reported the results as generated from the data from the two hospitals and highlighted where necessary that we were comparing our hospital-based data with other population-based studies. We have inserted a new paragraph on the limitations of this study and modified the conclusion. We are thankful for the suggestions.

References
1.   Gearing RE, Mian IA, Barber J, Ickowicz A. A methodology for conducting retrospective chart review research in child and adolescent psychiatry. J Can Acad Child Adolesc Psychiatry. 2006;15: 126–134. doi:10.1186/1478-4505-4-26
2.   Arya R, Antonisamy B, Kumar S. Sample size estimation in prevalence studies. Indian J

Pediatr. 2012;79: 1482–1488. doi:10.1007/s12098-012-0763-3

Competing Interests: We have no conflict of interest to declare