# ADDIS ABABA UNIVERSITY
# COLLEGE OF HEALTH SCIENCE
# SCHOOL OF PUBLIC HEALTH

# ASSESSMENT OF POPULATION BASED CANCER REGISTRY DATA QUALITY IN ADDIS ABABA, ETHIOPIA, 2019/20

BY

SHIBABAW YIRSAW(BSc)

Advisors

Girma Taye (Ph.D., Associate Professor)
Wondimu Ayele (MPH, Ph.D. Fellow, Assistant Professor)
Mathewos Asefa (MD, Oncologist)
Eva J Kantelhardt

A Thesis Proposal to be submitted for the Public Health Department, College Of Health Science, Addis Ababa University in Partial Fulfillment of the Requirement for the Degree of Masters of Public Health

November, 2020

Addis Ababa, Ethiopia

# Addis Ababa University
# College Of Health Sciences
# School Of Public Health

This is to certify that the thesis prepared by Shibabaw Yirsaw Akalu, entitled: *Assessment of Data Quality in Addis Ababa Population-Based Cancer Registry* and submitted in fulfillment of the requirements for the Degree of Masters of Public Health in General Public Health complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

---------------------------------------- ----------------- -------------------

Chairman, Examining Committee sign Date

----------------------------------------- --------------- --------------------

External Examiner sign Date

----------------------------------------- --------------- -------------------

Internal Examiner sign Date

----------------------------------------- --------------- --------------------

Advisor sign Date

----------------------------------------- ---------------- ---------------------

## Declaration

I the undersigned, declare that this thesis is my original work and it has not been presented in other universities, colleges, or institutions for a similar degree or another purpose. All sources of the materials used in this thesis have been duly acknowledged.

Name of the student: **Shibabaw Yirsaw Akalu**

Signature.______________________ Date__________________

**Approval of the primary Advisor**

Name of the primary advisor: **Dr Girma Taye ( PhD, Associate Professor)**

Signature.______________________ Date __________________

ADDIS ABABA UNIVERSITY,

COLLEGE OF HEALTH SCIENCES,

SCHOOL OF PUBLIC HEALTH

# MASTERS OF PUBLIC HEALTH

# RESEARCH THESIS SUBMISSION FORM

| | |
|---|---|
| Name of Investigator | Shibabaw Yirsaw |
| Name of Advisor/s | Girma Taye (Ph.D., Associate Professor)<br>Wondimu Ayele (MPH, Ph.D. Fellow)<br>Mathewos Asefa (MD, Oncologist)<br>Eva J Kantelhardt |
| The full title of the research | "Assessment of Data Quality in Addis Ababa Population-Based Cancer Registry, Addis Ababa, Ethiopia, 2019/2020." |
| Duration of the project | May 2020 to October 2020 |
| Study area | Addis Ababa, Ethiopia |
| The total cost of the project | 40,120.00 Birr |
| Source of budget | Addis Ababa University, DDCF project |
| Contact Address of Principal Investigator | Mobile Phone -0925334474<br>Email. shibabaw2063@gmail.com |

# ACKNOWLEDGMENT

Before everything, thanks to the Almighty God for the countless gifts you offered me!

***ለቅኖች ምስጋና ይገባል! መዝ 32፡1***

***For praise is comely for the upright! Psa 33:1***

I would like to express the deepest appreciation to my advisors Dr. Girma Taye and Wondimu Ayele for their unreserved constructive comments and generous suggestions throughout the development of this research work. Without your guidance and persistent support, the work of this research project would not have been possible.

It is a great pleasure to acknowledge Dr. Mathewos Asefa for his help to do my research project on this area, Eva J Kantelhardt for her constructive comments and suggestions, and Solomon Asmare, he helped me a lot throughout my project work.

I am grateful to thank all staff members in the Addis Ababa cancer registry, Key Informants, and medical registration unit workers of the Oncology department, Tikur Anbesa Specialized Hospital.

I would like to express my extreme sincere gratitude for my families: my wife Tirusew Malede (MA), without you nothing is possible, and my little one Yohana Shibabaw, your unlimited love makes me strong! Long live for both of you! And thank you all my friends.

My heartfelt thank and appreciation for Ministry of Agriculture; for its sponsorship of the master's degree program, Doris Duke Charitable Foundation (DDCF)" and "JSI Research & Training Institute, Inc." for their financial support.

Finally, I extend my appreciation to Addis Ababa University, College of health science, department of public health for creating an opportunity to conduct this research as part of the requirements of a Master's degree In Public Health.

# Contents

## Lists of Tables

## Lists of Figures

# ACRONYMS

| | |
|---|---|
| AAPBCR | Addis Ababa Population-Based Cancer Registry |
| ACRN | African Cancer Registry Network |
| ANC | Antenatal Care |
| ASR | Age Standardized Rate |
| CI5 | Cancer Incidence in five continents |
| Canreg5 | Cancer registration 5 |
| DCO | Death Certificate Only |
| DDCF | Doris Duke Charitable Foundation |
| HMIS | Health Information System |
| IACR | International Association of Cancer Registries |
| IARC | International Agency for Research on Cancer |
| ICD | International Coding of Disease |
| ICD-O | International Coding of Disease for Oncology |
| KII | Key Informant Interview |
| MV | Morphological Verification |
| NCD | Non Communicable disease |
| SSA | Sub Saharan Africa |
| TASH | Tikur Anbesa Specialized Hospital |
| WHO | World Health Organization |

# Abstract

**Background (problem statement):** Population-based cancer registries systematically collect information on all new cancer cases in a defined geographic area. The value of cancer registries is highly dependant on the quality of the data that it collects, but the level of data quality for the population-based cancer registry in Addis Ababa was not evaluated and well understood.

**Objectives:** the general objective was to assess data quality status and gaps in Addis Ababa population-based cancer registry, Addis Ababa, 2019/2020

**Methods:** Convergent parallel mixed study design with the methods proposed by Bray and Parkin, and Parkin and Bray (Eur J Cancer 2009;45:747–64) was applied to the Addis Ababa population-based cancer registry data from 2012 to 2016. Cancer cases registered in the database from 2012 to 2016 were included in the study, and data were sampled for re-abstracting audit and timeliness and the sample size was 408. Polynomial and linear regression, one-way ANOVA, Non-parametric test, kappa coefficient, and Intraclass correlation were performed using SPSS.

**Results:** The coding and classification in the registry agreed with international agreements. Annual trends in ASR for both sexes was not linear and exhibited fluctuations, age-specific incidence curves in both sexes showed a rapid rise above middle age groups and sharp decline at old age, ASR for three most common neoplasms in both sexes was lower than other SSA Countries and childhood incidence rate per 100,000 showed below world reference for males 0-4 and females age group 5-9 and 10-14. MV% was 87.2% in males and 90.2% in females, the agreement rate in the re-abstracting audit was reasonably high, DCO% was 0.11% in males, and 0.14% in females, the overall proportion of other and unspecified cases was 2.6%. The meantime between identification ad registration of cases was 92 days.

**CONCLUSION:** the AAPBCR was comparable in using international agreements, there was a reasonably high level of validity in the registry, with the variability of completeness and delay registration. Budget and ownership were major gaps identified. It was recommended to improve the data collection system to active, covering all data sources, and the FMoH to handover the registry from owners.

**Key Words**: Cancer Registry, Data Quality, Comparability, Completeness, Validity, Timeliness

# 1. INTRODUCTION

## 1.1. Background

The process of collecting and registering cancer cases using the cancer registry was first started in Humberg in 1926 followed by the Danish cancer registry and currently, there are more than 700 cancer registries worldwide(1–3). The first cancer registry in Africa, the Kampala cancer registry was established in 1954(4).

Population-based cancer registries systematically collect information on all new cancer cases in a defined geographic area(5), which are crucial to cancer data collection (incidence, prevalence, mortality, and survival), data analysis, understanding epidemiologic trends, development of a national cancer control program, research, and resource allocations(6). The most basic function of a cancer registry to cancer control is to assess the current magnitude of the cancer burden and its likely future evolution(7).

Cancer registries are basic institutions for controlling cancer disease and its principal purpose is systematic and continuous registration, compilation, analysis, and reporting of cancer data and as well as classification of cancer cases(8). Population-based cancer registries are foundational parts of the national cancer control plan, and essential for attaining results, and not only measuring rates, trends, and sex but also to evaluate the quality of care for cancer cases(5).

The new cancer data from population-based cancer registries were used for estimating global and national cancer burden, but most of the time the population-based cancer registries covered major cities and not accessible for the rural environments(9).

Population-based Cancer Registry in Ethiopia named “the Addis Ababa City Cancer Registry” was first established in Addis Ababa City Administration in Tikur Anbesa specialized university hospital within the radiotherapy center since September 2011 and collects data from the residents of the city Administration(10), and it is one of the 30 population-based cancer registries in Africa. The main source of information for the registries are public and private hospitals diagnosing and treating cancer cases, diagnostic/pathology laboratories, and death certificates in the registry area(1).

The registry has been used for cancer incidence and prevalence estimation, national cancer control plan, and considerably used for research purposes(11,12).

Health data are basic sources of information to assess national programs progress and performance and on an annual basis nationally or subnationally(13), and poor data quality can affect the level of data utilization efficiency by health facilities policymakers and even and may result in poor and wrong decision-making (14). Similarly, quality data in population-based cancer registries have a significant role in planning and evaluating cancer control programs(5).

## 1.2. Statement of the problem

Assuring the issue of data quality is becoming difficult due to the diversity of data sources, complex data structure, the difficulty of data integration, rapid change in big data, and the absence of data quality standards(14). In a study conducted in Eastern Ethiopia, only 75.3% of health facilities assured data quality(15), and a study conducted in Jimma also showed a low level of completeness and timeliness(16).

Quality data for the health information system in Ethiopia is below the expected standard and is the salient obstacle to improve health service delivery and data with poor quality is being used for decision(17). Data verification report generated by Ethiopian Public Health Institute showed that only 72% of health facilities had source document for their ANC report, 87% of facilities that used the HMIS report system had complete ANC data and 30% of health facilities had ANC report consistent with the source document, which showed the status of data quality in health facilities is below the standard(18).

The emphasis of a cancer registry should be on the quality of the data collected, rather than on the quantity(6), and the value of cancer registries is highly dependent on the quality of data that it collects(19).

Even though cancer registries are important for cancer data collection, epidemiological surveillance, incident determination, and national cancer control programs (8), the lack of high-quality data on cancer incidence makes it difficult to evaluate the exact burden of cancer(20). In Addis Ababa's population-based cancer registry only morphological verification was reviewed by Memirie et al., 2018 which showed an overall 89% (12), and by Parkin et al., 2018 which showed 84.3% in males and 86.8% in females(10).

The Addis Ababa cancer registry data is used to estimate incidence rates and burden of cancer in the country and there are studies conducted on the population-based cancer registry of Addis Ababa for this purpose which most are focused on the incidence of cancer and they didn’t evaluate the quality of data in the registry(11,12). As a result, the level of data quality for the population-based cancer registry in Addis Ababa is not currently evaluated and well understood comprehensively.

## 1.3. Significance of the study

This study aimed to assess data quality at a population-based cancer registry in Addis Ababa City using indicators of data quality (comparability, completeness, validity, and timeliness) proposed by Bray and Parkin, and Parkin and Bray (Eur J Cancer 2009;45:747–64)(21,22).

This study might contribute to the information revolution of the Health Sector Transformational Plan (HSTP)I by estimating the status of data quality in the cancer registry.

Similarly, as the purpose of the study was determining the level of data quality in the registry, it might lead researchers to study factors associated with data quality.

Even though the study was concerned about data quality, it will generate information about incidence rate, age-standardized incidence rates, and age-specific incidence rates which may indirectly showed the burden of cancer in the country as other studies did (11,12).

This study was intended to show the level of data quality in Addis Ababa population-based cancer registry which is important for policymaking, planning, and monitoring, and evaluation of cancer programs.

Finally, this study could be an input for new cancer registries that will be emerged in different regions and upgrading to a national population-based cancer registry.

# 2. LITERATURE REVIEW

## 2.1. Cancer and cancer Registries

### 2.1.1. Cancer and its types

Cancer is a disease caused by uncontrolled growth and division of cells due to cellular changes. Some cancer types cause rapid cell growth and others cause cells to grow and divide at a slower rate. Certain forms of cancer result in visible growths called tumors, while others, such as leukemia, do not (23). Cancer is the generic term for a group of diseases that can affect any part of the body. Malignant tumors of the brain, lung, breast, prostate, skin, and colon are among the diseases known as cancer(6).

A lifestyle like living and working conditions, smoking, unhealthy diet, and physical inactivity are the major factors for the cause of cancer, even if it is biologically caused by mutations in the genes responsible for cell multiplication and repair(6,24).

Cancer cells primarily started in a specific body part, and then spread through the lymph nodes to the remaining body cells (23). Lung and stomach cancers, and breast and cervical cancers are the most prevalent cancer types worldwide in males and females respectively(6).

### 2.1.2. Burden of Cancer

Global cancer statistics report an estimated 18.1 million new cases and 9.6 million deaths worldwide due to cancer. Among this burden Asia, Americas, Europe, Africa and Oceania shared 48.4%, 21.0%, 23.4%, 5.8% and 1.4% of new cases and 57.3%, 14.4%, 20.3%, 7.8% and 0.7% of mortality respectively(20).

Globally; Lung cancer, breast cancer, colon cancer, and non-melanoma of skin constitutes the top five causes of new cancer cases, and lung, stomach, liver, colon, and esophageal cancers were the five leading causes of new cancer deaths respectively in 2018(20). Prostate and liver cancers were the leading causes of male incidence and mortality in Africa, while breast and cervical cancers were the leading causes of incidence and mortality in females respectively. Similarly, the leading causes of male incidence in America were prostate cancer but the leading cause of mortality was lung cancer, and in females, it was breast and lung cancers.

Lung cancer was the leading cause of male incidence in Asia and Europe, differently in Oceania non-melanoma of skin cancer was the leading cause of male incidence and lung cancer was the leading cause of mortality, and breast and lung cancer were leading cause for females(25). Commonly seen cancer cases in Ethiopia hospitals were leukemia, lymphoma, retinoblastoma, Wilms' tumor (nephroblastoma), and bone and soft tissue sarcomas in children and breast cancer, cervical cancer, head and neck cancer, oesophageal cancer, sarcoma, colorectal cancer, liver cancer, Non-Hodgkin lymphoma and Skin cancer in adults. (26)

Non-communicable diseases in Ethiopia were responsible for the cause of 711 deaths per 100,000 people in 2015, becoming the leading cause of age-standardized death rates and Cancer also caused 123 deaths per 100,000 people(27). The 2018 global cancer estimate also showed that the burden of cancer in Ethiopia is increasing and figured as 67, 573 new cancer cases, and 47,974 deaths(20). In a study conducted in Addis Ababa population-based cancer registry, cancer of the breast(31.5%), cervix uteri(14.1%), and ovary(6.3%) were the most common cancers in the women, and colorectal(10.6%), non-Hodgkins lymphoma(10.2%) and prostate(6.9%) cancers were the commonest among males(11). Similarly, Memirie et al estimated 21,563 male and 42,722 female new cancer cases for 2015, and colorectal, non-Hodgkins lymphoma and prostate cancers were commonest in males and breast, cervical, and ovarian cancers were commonest in females respectively(12).

### 2.1.3. Cancer registries

#### 2.1.3.1. History of cancer registries

The first population-based cancer registry was established in Humberg in 1926 and followed by the Danish population-based cancer registry and currently, there are around 700 population-based cancer registries globally(1–3).

In Africa the first population-based cancer registry was the Kampala cancer registry which was established in 1954 followed by the Ibadan cancer registry in Nigeria established in 1960 and Addis Ababa and Abuja cancer registry were established lately(4). The Africa Cancer Registry Network(AFCRN) which was first founded in 2012 has currently members of around 30 population-based cancer registries. (10,27)

#### 2.1.3.2. Types of cancer registries

Leal et al, 2016 classified cancer registries based on their source of information and their purpose into three groups as population, histopathological, and hospital registries(5).

**Population registry**:- This type of registry is the most commonly used registry and systematically collects information on all new cancer cases in a particular geographic area and is determined by multiple sources like public and private hospitals and medical centers, pathology laboratories, and death certificates(1,5).

**Histopathological registry**:- “This type of registry collects information from one or more pathology laboratories and is useful for laboratory needs. It provides an incomplete and skewed cancer profile, essentially determined by the types of tissues that the laboratory can process” (1,5).

**Hospital registry**:- This type of registry collects information from all cancer patients treated at one or more hospitals. It is useful for administrative purposes because it aids in prioritizing hospital resources. Also, it facilitates monitoring of health programs and allows the detection of patterns or frequencies of different types of cancer treated in the hospital as well as monitoring of the outcomes of treatment, survival rates, quality of life, and adverse effects of treatment (1,5). The main difference of the three cancer registries is both the hospital-based cancer registry and histopathological or pathological cancer registries provide biased population sample and the two are less useful for cancer control planning, but population-based cancer registry can ascertain cancer incidence from multiple sources and important for cancer control planning(1).

#### 2.1.3.3. Source of Information for cancer registries

Most population-based cancer registries used similar source of information which include: public and private hospitals and medical centers, public and private outpatient surgery centers, public and private anatomical pathology laboratories, civil registry offices that issue death certificates, public and private specialty cancer diagnostic centers, public and private hospice centers, and public and private nursing homes(5), and the sources of information can be generally grouped into three main categories as hospitals, laboratories and death certificates(1).

#### 2.1.3.4. Importance of cancer registries

Population-based cancer registries are the most crucial information sources for cancer incidence in a geographically defined population and are helpful for studies concerned on possible causes of cancer in the community and to evaluate cancer control programs(1). Parkin, 2008 illustrated the most important purposes of cancer registries, which includes; assessment of the current magnitude of cancer burden through incidence measurement, estimation of mortality rates and prevalence rates of cancer cases, measurement of survival rate from cancer, an important source of information for any risk factors, source of cases for case-control and cohort studies, screening, and early detection and for evaluation of cancer control programs(7). Population-based cancer registries are highly desirable for the development of national cancer control programs and also are useful to follow the pattern of cancer in a given geographical area by evaluating trends of incidence for cancer cases over time(6).

## 2.2. Data quality in Cancer Registries

For the evaluation of data quality in population-based cancer registries, four data quality dimensions were proposed and described by IARC and Parkin and Bray, 2009; Bray and Parkin, 2009, which are Comparability, Completeness, Validity, and Timeliness (1,21,22), and most cancer registries were evaluated by using those methods and principles(10).

### Comparability

As described by Parkin and Bray, 2009; Bray and Parkin, 2009, comparability is “the extent to which coding and classification procedures at a registry, together with the definitions of recording and reporting specific data items, adhere to agreed international guidelines”(21,22) and is used in different cancer registries to evaluate the data quality. In Singapore, Zurich, and Zug population-based cancer registries coding and recording of all cancer cases were performed following international standards using ICD-9 for the period 1980-2002 and ICD- 10 since 2003(29,30). Similarly, the national cancer registry of Ukrain used ICD 10 and ICD-O-3 to record and code cancer cases and to collect information on all malignant and in situ neoplasms(31). The classification and coding of neoplasms in the Gambia cancer registry also followed ICD-O-3, rule of incident date registration, and multiple primary tumors recording is performed according to the recommendation proposed by the International Association of Cancer Registries(32).

The studies conducted on Addis Ababa population-based cancer registry which concerned cancer incidence and cancer data report didn't include any report on comparability(11,12).

## Validity

Validity is described by Parkin and Bray, 2009; Bray and Parkin, 2009, as "the proportion of cases in a dataset with a given characteristic (e.g. site and age) which truly have the attribute"(21). The main methods used to evaluate the validity of cancer registry in Gambia cancer registry were re-abstracting and recoding audit which had 94% agreement with original, morphologically verified cancer cases varied little with time, and the proportion of death certificate only cases 6.6% for men and 3.6% for women(32). In the Finland cancer registry 93% of all registered cancer cases were morphologically verified and the proportion of death certificate only cases were 2.6% of all tumors(33). For Singapore cancer registry the proportion of morphologically verified cancer cases were 88.6% for males and 93.0% for female, and 1.5% of cases were death certificate only(30). The proportion of morphologically verified cases increased from 73.6% in 2002 to 82.3% in 2012 for the cancer registry of Ukraine, but DCO was very low accounting for only 0.1% for both sexes(31). The proportion of morphologically verified cases of Addis Ababa cancer registry were relatively low which accounts for 84.3% for males and 86.8% for females, and the registry doesn't use death certificate only verification method(10).

## Completeness

Completeness is the degree to which all diagnosed neoplasms within a registry's catchment population are included in the registry database. Several methods can be used to evaluate the level of completeness of the enumeration of cases within a catchment population(22,34).

The semi-quantitative methods used in the evaluation of Iceland cancer registry showed a relatively high degree of completeness; the trend of incidence rate was stable over time and comparable with other Nordic countries, childhood cancer incidence rate was above the given reference range(35). Data quality evaluation in Singapore cancer registry showed a stable annual incidence rate trend for the period 1968-2003, the age-specific incidence curves for males and females in the registry indicated no variability for completeness, and age-specific incidence rates for childhood cancer in Singapore, 2008–2012, was compared with the upper and lower deciles for childhood cancer published in CI5 Volume X, indicated the values were within range except for boys and girls aged 0–4 years which exceeded the upper limit of the reference interval(30).

## Timeliness

Rapid reporting of information on cancer cases is the fourth method of data quality evaluation for cancer registries(21). The median time from diagnosis to availability for research is eight months in the Iceland cancer registry and cancer data were reported within 15 months, which was in the range of recommended guidelines (35). But the Addis Ababa population-based cancer registry timeliness was not estimated and well understood.

## 2.3. Cancer registry data quality Conceptual Framework, Addis Ababa, 2012-2016

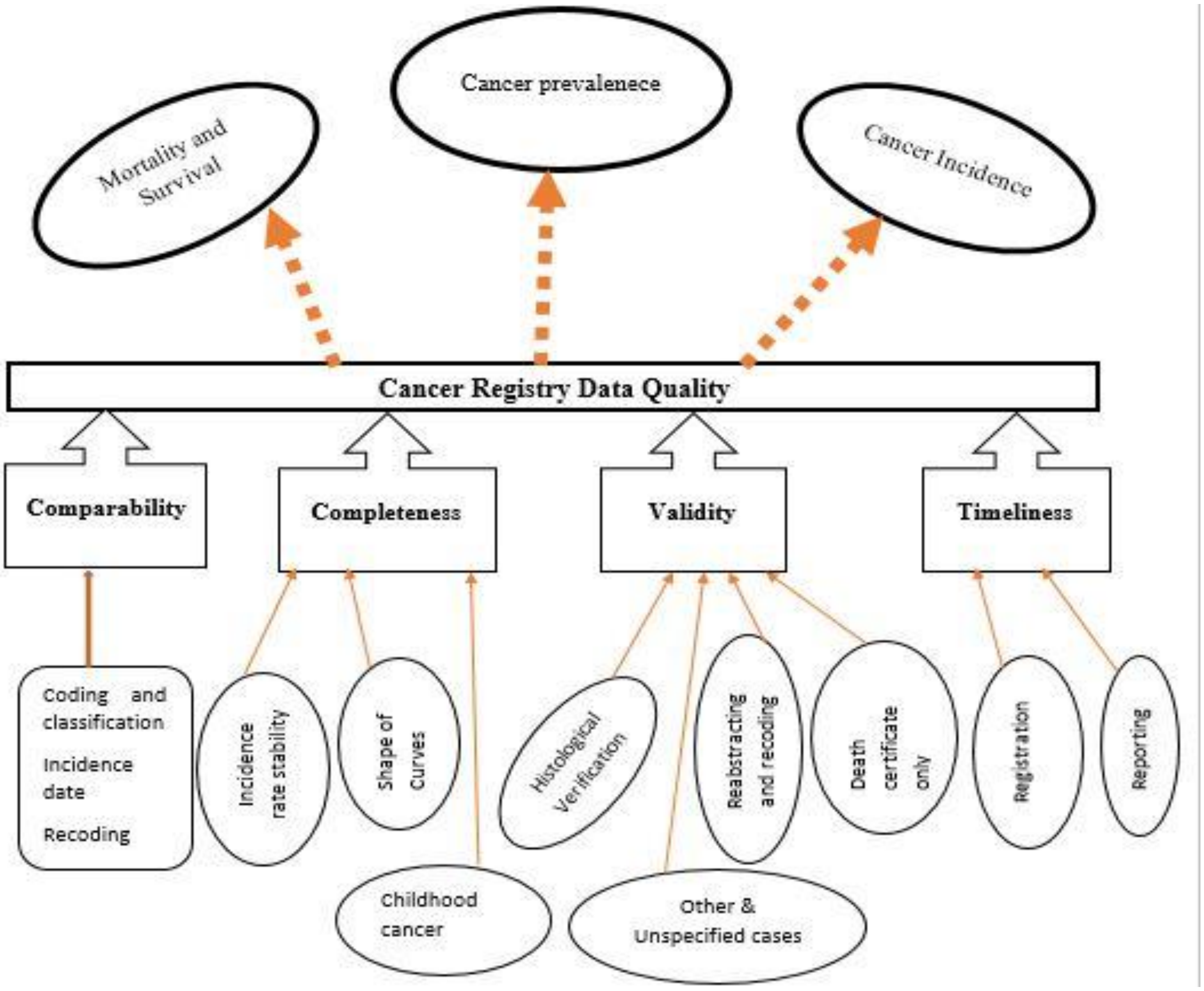

Figure 1: Conceptual framework of data quality assessment in the cancer registry

This population-based cancer registry data quality assessment conceptual framework is developed based on a literature review and dimensions of data quality assessment and methods proposed by those reviewed literature. The first stage before/below the dimensions are methods applied to measure the four dimensions, each indicated by arrow to the corresponding dimensions and next are four dimensions used for data quality evaluation in cancer registries. The expected outputs of quality cancer data are Incidence rate, prevalence rate and mortality. Quality data in population-based cancer registries are useful for the generation of information about the incidence, prevalence, and mortality, and survival of cancer. Outputs indicated by the broken arrow are not included in the study or the study objective was not to produce these rates.

# OBJECTIVES

## GENERAL OBJECTIVE

To assess data quality status and gaps in Addis Ababa population-based cancer registry, Addis Ababa, 2019/2020

## SPECIFIC OBJECTIVES

1. To estimate the status of data quality in Addis Ababa population-based cancer registry, Addis Ababa, Ethiopia, 2019/2020
2. To identify data quality gaps in Addis Ababa population-based cancer registry, Addis Ababa, Ethiopia, 2019/2020

# 3. METHODS and MATERIALS

## 3.1. Study area and study setting

Addis Ababa city administration is the largest city in the country and is the capital city of Ethiopia, administratively divided into 10 sub-cities: Addis Ketema, Akaky Kaliti, Arada, Bole, Gullele, Kirkos, Kolfe Keranio, Lideta, Nifas Silk-Lafto, and Yeka. and had a total population of 3,435,028 according to the central statistics agency projection for 2017(36).

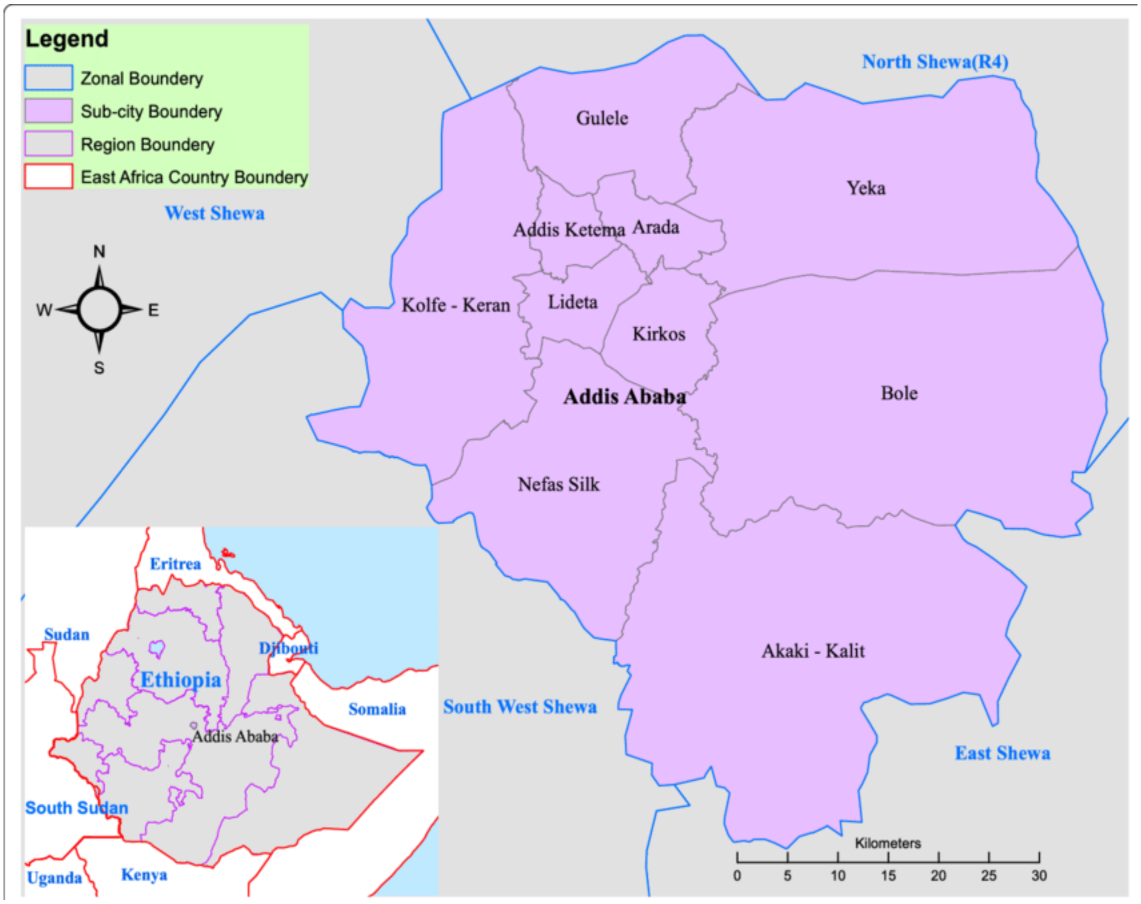

Figure 2: Administrative map of Adds Ababa

The Addis Ababa Population-Based Cancer Registry (AAPBCR) is one of the 30 registries in Africa(28), and the only population-based registry in Ethiopia, first established in 2011 under the radiotherapy unit of Tikur Anbesa Specialized Hospital (TASH) Addis Ababa University and serves the Addis Ababa population, the residents of Addis Ababa (as defined by ≥ 6 months of residence in the city).

The AAPBCR was run by four full-time employees, who received assistance with data collection from 22 staff members at selected hospitals. The main sources of information for the registry were government-owned hospitals, private hospitals, and diagnostic laboratories.

## 3.2. Study design

The mixed methodology is an emerging and widely used study method that systematically combined qualitative and quantitative data within the same examination of a research idea(37).

Convergent parallel research design which is the frequently used mixed research design and combines qualitative and quantitative approaches at a time (38) was used for a more complete understanding of data quality in the registry both for secondary data analysis and key informant interview to assess cancer registry data quality status and gaps in Addis Ababa population-based cancer registry.

Convergent parallel research design is the most familiar and advanced mixed research design in which qualitative and quantitative data can be collected and analyzed differently to provide different information and combined with the assumption of the same final result(38).

This research design was most appropriate for the study, to assess data quality status and identify data quality gaps in the registry. In this case, both quantitative data from the cancer registry and qualitative data obtained from key informants were analyzed independently or separately and combined to get more complete information.

Cancer data from 2012 to 2016 were obtained from the cancer registration database (canreg5) in the main office of Addis Ababa Cancer registry, and the other was obtained from independent re-abstaction of data from source documents specifically from patient medical registration/ cards.

Key informant interviews were conducted for cancer registrars, data collectors, and supervisors from the registry.

Secondary data analysis in the cancer registry from the canreg5 database was employed for the data from 2012 to 2016 and data quality of the registry were estimated using Bray and Parkin, and Parkin and Bray approach (Eur J Cancer 2009;45:747–64) to assess comparability, completeness, validity, and timeliness. (20,21)

### 3.2.1. Comparability:

Comparability is the degree to which coding of cases and data collection practices in the registry cohere with standards and international guidelines. The comparability of AAPBCR data was assessed by reviewing the data collection practices, standards and definitions used for coding and classification of neoplasms, guidelines used for coding of topography and morphology, coding of multiple primaries, incident date definitions, and staging in the registry, including a specification of the standards in coding, classifying and definitions that have been followed.

### 3.2.2. **Completeness**:

Completeness is the magnitude to which all new and diagnosed cancer cases occurring in Adis Ababa are included in the registry database. Semi-quantitative historic methods were used in this study to estimate the completeness of the registry. Before analysis data was converted from ICD-O to International statistical classification of disease and health-related conditions, 10th revision(ICD-10)(39).

**Semi-quantitative methods**

The semi-quantitative methods are methods that are mostly historical and database based calculations indicative for cancer registry completeness. Semi-quantitative methods showed some evidence about the registry completeness, don’t quantify the percentage of the completeness of the registry. These four semi-quantitative methods were estimated.

1. Trends of annual age-standardized (world population) incidence rates:- Stability of age-standardized incidence rates over five years from 2012 to 2016 or annual trends in the registry for most common cancers were examined.
2. Age-specific incidence curves:- age-specific incidence curves for the top three cancers in each sex with 5 years age difference groups were seen and observed for abnormal fluctuations of incidence curves.
3. Age-standardized incidence rates comparison:- Comparison of ASR from 2012 to 2016 for common cancers with five Sub-Saharan African countries(40) were made.
4. Childhood age-specific incidence rate:- age-specific incidence rates of childhood age groups from 0-4, 5-9, and 10-14 were compared with reference ranges published in cancer in five continents volume X(19).

### 3.2.3. **Validity**:

Validity is the accuracy of the work of the registry and is defined by Bray & Parkin as the proportion of cases in a dataset with a given characteristic (e.g. site and age) which truly have the attribute. In AAPBCR, to verify codes for sex and topography IARC CHEK has been used. The four methods below were used to estimate the validity of the registry.

1. Re-abstracting audit: which is very useful to check agreements with source medical record was performed. A total of 408 samples of patient cards were withdrawn from the Oncology Medical Registration Unit, cancer cases abstracted from the records, and entered into an excel tool prepared for this purpose for agreement determination.

   ICD-O coding system(41) was used to code topography and morphology of cancer cases, multiple primaries were coded according to the International Association of Research on Cancer(IARC) coding guideline, and the definition of an incident date was according to the guideline.

   The agreement rate between the data abstracted from sample source medical records/patient records and extracted from the database was calculated and inter-rater reliability was estimated using, kappa coefficient (κ)(42) for each variable with a p-value less than 0.05. All statistical analyses were performed using SPSS version 23. Similarly, an intraclass correlation was used to estimate uniformity among registry and re-abstracted data.
2. Percentage of morphologically verified cases:- the proportion of cases diagnosed using histology and cytology or cases diagnosed microscopically were calculated for common cancers and overall morphologically verified cases (MV)% were compared with Six SSA countries registries.
3. The proportion of other and unspecified cases:- the proportion of neoplasms with no specific primary sites from C76 to C 80 in ICD-10(39) were calculated.
4. The percentage of death certificate only cases was assessed.

### 3.2.4. **Timeliness**:

Timeliness is the time with which a registry can collect, process, codify, and report data with a high level of completeness, validity, and accuracy following time frames previously established by the registry was estimated. The mean time from diagnosis to registration and from registration to reporting date was calculated. The standard registration time frame in the registry was one month and 5 years for publication.

A key informant interview is an in-depth interview with persons who are familiar with a specific issue or work and who have knowledge of the issue and is important to fetch information from different experts and professionals. (42)

Key informant interviews were conducted using interview guides developed for this purpose for purposively selected key informants based on information saturation from the cancer registry and from the source medical records to identify data quality gaps in the registry.

## 3.3. Study population

The study population was all cancer cases included in the Addis Ababa Population-Based Cancer Registry database from 2012 to 2016. The registry only included new cancer cases from residents of Addis Ababa, and excluded any case from outside of Addis Ababa, even if diagnosed in Addis.

## 3.4. Sample Size and Sampling Procedure

The total cases registered from 2012 to 2016 was used in the study for the methods which need whole data analysis. The registry was established first in September 2011, but organized registration was started in 2012, and the data after 2016 was not available for research.

For those methods which need a sample, the sample size was determined using a single population proportion formula considering the prevalence rate of 50% with a 5% margin of error and a 95% confidence level. The actual sample size was computed using Epi info sample size calculation and was resulted in 371 from total registered 11,260 cancer cases in the registry, adding 10 % to the sample size to compensate for any lost medical records, the total sample size became 408. Case number (medical registration number) list from 2012 to 2106 was obtained from the main registry office and sampled cases were selected randomly using excel random sample selection function. Selected case numbers were given for Oncology medical registration unit and patient files were withdrawn.

The sampling was used only for re-abstracting audit (validity) and timeliness. Furthermore, whole data from 2012 to 2016 in the database was used for comparability, Validity (the proportion of morphological verification (MV%), death certificate only percentage (DCO%), the proportion of other and unspecified cases (O&U), and completeness (rates, trends, and comparisons).

For a qualitative exploration of data quality gaps in the registry, participants were selected purposively using the assumption of a heterogeneous sample, from the registry and source medical records, and 6 were interviewed according to information saturation assumption.

## 3.5. Data collection tool and procedures

### 3.5.1. Data collection tool

Data were re-abstracted from a sample of a patient medical record and coded and adjacent data extracted from the canreg5 database and entered to excel with the tool developed for this purpose which was adapted according to the AFCRN cancer data collection tool(44). The tool comprises all important variables listed under the data collection form used by the registry and was the standard tool used by all registries.

Interview guides were developed for cancer registrars and data collectors separately for an interview regarding data quality gaps and the process of data collection and registration of newly diagnosed c ancer cases.

### 3.5.2. Data collection procedure

Cancer data is being entered into the database by using software called can reg5 in the registry routinely, and for this study, data were exported from the database to an excel extraction tool developed for this purpose for further analysis in the registry main office.

For re-abstracting audit Cases with variables were identified using sampled medical registration number (MRN) from the database and extracted to the excel tool.

Sampled patient's files or cards were withdrawn and the corresponding cases with variables from patient medical records were re-abstracted using the excel extraction tool for comparison.

Qualitative data were collected using a telephone interview to identify gaps with the help of interview guides. An audio recording was taken according to participants' agreement from all six participants.

## 3.6. Variables

### 3.6.1 Dependent variable

- Data quality

### 3.6.2 Independent variables

- age, sex, usual residence address,
- date of diagnosis,
- basis of diagnosis( death certificate only, clinical investigation, specific tumor markers, cytology/hematology, histology, histology of primary),
- the primary site of the tumor,
- morphology,
- behavior, and stage

## 3.7. Operational definition

**Data quality in the registry**: is defined as a registry that follows international agreements and guidelines to register and code, captured all cancer cases in a defined geographic area, fulfilling accuracy recommended by IACR, and reporting data within the recommended time frame.

**Comparability** is the level to which the whole procedure and practice of registration coding and classification of neoplasms in the registry, adhere to agreed international guidelines.

**Completeness** is the degree to which all diagnosed neoplasms within a registry's catchment population are captured and registered in the registry database.

**Validity** is the accuracy of the registry described by a given characteristic (e.g. site and age) which truly have the attribute.

**Timeliness** is the recommended time frame for registering and reporting diagnosed cancer cases in the registry.

**Re-abstracted** data is the data abstracted from sample patients' medical records and coded into the data extraction tool.

**Adjacent data** is the sample data extracted from canregs5 and entered into a data extraction tool for agreement analysis.

## 3.7 Data analysis procedures

Comparability of the registry was evaluated by reviewing definitions, guidelines, and standards used in the registry.

A total of 11260 cancer data from 2012 to 2016 were Extracted from the registry(CanReg5) and exported to excel. Annual age-standardized rates(ASR) were calculated using the world standard population and checked for abnormal trends for five years diagnosis period for the top five neoplasms in each sex. Overall ASR was compared with neighboring African countries(40).

Age-specific incidence curves of the most common cancers in each sex were presented for five years period from 2012 to 2016 and observed for abnormal fluctuations. Polynomial and linear regression, one-way ANOVA, Non-parametric test were employed for historic methods.

Childhood age-specific incidence rates for age groups 0-4, 5-9, 10-14 were determined and compared with standards published in cancer incidence in five continents Volume X(CI5 V.X)(19).

The agreement rate between the re-abstracted data and database was calculated, and the k coefficient or Kappa statistics and intraclass correlation coefficient were estimated using SPSS version 23 for each variable.

Percentage of morphologically verified cases, death certificate only diagnosis, and other and unspecified cases ( C76 to C80) were obtained.

The time difference between diagnosis and registration of cases was computed and the meantime was determined.

Based on information saturation and maximum variety assumption, key informant interviews were conducted for six respondents from the registry and information source hospitals.

Verbatim transcription was done for audio records obtained from key informant interviews, word by word transcription in Amharic language and translated to English language and thematic analysis in line with the quantitative data analysis were undertaken. Qualitative analysis was undertaken using OpenCode version 4.02 software.

## 3.8 Data Quality Management

The canreg5 software program applied in the registry handled automatic checks in assuring the internal validity of cancer data like sex and topography discrepancies.

Re-abstracted data were entered into the excel tool by checking the database and discrepancies were resolved using the ICD-O coding system and guidelines used in the registry.

Finally, after data re-abstraction was completed, data were verified by cancer data registrars in the main office of Addis Ababa Population-based cancer registry.

## 3.9 Ethical consideration

This study was approved by the Institutional Health Research Ethics Review Committee of Addis Ababa University College of Health Science.

All data collection methods, data use rules, and data abstraction techniques were governed according to the data sharing policy of the registry. All information extracted from the registry was kept confidential.

## 3.10 Dissemination of the results

The findings of this study will be presented to the community of Addis Ababa University during open defense, submitted to Addis Ababa University DDCF project, and AAPBCR. Lastly, it will be published in peer-reviewed journals for further use.

# 4. RESULTS

## 4.1. SOCIO-DEMOGRAPHIC CHARACTERISTICS of KII RESPONDENTS

Three respondents were females aged between 26 and 45, and the remaining three respondents were males aged from 25 to 46.

Two of the KII respondents were laboratory technologists by profession, one nurse, one public health, one computer science, and the other one had statistics, computer science, and Laboratory professions. Two respondents were permanent employees of the registry, and four were doing data collection on voluntary bases or they were not accountable for the data collection they did.

Four of the respondents were contact persons or data collectors of the registry and two respondents were central supervisors and data managers.

## 4.2. COMPARABILITY

Comparability is about the systems and procedures used for coding of cancer cases, classifications, recoding and reporting definitions, incidence case and date definitions, the difference between new and recurrence cases.

Topography, morphology, behavior, and grade coding and classification of incident cancer cases in AAPBCR were according to the International Classification of Disease for Oncology(ICD-O3)(41) and converted to ICD-10 for analysis purpose(39).

Incidence date definition and basis of diagnosis were according to guidelines proposed by the International Association for Cancer Research(IARC). Rules for coding of multiple primary cancers in the registry followed principles developed by the IACR(8).

Two KII respondents also confirmed the comparability of the registry; the whole data collection and registration process of the registry were according to international agreements, guidelines, manuals, and standards.

Anonymous said, " *Any new registry is expected to be a member of regional cancer registry network, for the membership, the registry needs to be comparable, should follow guidelines proposed by IARC and IACR for coding"*. Another respondent also has confirmed that "*the registry has been using ICD-O coding guidelines for coding and registering any neoplasms in the registry"*.

Two respondents agreed that the Coding of multiple primaries was according to the manual suggested by IARC. A respondent described that he never faced multiple primaries and he doesn’t know how he registered if happened.

Data collectors wrote the diagnosis on the form if it is clinical and attach biopsy result report if the basis of diagnosis were pathological or histological, they didn’t code, coding was taken place by the main registrars.

Anonymous said “*I don’t write the date of diagnosis rather just I printed the biopsy result and attached with the data collection form having background information.*

*Generally, I thought the registry is comparable; the basis of diagnosis, staging, topography, morphology, multiple primaries, etc…are coded according to ICD-O coding guideline*. Stated a respondent.

## 4.3. COMPLETENESS

Completeness of the registry was evaluated using semi-quantitative methods proposed by Bray and Parkin(22) and IARC(34).

### 4.3.1. Annual trends in Age Standardized Incidence Rate (ASR) for common cancers

Annual trends in ASR of the top five cancers in the registry did not exhibit abnormal variation for the period 2012 to 2015, but showed a minor decrease in Colorectum, NHL, and all sites combined in 2016 for males and showed observable fluctuations in all sites combined cancer cases for females for the same period, but the trend for other cases showed no increments in ASR with years of diagnosis. (Fig. 3&4)

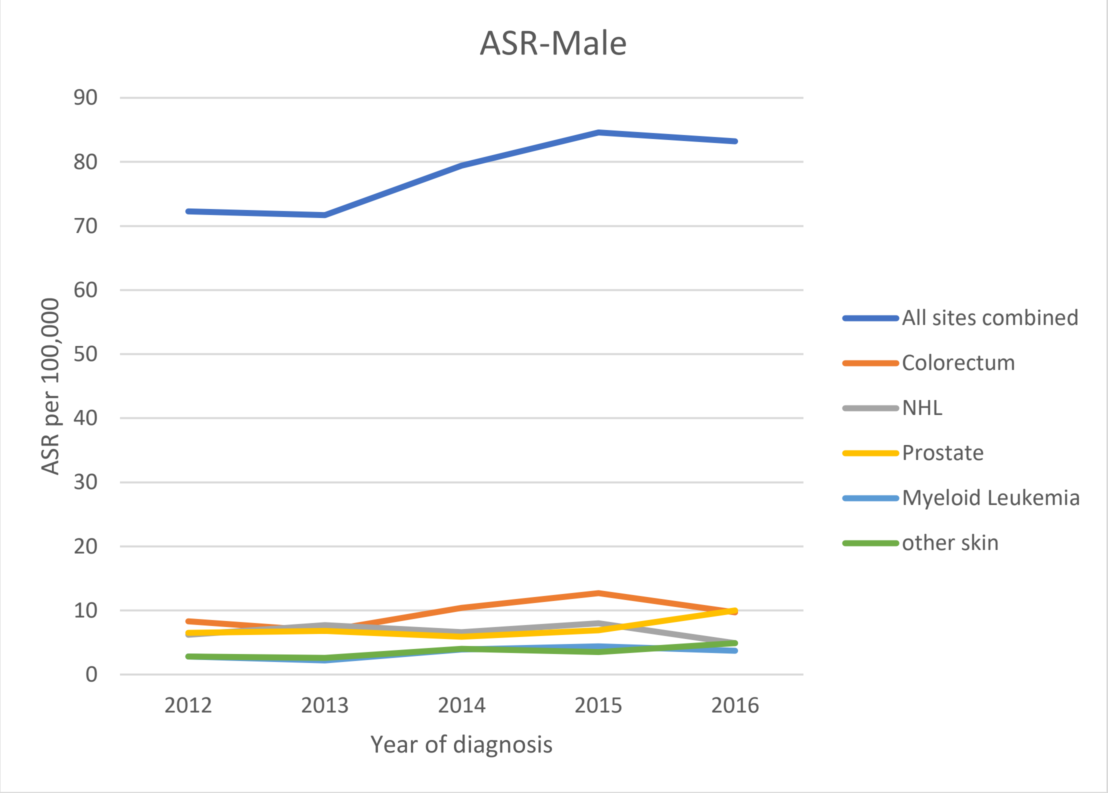

Figure 3: Annual trends in ASR for common cancer cases in Males, Addis Ababa, 2012-2016

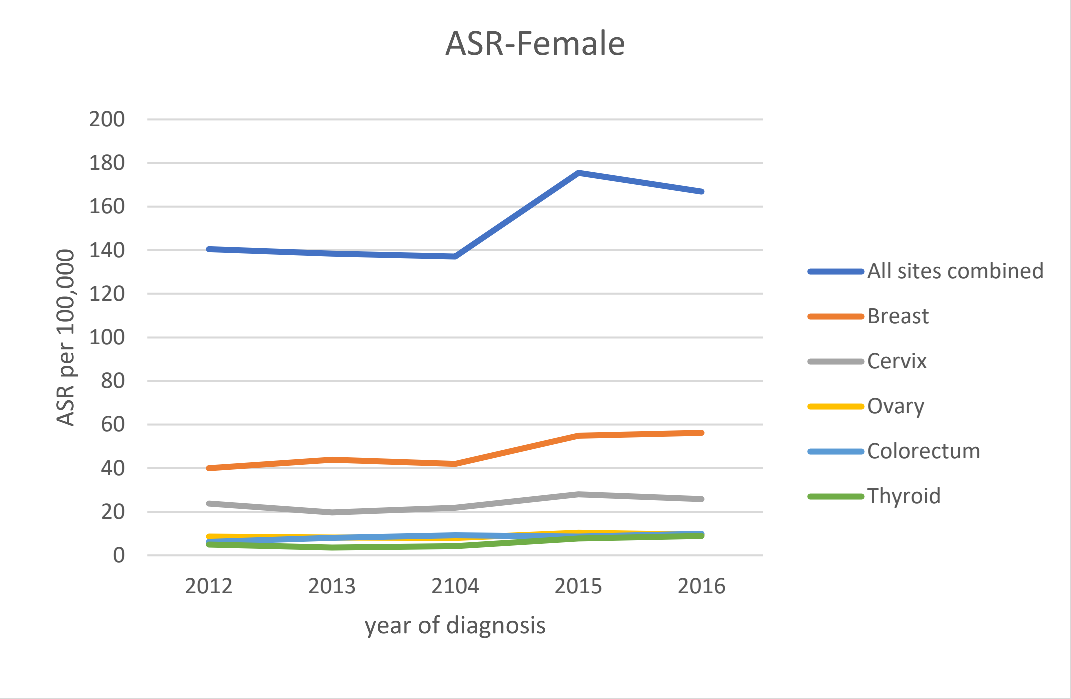

Figure 4: Annual trends in ASR for common cancer cases in females, Addis Ababa, 2012-2016

Linear Regression Analysis Result

The result of linear regression analysis between the year of diagnosis and Age Standardized Rates for the top five cancer cases in both sex showed no significant linear relationship or could not be estimated using linear model, except all site combined cancer cases for males (p-value 0.03). Similarly, non of ASR for common neoplasms in Females had a significant linear relationship with the year of diagnosis, or could not be estimated using a linear model.

The result from linear regression analysis indicated that there was a fluctuation of ASR with years of diagnosis for common cancer cases in both sexes except all combined neoplasms in males.

### 4.3.2. Age–Sex-specific incidence curves

The shape of age-specific incidence curves of breast, cervix, and ovary cancers in females showed a sharp increase between age 45 and age 70 and decline after age 70 and colorectum, non-Hodgkin lymphoma and prostate cancers in males showed increment between age 60 and 80 and decline after age 80, which showed variability in the completeness of the registry. (Fig. 5&6)

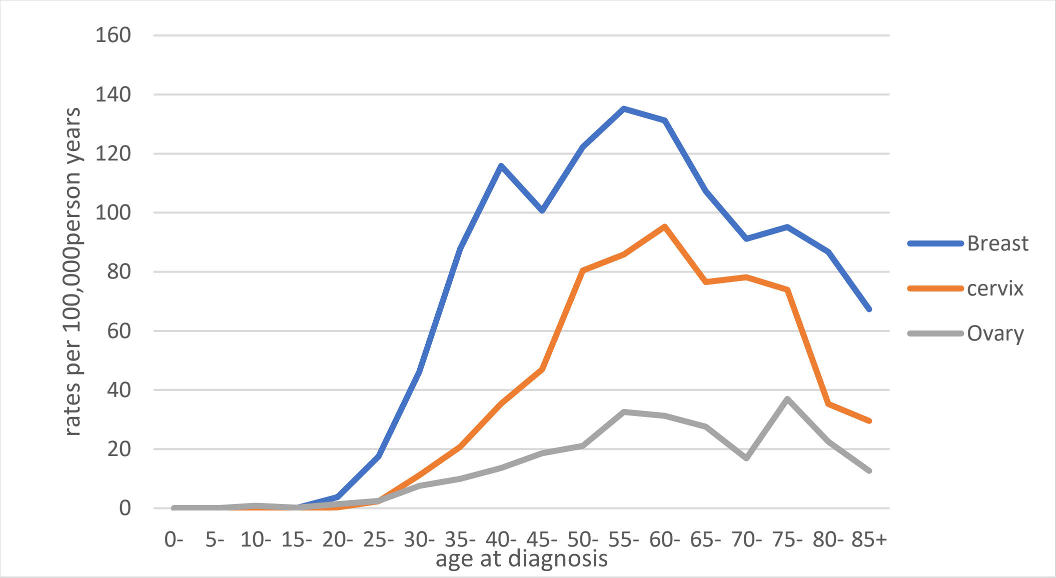

Figure 5: Age-Specific Incidence curve Female, Addis Ababa, 2012-2016

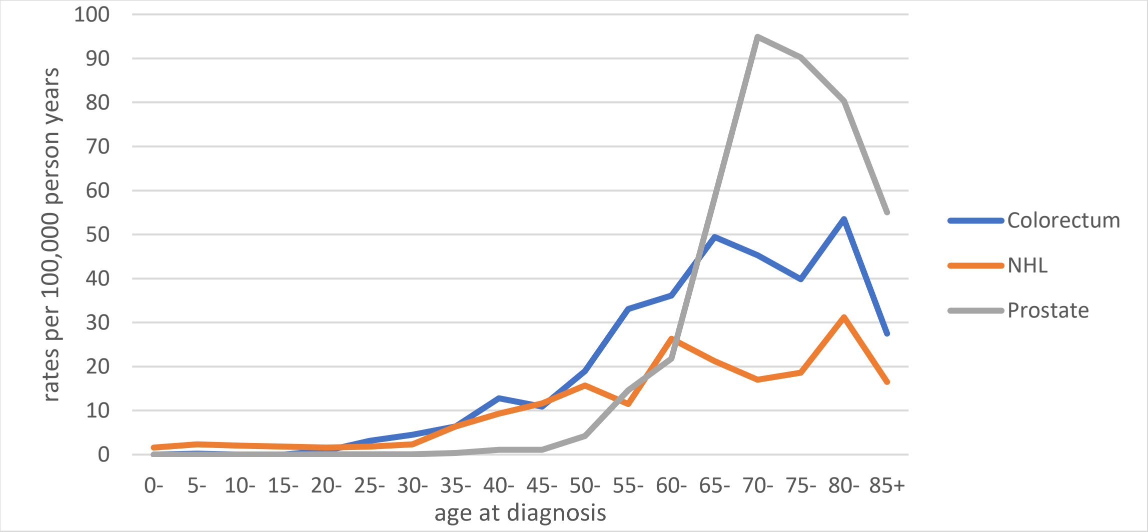

Figure 6: Age-Specific Incidence curve-Male, Addis Ababa, 2012-2016

## Polynomial Regression Analysis Result

The relationship between age group and ASR for common cancer cases in females was not linear and best described by polynomial regression of third-degree or cubic function. Similarly, the relationship between age group and ASR for common cancer cases was not linear and best estimated by the cubic model for colorectal Cancer and Non-Hodgkins Lymphoma(NHL), but quadratic or second degree for Prostate Cancer. The relationship between the age group and the most common neoplasms in both sexes was significant with an α-value of less than 0.01.(Table 1)

Table 1: Polynomial regression of third-degree analysis result

| Male | | | | | | Female | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Cases | $R_2$** | $B_1$ | $B_2$ | $B_3$ | | Cases | $R_2$** | $B_1$ | $B_2$ | $B_3$ |
| Colorectal | 0.907 | -2.252 | 7.687* | -4.685* | | Breast | 0.912 | -0.482 | 6.092 | -5.090* |
| Non-Hodgkin Lymphoma | 0.841 | -1.716 | 6.373 | -2.440* | | Cervix | 0.942 | -2.904 | 11.413* | -8.083* |
| Prostate | 0.782 | -0.603 | 1.455* | | | Ovary | 0.872 | -1.987 | 8.704 | -6.152* |

*-- p-value less than 0.01

** $R^2$ for the Cubic model

Polynomial regression of third-degree (Cubic Model) formula;

$Y=B_0 + B_1X + B_2X^2 + B_3X^3 + \varepsilon$

Y-dependent variable (ASR), X- independent variable (age group)

$B_0$- constant term

$B_1$, $B_2$, $B_3$-coefficients for linear, second degree, and third-degree respectively

$\varepsilon$- an unobserved random error

### 4.3.3. Comparison of Age Standardized Rate for Sub Saharan African Countries

Comparison of ASR(world) per 100,000 for most common cancers(11,40,45) was made with other Sub Saharan African countries, namely; Kenya(Nairobi registry), Mozambique(Beira registry), Uganda(Gulu registry), Zambia(Lusaka registry), and Zimbabwe( Harare registry)(40).

The findings from the registries indicated that the registry was comparable with those SSA countries for colorectum cancers in males and breast, ovary, and all sites in females. The ASR in the registry was relatively low for liver, prostate and all sites in males and cervical cancer in females from the numerical comparison. The ASR for SSA countries was found from publications(40). (Table 2)

**Table 2: Overall ASR for Common Neoplasms in SSA Countries**

| Sites | Ethiopia, Addis Ababa (2012-2016) | Kenya, Nairobi (2012-2014) | Mozambique, Beira (2014-2017) | Uganda, Gulu (2013-2015) | Zambia, Lusaka (2011-2015) | Zimbabwe, Harare (2013-2015) |
|---|---|---|---|---|---|---|
| Male | | | | | | |
| Colorectum | 9.4 | 13 | 2.5 | 1.9 | 4.4 | 16.1 |
| Liver | 2.9 | 4 | 7.2 | 13.6 | 3 | 17.4 |
| Prostate | 7.1 | 49.8 | 18.5 | 23 | 45.5 | 118.6 |
| All sites | 76.5 | 160 | 128 | 109.9 | 112.9 | 327.9 |
| Female | | | | | | |
| Breast | 46.3 | 51.3 | 17.9 | 10.5 | 20.2 | 43 |
| Cervix uteri | 23.3 | 35.5 | 56.9 | 53.6 | 64.7 | 81.6 |
| Ovary | 8.7 | 10.4 | 1.4 | 2.7 | 2.4 | 10.6 |
| All sites | 148.1 | 207.3 | 149.8 | 110.7 | 141.0 | 313.4 |

Statistical comparison of means of total ASR for the top three neoplasms in both sexes using one way ANOVA showed that there was no significant difference of means of ASR between Sub Saharan African countries with a p-value less than 0.05. (Table 3)

Table 3: Comparison of overall ASR in SSA Countries(40)- one way ANOVA, both sexes

| | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| Colorectum | Between Groups | 86.454 | 5 | 17.291 | 0.38 | 0.846 |
| | Within Groups | 272.895 | 6 | 45.483 | | |
| Liver | Between Groups | 93.684 | 5 | 18.737 | 0.392 | 0.838 |
| | Within Groups | 286.485 | 6 | 47.747 | | |
| Prostate | Between Groups | 4026.767 | 5 | 805.353 | 0.495 | 0.772 |
| | Within Groups | 9768.955 | 6 | 1628.159 | | |
| Breast | Between Groups | 748.487 | 5 | 149.697 | 0.241 | 0.93 |
| | Within Groups | 3731.54 | 6 | 621.923 | | |
| Cervix uteri | Between Groups | 1078.9 | 5 | 215.78 | 0.138 | 0.977 |
| | Within Groups | 9379.18 | 6 | 1563.197 | | |
| Ovary | Between Groups | 46.407 | 5 | 9.281 | 0.358 | 0.86 |
| | Within Groups | 155.61 | 6 | 25.935 | | |

### 4.3.4. Childhood age-specific incidence rates

Childhood age-specific rates(per 100,000) were compared with the lowest and highest deciles of incidence rate published in cancer in five continents Volume X(19). The rates were within the recommended range for boys age group 5-9, and 10-14, and for females only age group 0-4, but the value for males age group 0-4, and females age group 5-9 and 10-14 were below the range. (Table 4)

Table 4: Childhood Age-specific incidence rates per 100,000

| Age group | Boys | Reference | | Girls | Reference |
|---|---|---|---|---|---|
| 0-4 | 11.8 | <12.6 to >26.4 | | 12.2 | <12.1 to >23.7 |
| 5-9 | 11.5 | <8.9 to >17.9 | | 6.3 | <7.0 to >13.0 |
| 10-14 | 9.0 | <9.0 to >17.2 | | 6 | <8.2 >16.0 |

Key informants were asked about the population covered by AAPBCR and two respondents agreed that from the population covered by the registry which is the Addis Ababa population there might be missed health facilities with cancer services as the registry included only ten institutions.

The registry did not cover the expected population, there were unincluded health institutions that could be a source of information for the registry.

A respondent said, "*we might miss cases visited in health facilities which are not included to the contact list of the registry and to estimate missing cases in number, approximately 7% to 5% cases might be missed*".

Data collection was interrupted in health facilities due to data collectors turnover, as a result, there were missing cancer cases to include in the registry. This case was explained by a respondent as "*It was started a week ago after more than a year interruption, no data collected and send to the registry for a year, we were doing the work, but not sent to the registry*".

Four respondents concurred that all data collected were from Addis Ababa residents and the population covered by the registry was the Addis Ababa population.

In other words regarding registering cancer cases visited at health facilities, most agreed there is a very low chance of missed cases unregistered because of the nature of the disease, repeated appointment for treatment, and follow-up. But a respondent stated that there was some discontinuation of registering cancer cases because of data collectors turn over.

Anonyms described registering all cancer cases visited the health facility as "*as far as I know in my hospital because it is private hospital and flow of patients is low, there is no chance of missing cancer cases".*

According to the respondents, there were mandatory and optional data elements to be collected using the standard form, and most key informants assured that they filled all required data elements on the form. In contrast, one respondent argued that some data elements like TNM staging, topography code, and morphology codes were completed by physicians and if not completed sent incomplete. A respondent said, "*I have used a one-page data collection form and I complete patients demography and Tumor related information, but TNM staging is difficult to complete because physicians didn't complete TNM staging, this was filled by the registrar".*

Generally, there was missing information agreed by most data collectors, in most cases missing TNM staging was the most common finding from key informant interviews and a respondent stated that the basis difficulty to filled information was ripped and missed patients' medical record.

## 4.4. VALIDITY

The validity or accuracy of AAPBCR was assessed using a re-abstracting audit, the proportion of Morphologically Verified caes and death certificate only case, and the proportion of other and unspecified cases suggested by Bray and Parkin(21) and IARC(34).

### 4.4.1. Reabstracting audit

Reabstracting audit is a method used to check the agreement between source documents and databases. A total of 11260 cancer cases were registered in Addis Ababa Population-Based Cancer Registry(AAPBCR) from 2012 to 2016. Reabstracting was taken place from patients' medical records, from 408 sample size, a total of 390(95.6%) sampled patients' records were re-abstracted and compared with adjacent data or the same cases registered in the registry.

Among 408 sample size, 18(4.4%) of sampled records were not found from the medical registration unit and re-abstracting was continued with 390 patients card. Seven mandatory and tumor-related variables were selected for comparison of re-abstracted and registered data.

From total registered neoplasms from 2012 to 2016, 7582(67.3%) were females, the highest registration occurred in 2016(22.5%), and lowest in 2013 (18.7%). From a total of 390 re-abstracted neoplasms, 277(71%) were females, and the highest was from 2014 (31%) and the lowest in 2012 (12.1%). (Table 5)

Table 5: Re-abstracted data distribution by year of diagnosis and sex

| Study variable | Cases in canreg5 | | Reabstracted cases | |
|---|---|---|---|---|
| | No | % | No | % |
| Year of diagnosis | | | | |
| 2012 | 2120 | 18.8 | 47 | 12.1 |
| 2013 | 2106 | 18.7 | 78 | 20.0 |
| 2014 | 2186 | 19.4 | 121 | 31.0 |
| 2015 | 2319 | 20.6 | 76 | 19.5 |
| 2016 | 2529 | 22.5 | 68 | 17.4 |
| total | 11260 | 100 | 390 | 100 |
| Sex | | | | |
| Male | 3678 | 32.7 | 113 | 29 |
| Female | 7582 | 67.3 | 277 | 71 |
| total | 11260 | 100 | 390 | 100 |

Data accuracy or reliability was presented using two indicators: data agreement rate which is determined using simple mathematical calculation and intercoder reliability with Cohen's κ coefficient, using Kappa statistical analysis. The range of overall agreement rate among variables was from 74.6% to 97.4%, the lowest agreement rate was observed for the date of diagnosis and the highest was for behavior and sex, Kappa coefficient was ranged from 0.438 to 0.963 and there was a significant agreement rate between the database and re-abstracted data with a p-value less than 0.05. (Table 6).

Table 6: Agreement rate and k coefficient between re-abstracted and registered data

| Variables | Observed(No.) | Agreement rate (%) | k Coefficient | P-value |
|---|---|---|---|---|
| Sex | 390 | 97.4 | 0.937 | <.0001 |
| Age | 390 | 96.4 | 0.963 | .000 |
| Date of diagnosis | 390 | 74.6 | 0.745 | .000 |
| Basis of diagnosis | 390 | 89.2 | 0.707 | 000 |
| Topography | 390 | 89.5 | 0.882 | .000 |
| Morphology | 390 | 85.4 | 0.837 | .000 |
| Behavior | 390 | 97.4 | 0.438 | .000 |

The reliability between re-abstracted data and data extracted from the database was also determined using the intraclass correlation coefficient. The intraclass correlation coefficient for the relationship between abstracted data and data extracted from the database for all variables were high near to one, but low for only Behavior code. The relationship between abstracted data and data from the database was significant for all variables with a p-value of less than 0.5. (Table 7)

Table 7: Intraclass correlation coefficient for data re-abstracted and extracted data from the database

| Variables | Intraclass Correlation Coefficient | 95% Confidence Interval | | P-value |
|---|---|---|---|---|
| | | Lower Bound | Upper Bound | |
| Sex | 0.967 | 0.960 | 0.973 | 0.001 |
| Age | 0.999 | 0.998 | 0.999 | 0.001 |
| Date of Diagnosis | 0.989 | 0.987 | 0.991 | 0.001 |
| Basis Of Diagnosis | 0.851 | 0.818 | 0.878 | 0.001 |
| Topography | 0.763 | 0.711 | 0.806 | 0.001 |
| Morphology | 0.966 | 0.959 | 0.972 | 0.001 |
| Behavior | 0.391 | 0.256 | 0.501 | 0.001 |

### 4.4.2. The proportion of Morphological verification and Death certificate only cases

The proportion of morphologically verified cases (MV%) and death certificate cases only (DCO) were calculated for common neoplasms in the registry from the year 2012 to 2106. Overall cancer cases diagnosed with morphological verification(MV%) except non-melanoma of skin were 87.2% for males and 90.2% for females. The proportion of morphologically verified cases was very high for melanoma of the skin in males and larynx in females which were 100 %, and lowest was for pancrease in males and for brain and CNS in females which were 52% and 42.9% respectively Overall cancer cases identified using death certificate(DCO%) except for non-melanoma of skin were 0.11 and 0.14% for males and females respectively.

The mean difference between MV% for Males and Females was 0.2824, and P-value 0.882. This showed no significant difference between MV% for Males and Females at α-value less than 0.05. Multiple comparisons between sites using Kruskal Wallis independent t-test showed a significant difference with p-value 0.017, but a pairwise comparison showed no significant difference between sites for MV%. Table 8)

Table 8: MV% and DCO% for common cancers in both sexes, Addis Ababa, 2012-2016

| Sites | ICD10 | Male | | | Female | | |
|---|---|---|---|---|---|---|---|
| | | Cases | MV% | DCO% | Cases | MV% | DCO% |
| Mouth and pharynx | C00-14 | 190 | 97.4 | 0 | 174 | 96.6 | 0 |
| Esophagus | C15 | 83 | 91.6 | 0 | 118 | 80.5 | 0 |
| Stomach | C16 | 161 | 80.8 | 0 | 160 | 79.4 | 0 |
| Colon, rectum& anus | C18-21 | 461 | 90 | 0 | 442 | 85.8 | 0 |
| Liver | C22 | 126 | 55.6 | 0 | 131 | 60.3 | 0 |
| Pancreases | C25 | 50 | 52 | 2 | 53 | 66 | 0 |
| Larynx | C32 | 50 | 92 | 0 | 8 | 100 | 0 |
| Lung, trachea & bronchus | C33-34 | 165 | 71.5 | 0.61 | 136 | 77.9 | 0 |
| pleura & other thoracic | C73-38 | 32 | 87.5 | 0 | 38 | 97.4 | 0 |
| Melanoma of Skin | C43 | 19 | 100 | 0 | 18 | 94.4 | 0 |
| Breast | C50 | 149 | 96 | 0 | 2498 | 95.9 | 0.04 |
| Cervix Uteri | C53 | NA | NA | NA | 1061 | 94.3 | 0.09 |
| Corpus uterus NOS | C54-55 | NA | NA | NA | 178 | 88.2 | 0 |
| Ovary & adnexa | C56 | NA | NA | NA | 444 | 77 | 0.23 |
| Prostate | C61 | 260 | 81.2 | 0 | NA | NA | NA |
| Testis | C62 | 41 | 87.8 | 0 | NA | NA | NA |
| Kidney, Urinary NOS | C64-66 | 101 | 77.2 | 0 | 102 | 71.6 | 0 |
| Bladder | C67 | 154 | 81.8 | 0 | 66 | 80.3 | 0 |
| Brain, CNS | C70-72 | 59 | 61 | 0 | 63 | 42.9 | 3.17 |
| Thyroid | C73 | 103 | 93.2 | 0 | 343 | 94.8 | 0 |
| Lymphoma | C81-88, C90 | 434 | 94.5 | 0 | 292 | 93.5 | 0.34 |
| Leukemia | C91-95 | 355 | 98.3 | 0.56 | 295 | 98.3 | 0.34 |
| All sites but C44 | C00-96, but C44 | 3510 | 87.2 | 0.11 | 7328 | 90.2 | 0.14 |

The proportion of morphological verification for both sexes in AAPBCR was better than registries in other six SSA countries; Zimbabwe 9Harare), Zambia(Lusaka), Uganda(Gulu), Tanzania(Kilimanjaro), Mozambique(Beira), and Kenya(Nairobi). (Fg. 7)

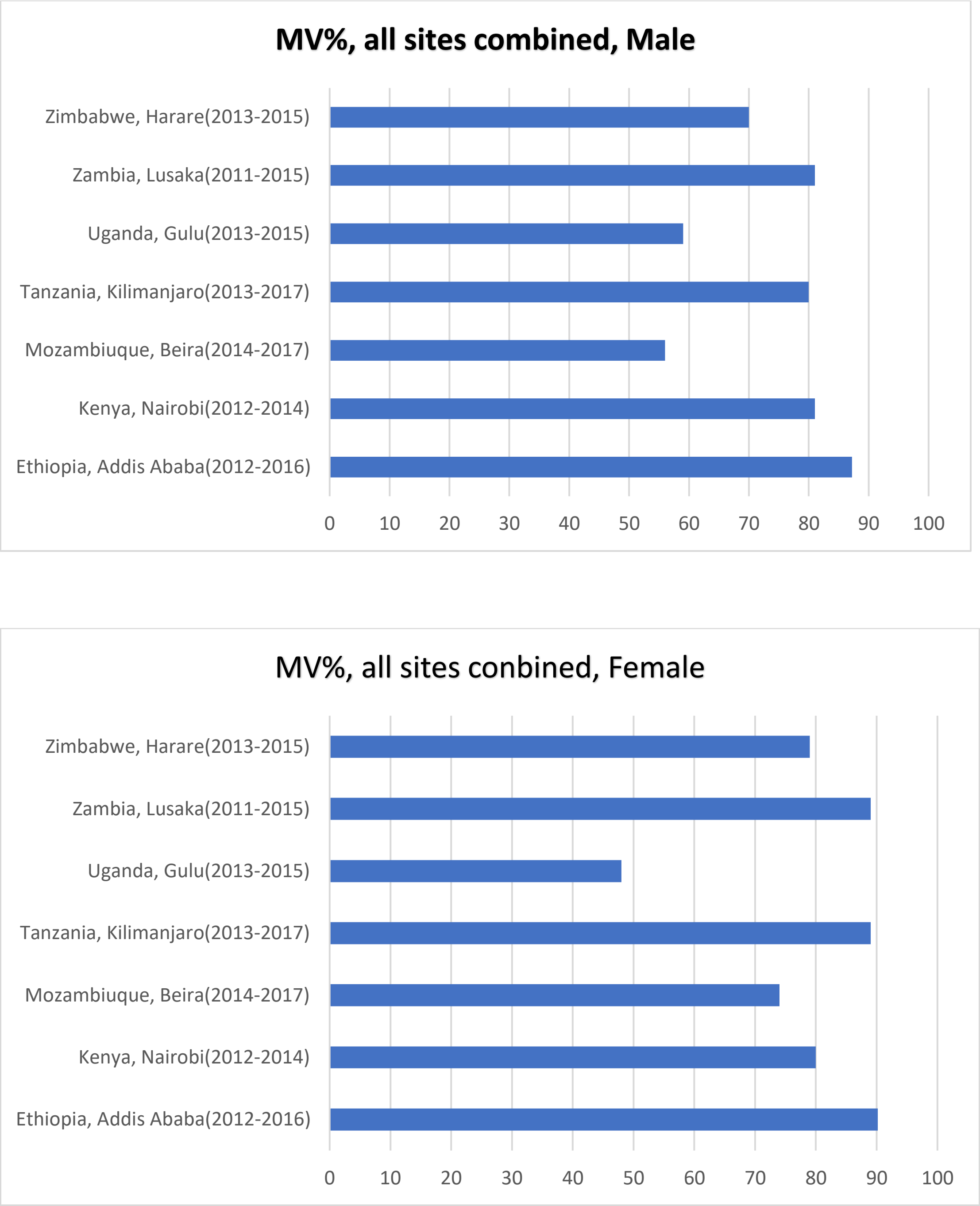

Figure 7: MV% Comparison with Sub Saharan African Countries

### 4.4.3. Other and Unspecified cases

The summarized Proportion of other and unspecified cases were 2.6 % and decreased from 4.9% in 2012 to 1.6% in 2016. (Table 9)

Table 9: The proportion of other and unspecified illness by year of diagnosis

| Year of Diagnosis | Total Cases | O & U cases | % |
|---|---|---|---|
| 2012 | 2120 | 104 | 4.9 |
| 2013 | 2106 | 44 | 2.1 |
| 2014 | 2186 | 61 | 2.7 |
| 2015 | 2319 | 45 | 1.9 |
| 2016 | 2526 | 40 | 1.6 |
| Total | 11260 | 294 | 2.6 |

Two key informants were interviewed for the mechanisms of checking the accuracy of the registry, and both respondents agreed that cross-checking with data collection form and source documents were conducted before registration of the case into the database. In addition to cross-checking, the registry provided training for data collectors to minimize accuracy issues.

Anonyms said "*we have contact persons, they registered cases if from Addis Ababa, according to the training we gave. They registered demographic related data, tumor-related data, and treatment-related data separately, experts from the registry checked the registered data. Assuring data quality onsite was made by registrars before submission. I think the data we registered is accurate*".

Data collectors interviewed for the basis of diagnosis had a similar understanding of the selection of basis of diagnosis and they followed the same coding. Three respondents concurred the chronology of selection of the method of diagnosis first microscopic(histology, cytology) and then nonmicroscopic ( hematology, CT-scan, clinical). In contrast, one data collector did not complete the basis of diagnosis, rather the physician did.

A participant described the basis of diagnosis as "*Most of the time pathology/histology is the basis of diagnosis especially for breast cancer after mastectomy, a sample taken from the site and if the basis is biopsy I choose histology as a basis of diagnosis. Otherwise, secondary choice next to biopsy is the CT scan, I wrote CT scan".*

Concerning the date of diagnosis among four respondents, two followed the same instruction of taking the date sample collected as the date of diagnosis and inconsistent with this, the other one did not write the date of diagnosis rather print and attached biopsy result with the data collection form. One respondent used date laboratory results reported as the date of diagnosis before he got the training and after training, he also agreed with the first respondents.

*"I took from history that the physician diagnosed and collect pathology sample, pathology confirmed date" a*ssured a participant.

Inconsistency of data in the registry was described by missing data elements due to ripped and lost documents like patients' card and laboratory results because file handling is still paper-based in almost all facilities.

Participant from the registry enlightened this as "*Data handling---most facilities still used paper-based data compilation. Due to these documents may be lost, incomplete, unreadable documents some may cause accuracy and completeness problems"*.

Participants certain that most neoplasms were diagnosed using histology and cytology which means identification of cancer cases in most cases were by morphological verification.

## 4.5. TIMELINESS

The mean time difference from diagnosis to the registration of neoplasms in the AAAPBCR was calculated from 390 sampled data collection forms and was 96.8 (80.6,113) days with a 95% confidence interval. The defined period of the report or notification from contact institutions to the registry was one month, and publication of summary results was sent only once.

From a total of 390 cases, 256 (65.6%) of cases were registered within a month, and 51 (13.%) were registered to the registry after four months. (Table 10)

Table 10: The proportion of reporting time, Addis Ababa, 2012-2016

| Reporting time | No of cases | % |
|---|---:|---:|
| Less than 1 month | 256 | 65.6 |
| 1 t0 2 months | 50 | 12.8 |
| 2 to 3 months | 17 | 4.4 |
| 3 to 4 months | 16 | 4.1 |
| More than 4 months | 51 | 13.1 |
| Total | 390 | 100.0 |

Key informant interview findings revealed that the average period of the report or notification from contact institutions to the registry was around one month, and the publication of summary results was sent only once.

Participants from contact institutions were also asked about the time of registration of cases and notification to the AAPBCR and registration were taken place as cancer cases are identified. Except for one who said notification may be delayed up to three months, others agreed that they submit collected data monthly to the registry.

Anonyms explained notification or reporting as "*I collect the data and submit to the registry in person monthly, but sometimes there may be delayed submission*".

Data from the registry was reported for those who need the data and it was published only one time in 2016. Participant from the registry assured that "*we published cancer data every five years, we have already published it by 2016 and now its time for the second publication*".

## 4.6. Data Quality Gaps and Challenges in AAPBCR

The key informant interview identified data quality gaps and challenges in the AAPBCR, and key informants described different quality gaps. Training for data collectors was the main gap in the registry raised by almost all data collectors. Three of four respondents explored that they didn't have any training before cancer data collection. In contrast, one participant took initiation training abroad. Once started there was no training and information update in the registry. A participant confirmed this and she described her start as "*No training I received, but only one short training before three years. The main registrar gave me an onsite type training and data collector before I showed me some issues".*

The use of standardized data collection form was another gap obtained from key informant interview participants. Two of four participants used standard data collection form developed by the AFCRN, one participant wrote variables on white paper and the last participant printed a biopsy report and attached with patients background information.

A participant agreed "*There are minor challenges like availability of data collection form, I got it from the pathology unit, if the unit is not there, I obliged to register on white paper. I can't print the form, even if it is not as such difficult, it is just to tell you*".

Superimposed responsibility and accountability were other major gaps recognized by participants. All except participants from the registry were superimposed this data collection responsibility over their permanent job and there was no accountability for the data collection process. Data collectors were not accountable. All four data collectors were doing data collection voluntarily.

Anonyms affirmed that "*In addition to my responsibility as a staff nurse in the hospital, I am collecting cancer data from Addis Ababa residents coming to our hospital".* Adding on this she said "*The other challenge for me is this responsibility is superposed over my responsibility, sometimes it may take my time. Perhaps, sometimes if I am loaded with duties, I record the card number and withdrawn again".*

The missing health facilities providing cancer diagnosis ad treatment was a significant gap in population coverage of the Addis Ababa population which affect the completeness of the registry.

A participant described missing health facilities and cancer cases as " *We might miss cases visited in health facilities which are not our contacts to estimate in number, 7% to 5% cases might be missed"*.

Missing cases because of data collectors turnover was also an indispensable data quality gap in the registry explored by participants. Cancer data collection was major responsible for non of data collectors and collecting data is only their willingness.

Respondents from the registry confirmed, "*we recruit and trained contact persons, then because of any reason the data collector might stope data collection, and in between, there might be missing cases"*.

The last gaps identified were delayed submission of collected data to the main registry and incomplete information on the form. TNM staging was not completed by some physicians and data collectors leave incomplete.

The agreed time of submission for the registry was monthly as respondents, but sometimes there was delay submission.

The delayed submission was described by a respondent as "*We send completed forms, they collect monthly, and sometimes they might collect after three months"*.

Another respondent confirmed that "*I collect the data and submit to the registry in person monthly, but sometimes there may be delayed submission"*.

Challenges for collection and registration of neoplasms were narrated by participants, and they sketched major challenges in the registry.

The shortage of budget was the main challenge explored by participants and agreed all that budget constraint is the main challenge to cover all target population in Addis Ababa. The source of the budget for the registry was donor-based, not enough even to recruit data collectors.

The budget shortage was described by a participant as *"There is no enough budget to collect data from all Addis Ababa population, and we want to add useful variables like comorbidity with HIV, we can see Kaposi sarcoma, non-Hodgkin lymphoma, but we can't expand because of budget"*.

Another challenge was data collectors turnover since the data collection was voluntarily no consistent data collection. The cases were narrated by a respondent as "*data collector's incentive is very low, and data collectors' turned over every time due to this data collectors might stop collecting data, and cases missed during this time".*

The more focused challenge by respondents was not computerized data collection system, and still using a paper-based collection, due to this difficult to retrieve lost medical records and patient files. Most hospitals used paper-based registration of cases and data collection was from medical records and patient files.

Finally, the registry office was very narrow which can't accommodate staff and files. The office was described by a participant as "*the registry office is about 20- 25 m2 spaced room, staff, documents, very crowded, lack of space… crowding might cause the quality of data and difficulty to retrieve cases easily*".

# 5. DISCUSSION

The assessment of data quality in the Addis Ababa population-based cancer registry was performed according to the dimensions suggested by IARC. The four dimensions used to assess the data quality were Comparability, Completeness, Validity, and Timeliness. Furthermore, key informant interview was employed to identify data quality gaps in the registry.

Comparability the scope of using systems and procedure for coding and classification in the registry were according to international agreements and guidelines. The coding of neoplasms in the registry was following the ICD-O coding system. Incidence date definition, the basis of diagnosis, and rules for coding of multiple primary cancers in the registry followed principles proposed by IARC. Qualitative findings also guaranteed that the whole process of data collection and registration were taken place according to international guidelines and manuals. There was the same finding in Singapore, Ukraine, and Gambia cancer registries (29–32), the registries were using international guidelines and agreements to register and report neoplasms. This is because registries are obliged to follow international guidelines and agreements to be a member of cancer registry networks in their respective regions since they were established.

Assessment of completeness of the registry using semi-quantitative methods of historic methods disclosed that there was some variability of completeness. First, the annual trends in Age Standardized Incidence Rate (ASR) for common cancers showed no linear relationship between year of diagnosis and annual ASR, except all combined cases in males. This indicated that there were heterogeneity and variability in the completeness of the registry for cancer cases in both sexes Similarly, the Gambia cancer registry showed some fluctuation in the trend of ASR from 1990 to 2009(32). This finding contradicts the fact that the number of cancer cases increased as the population increased. Different from the finding in AAPPCR, completeness assessment of the Iceland, Ukraine, and Singapore cancer registries showed relatively stable ASR for the evaluation period(30,31,35). This can be justified by the qualitative finding as the registry was not able to cover all the Addis Ababa population due to budget constraints. And further explained by the low healthcare-seeking behavior of cancer patients in the setting(46,47).

In other cases, discontinuation of data collection in some health facilities for a year due to data collectors turnover could be contributing to the variability of completeness manifested in the annual trend.

The shape of age-specific incidence curves for breast, cervix, and ovary cancers in females showed a sharp increase between age 45 and age 70 and decline after age 70, and colorectum, non-Hodgkin lymphoma, and prostate cancers in males showed increment between age 60 and 80 and decline after age 80, which indicated in variability in the completeness of the registry before age 25 and after age 70 in females and before age 45 in males. This is because early detection and diagnosis of cancer disease are low and the rate of health-seeking after the advanced stage is increased(48).

Even though the ASR curve is expected to increase with age, in females there was a sharp decline after age 70 due to decreased medical seeking al old age(49), indicated incomplete ascertainment for the group. The atypical shape of the age-specific incidence rate curve was similar to the finding in the Gambia cancer registry, in which there was a rapidly declining incidence rate among the elderly in both sexes(32). In contrast to this, the shape of the age-specific incidence curve in Singapore for both sexes was typical, showing no variability in the completeness of the registry(30).

Mathematical comparison of Age Standardized Incidence Rate for most common cancer cases in both sexes with Sub Saharan African Countries, namely; Kenya(Nairobi registry), Mozambique(Beira registry), Uganda(Gulu registry), Zambia(Lusaka registry), and Zimbabwe( Harare registry) revealed that very low ASR for Liver(2.9), prostate (7.1), all sites combined (76.5) in males compared to the above registries in SSA. Similarly, ASR for cervix uteri (23.3) in females was relatively low than other registries. But ASR for colorectum in Males and breast, ovary and all sites combined in females were as comparable as other registries in SSA(40).

Statistical comparison of means of total ASR for the top three neoplasms in both sexes using one way ANOVA showed that there was no significant difference of means of ASR between Sub Saharan African countries with a p-value less than 0.05. There was a difference in ASR between AAPBCR and other five SSA countries, which is an indication of the variability of completeness for Liver, Prostate, and aa sites combined cancer cases in males and cervix uteri in females. But this difference was not significant. Different from this, the Gambia cancer registry was comparable with other west African countries for liver, uterine cervical cancer, and non-Hodgkin lymphoma, but relatively low for prostate, breast cancer, and overall cancer cases(32).

Childhood age-specific rates(per 100,000) comparison with the lowest and highest deciles of incidence rate published in cancer in five continents Volume X(19), showed that the rates were within the recommended range for boys age group 5-9, and 10-14, and for females, only age group 0-4, but the value for males age group 0-4 and females age group 5-9 and 10-14 were below the range. In these age groups, the result indicated that there was a decrease in the completeness of female childhood cancer. This could be further explained by key informant interviews that there were missing of health facilities providing cancer services, due to this cancer cases were left unregistered by the registry. The registry did not cover all health facilities which provide cancer services in Addis Ababa. In another way, there was data collection interruption due to data collectors turnover, this enter caused missing of cancer cases for registration.

The agreement rate between re-abstracted data and data extracted from the database were reasonably high, which can be described by the use of standards and guidelines for coding in both the re-abstracting audit and the database and showed the registry had acceptable accuracy. Similarly, the intraclass correlation association measurement for the relationship between data abstracted and database ranged showed a very significant association with a p-value of less than 0.01. This discovery was also an indication of the very high accuracy of the registry regarding the coding of neoplasms which declared that the registry was using ICD-O coding guidelines to code neoplasms. The Gambia cancer registry recording audit also revealed a similar agreement with the source document, and the average agreement rate was 94%(32).

Acceptable percentage of overall morphological verification (MV%) except non-melanoma of skin was a good indication of a better use of histological identification as the basis of diagnosis for neoplasms in the registry. This results of histological diagnosis confirmed the true attribute for the code of the neoplasms. The proportion of morphologically verified cases was very high for melanoma of the skin in males and larynx in females which were 100 %, and lowest was for pancrease in males and for brain and CNS in females which were 52% and 42.9% respectively. This is because according to the registry rule brain and CNS cancers were diagnosed clinically and there was a rule to register brain and CNS neoplasms whatever the behavior was. There was no significant difference between MV% for Males and Females which indicated a uniform basis of diagnosis in both sexes.

The proportion of morphological verification of AAPBCR was better than SSA Countries(40), and the Ukraine registry, (31) from 2002 to 2012.

But the registry's MV% was less compared with the registries in the Gambia with 94% overall MV% from 2000 to 2009 and Singapore with 88.6% in males and 93.0% in females from 2003 to 2007. This better proportion of morphological verification indicated that neoplasms registered in the registry were representing the true value of the actual cases, which means the registry has a high level of accuracy and also indicated that pathological diagnosis of neoplasms was better than other cancer registries, especially in SSA Countries.

Five years of overall cancer cases identified using death certificate (DCO%) except for non-melanoma of skin were 0.11 and 0.14% for males and females respectively, which indicated very low DCO% and this is an implication of efficient tracing back of cancer cases missed by the normal case finding procedure. Cancer registries in Gambia and Singapore have better DCO% than AAPBCR, 6.6% in males, and 3.6% in females from 2000 to 2009 in the Gambia(32) and, 1.5% in males and 1.3% females for the period 2003 to 2007 in Singapore(30). The low use of the death certificate only (DCO) system might be due to the low availability of death certificate in the Country, and the absence of a system to trace back case from death information.

The five-year Proportion of other and unspecified cases were 2.6 % and decreased from 4.9% in 2012 to 1.6% in 2016. This revealed that pathological diagnosis of cancer cases and identification of specific site had good progress from 2012 to 2016, and important to specify the true value of the code to the topography. Similarly, this indicated that there was an increase in the confirmation of primary sites of cancer cases.

The validity of the registry was also affirmed by key informants that there was cross-checking of data collection forms before coding and registration to the main database, data collectors followed guidelines for the selection of basis of diagnosis, and most neoplasms were investigated using histology and cytology.

Even though the registry had a standard for registration and reporting, there was a delay in registration of cases after identification. It is expected to be registered a case within a month, but the average time of registration to the database was more than three months. This finding was the same as the finding from the key informant interview, that there was a delay in the submission of data to the main registry for up to three months. The reason behind the delay in registration of cases to the database might be the data collection system, which was a passive data collection system.

Major data quality gaps identified by key informant interviews were training for data collectors, use of standard data collection form, superposed responsibility, missing of health facilities which provide cancer services or population coverage, missing cases because of data collectors turnover, and delayed submission of collected data. The underlying cause for these data quality gaps was government commitment for the area and the government gave minimal attention to the cancer registry. Volunteers and donors owned the cancer registration activity and there was no follow-up from the government.

In addition to gaps, challenges in the registry were identified by key informant interviews like; budget, the registry had a very low budget to expand and cover all populations in Addis Ababa, data collectors turnover, paper-based data management, and very narrow registry office. Similar to the gaps, the underlying causes of challenges and constraints were budget and government focus for the registry.

Strengths and limitation of the study

strength

- Used all the four dimensions to assess data quality suggested by IARC,
- Used methods and statistical tests beyond the proposed methods by IARC, and
- Used key informant interviews to triangulate data quality gaps.

Limitations

- Used only historic methods to assess the completeness of the registry

# 6. CONCLUSION

In conclusion, the whole data collection, registration, and coding of neoplasms in AAPBCR were according to international agreements and guidelines, and the APBCR was comparable concerning using international agreements, standards, guidelines, and manuals. The findings in the validity assessment also suggested that the registry was fairly accurate. Despite using only historic methods to assess the completeness of the registry, findings of historic methods and key informant interviews were strongly indicative of the presence of variability in the completeness of the registry. Despite the absence of a standard for timeliness, there was no timely report and registration in the registry. Finally, budget and ownership were the major data quality gaps and challenges.

# 7. RECOMMENDATION

Based on the findings from the assessment of data quality and data quality gaps in the registry, I suggest improving the data collection and registration process from passive to active data collection system in the registry, Further, I recommend;

**Data collectors** to use standard data collection tool developed by AFCRN and to submit collected data to the registry.

**Cancer registry** to completely registering cancer cases identified in health facilities and to expand sources of information or contact health facilities not to miss cancer cases.

**Ministry of health** to owned the cancer registry and to expand the population covered by the AAPBCR and to upgrade the registry to the national level.

Finally, to the scientific community, I recommend further research on quantitative determination of the completeness of the registry, benefits of the cancer registry for cancer prevention, and generating evidence to upgrading the registry to the national level.

# 8. REFERENCES

1. Korir A, Swaminathan R, Ullrich A. PUBLICATIONS POPULATION-BASED CANCER REGISTRATION IN LOW- AND MIDDLE-INCOME SETTINGS PUBLICATIONS POPULATION-BASED CANCER. (43).

2. Storm HH, Michelsen EV, Clemmensen IH, Pihl J. The Danish cancer registry history, content, quality and use. Vol. 44, Danish Medical Bulletin. 1997. p. 535–8.

3. Dimitrova N. T He Process of Cancer Registry.

4. Gakunga R, Parkin DM. Cancer registries in Africa 2014: A survey of operational features and uses in cancer control planning. Int J Cancer. 2015;137(9):2045–52.

5. Leal YA, Fernández-Garrote LM, Mohar-Betancourt A, Meneses-García A. The importance of registries in cancer control. Salud Publica Mex. 2016;58(2):309–16.

6. WHO. National Cancer Control Programmes (page 2556-2560). 2556;

7. Parkin DM. The role of cancer registries in cancer control. Int J Clin Oncol. 2008;13(2):102–11.

8. Cancer registration: principles and methods. IARC Sci Publ. 1991;(95):1–288.

9. Jemal A, Bray F, Ferlay J. Global Cancer Statistics. 2011;61(2):69–90.

10. Parkin D, Ferlay J, Jemal A, Borok M, Manraj SS, N'da G, et al. Cancer in Africa Cancer in Sub-Saharan Africa. IARC Scientific Publication 167. 2018;

11. Timotewos G, Solomon A, Mathewos A, Addissie A, Bogale S, Wondemagegnehu T, et al. First data from a population based cancer registry in Ethiopia. Cancer Epidemiol [Internet]. 2018;53(February):93–8. Available from: https://doi.org/10.1016/j.canep.2018.01.008

12. Memirie ST, Habtemariam MK, Asefa M, Deressa BT, Abayneh G, Tsegaye B, et al. Estimates of Cancer Incidence in Ethiopia in 2015 Using Population-Based Registry Data. J Glob Oncol. 2018;(4):1–11.

13. World Health Organisation. Data quality review: atoolkitfor facility data quality

assessment.Module1. framework and metrics. 2017.

14. Cai L, Zhu Y. The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. 2015;1–10.

15. Teklegiorgis K, Tadesse K, Mirutse G, Terefe W, City DD, Dawa D, et al. Level of data quality from Health Management Information Systems in a resources limited setting and its associated factors , eastern Ethiopia Previous literatures. 2004;1–8.

16. Ouedraogo M, Kurji J, Abebe L, Labonte R. A quality assessment of Health Management Information System ( HMIS ) data for maternal and child health in Jimma Zone , Ethiopia. 2019;1–12.

17. Roadmap IR. 7 refernce Information Revolution Roadmap. 2016;(April).

18. EPHI. Ethiopia Health Data Quality Review : System Assessment and Data Verification for Selected Indicators Ethiopia Health Data Quality Review : System Assessment and Data Verification. 2016;

19. Forman D, Bray F, Brewster DH, Gombe Mbalawa C, B Kohler, Piñeros M, Steliarova-Foucher E R y SFJ. Cancer incidence in five continents, vol. X. Scientific Publication [Internet]. Vol. 164, IARC Scientific Publications. 2014. 1436 p. Available from: http://publications.iarc.fr/Book-And-Report-Series/Iarc-Scientific-Publications/Cancer-Incidence-In-Five-Continents-Volume-X-2014

20. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.

21. Bray F, Parkin DM. Evaluation of data quality in the cancer registry: Principles and methods. Part I: Comparability, validity and timeliness. Eur J Cancer [Internet]. 2009;45(5):747–55. Available from: http://dx.doi.org/10.1016/j.ejca.2008.11.032

22. Parkin DM, Bray F. Evaluation of data quality in the cancer registry: Principles and methods Part II. Completeness. Eur J Cancer [Internet]. 2009;45(5):756–64. Available from: http://dx.doi.org/10.1016/j.ejca.2008.11.033

23. Cancer: Overview, causes, treatments, and types [Internet]. [cited 2019 Sep 5]. Available from: https://www.medicalnewstoday.com/articles/323648.php

24. Irigaray P, Clapp R, Hardell L. Lifestyle-related factors and environmental agents causing cancer : An overview. 2008;(May 2019).

25. Ferlay J. Estimating the global cancer incidence and mortality in 2018 : GLOBOCAN sources and methods. 2019;

26. Woldeamanuel YW, Girma B, Teklu AM. Cancer in Ethiopia. Lancet Oncol [Internet]. 2013;14(4):289–90. Available from: http://dx.doi.org/10.1016/S1470-2045(12)70399-6

27. Misganaw A, Haregu TN, Deribe K, Tessema GA, Deribew A. National mortality burden due to communicable , non-communicable , and other diseases in Ethiopia , 1990 – 2015 : findings from the Global Burden of Disease Study 2015. 2017;1–17.

28. African Cancer Registry Network [Internet]. [cited 2019 Sep 17]. Available from: https://afcrn.org/index.php/about-us

29. Wanner M, Matthes KL, Korol D, Dehler S, Rohrmann S. Indicators of Data Quality at the Cancer Registry Zurich and Zug in Switzerland. Biomed Res Int. 2018;2018.

30. Fung JWM, Lim SBL, Zheng H, Ho WYT, Lee BG, Chow KY, et al. Data quality at the Singapore Cancer Registry: An overview of comparability, completeness, validity and timeliness. Cancer Epidemiol [Internet]. 2016;43:76–86. Available from: http://dx.doi.org/10.1016/j.canep.2016.06.006

31. Ryzhov A, Bray F, Ferlay J, Fedorenko Z, Goulak L, Gorokh Y, et al. Evaluation of data quality at the National Cancer Registry of Ukraine. Cancer Epidemiol [Internet]. 2018;53(February):156–65. Available from: https://doi.org/10.1016/j.canep.2018.02.002

32. Shimakawa Y, Bah E, Wild CP, Hall AJ. Evaluation of data quality at the Gambia national cancer registry. Int J Cancer. 2013;132(3):658–65.

33. Leinonen MK, Miettinen J, Heikkinen S, Pitkäniemi J, Malila N. Quality measures of the population-based Finnish Cancer Registry indicate sound data quality for solid malignant tumours. Eur J Cancer. 2017;77:31–9.

34. Shin HR, Curado MP, Ferlay J, Heanue M, Edwards B, Storm H. Chapter 5 : Comparability and quality of data. Changes. 2004;67–94.

35. Sigurdardottir LG, Jonasson JG, Stefansdottir S, Jonsdottir A, Olafsdottir GH, Olafsdottir EJ, et al. Data quality at the Icelandic Cancer Registry: Comparability, validity, timeliness and completeness. Acta Oncol (Madr). 2012;51(7):880–9.

36. Central Statistical Agency. Population Projections for Ethiopia from 2007 to 2037. 2013;(July).

37. Wisdom J, Creswell JW. Integrating quantitative and qualitative data collection and analysis while studying patient-centered medical home models. Agency Healthc Reseach Qual [Internet]. 2013;(13-0028-EF):1–5. Available from: http://pcmh.ahrq.gov/sites/default/files/attachments/MixedMethods_032513comp.pdf

38. JOHN W. CRESWELL. Qualitative, Quantitative and Mixed Methods Approach. SAGE publications; 2014.

39. World Health Organization. International statistical classification of diseases and related health problems, 10th revision (ICD-10), Fifth version. World Heal Organ [Internet]. 2016;1:332–45. Available from: http://www.who.int/classifications/icd/icdonlineversions/en/

40. Parkin DM, Jemal A, Bray F, Korir AR, Kamate B, Singh E, et al. Cancer in Sub-Saharan Africa- Volume III. uicc. 2019;III.

41. Sobin L, Parkin DM. International Classifi cation of Diseases for Oncology. 2000;240 pages. Available from: http://whqlibdoc.who.int/publications/2000/9241545348_eng.pdf

42. McHugh ML. Lessons in biostatistics interrater reliability : the kappa statistic. Biochem Medica [Internet]. 2012;22(3):276–82. Available from: https://hrcak.srce.hr/89395

43. UCLA Center for Health Policy Research. Using Key Informants. Key Informants. 2016;10.

44. Cancer A, Cancer R, Form N. Afcrn Cancer Registry. (1):1–4.

45. Adedeji OA. Cancer in Sub-Saharan Africa: Current Practice and Future. Cancer in Sub-

Saharan Africa: Current Practice and Future. 2017. 1–217 p.

46. Mesafint Z, Berhane Y, Desalegn D. Health Seeking Behavior of Patients Diagnosed with Cervical Cancer in Addis Ababa, Ethiopia. Ethiop J Health Sci. 2018;28(2):111–6.

47. Gebremariam A, Addissie A, Worku A, Assefa M, Kantelhardt EJ, Jemal A. Perspectives of patients, family members, and health care providers on late diagnosis of breast cancer in Ethiopia: A qualitative study. PLoS One. 2019;14(8):1–14.

48. Solomon S, Mulugeta W. Diagnosis and Risk Factors of Advanced Cancers in Ethiopia. J Cancer Prev. 2019;24(3):163–72.

49. Tesfaye Falaha A. Health Care Seeking Behaviour of Elderly People in Rural Part of Wolaita Zone, Southern Ethiopia. ICUS Nurs Web J. 2016;10(4).

# Annex I:  Data Extraction Tool

Data extraction tool from CanReg5, AAPBCR

| | | | | | Basis of Diagnosis | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| code | sex | age | address | date of diagnosis | death certificate only | clinical investigation | specific tumor markers | cytology/ hematology | histology | primary site of tumor | morphology | behaviour | stage | source of information |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |

## Annex II: Interview Guide for participants from the registry and source medical records

### Interview questions for participants from the registry

1. Would you tell me your background; age, profession, and responsibility in the registry?
2. When and how is the registry started?
3. What do say about the comparability of the registry? How it follows international guidelines to code and classify new cancer cases?
4. How do you describe the completeness of the registry, the population covered by the registry, and how do you assure it?
5. What are the information sources of the registry and how can you validate the accuracy of the registry?
6. How do you collect cancer cases identified in other health facilities in the city?
7. Would you tell me the types of cancer cases included in the registry?
8. Would you mention and explain major data quality gaps in the registry?
9. Do you have special training related to your position here? If yes what type of training are there, and if you have comments.
10. Do you feel there are adequate infrastructure and personnel for the registry to function well? If not why?
11. If you have anything to add…? Thank you for your time!

### Interview questions for participants from the source medical records

1. Would you tell me your background; age, profession, and responsibility in this organization?
2. How do you notify cancer cases to the Addis Ababa cancer registry? How frequent is it?
3. Would you explain to me which cancer cases and when do you notify the registry?
4. What are the challenges in your working area to notify cancer cases to the registry as early as possible?
5. How do you check whether your notification is received and included in the registry?
6. Do you have special training related to your position here? If yes what type of training are there, have your comments.
7. If you have anything to add…? Thank you for your time!

## Annex III: Information sheet and informed voluntary consent form for the head of cancer registry office

My name is__________________________, I am conducting a research in Addis Ababa population-based cancer registry for my Master's Degree in Addis Ababa University College of Health and Medical Sciences. I kindly request you to help me by giving your attention to explaining to you about this study.

**Title of the study**: Assessment of data quality in Addis Ababa population-based cancer registry, Addis Ababa, Ethiopia, 2019/20.

**Purpose**: The purpose of this study is to write a thesis as a partial requirement for the fulfillment of a Master's Degree in public health, and the study results on the status of data quality and data quality gaps in the registry.

**Procedure and duration**: cancer data from the canreg5 database used in the registry will be extracted and entered to excel and this may take not more than two weeks.

**Risk and benefits**: this study will be conducted on secondary data, there is no risk on clients, and there is the only minimal risk on key informants, it may cost some time.

**Confidentiality and Anonymity**: The information that I will get from the registry and collect from participants will be kept confidential. There will be no information that will identify the participants in particular. The findings of the study will be general for the study and will not reflect anything particular of individual persons. The interview guide will be coded to exclude showing names. No reference will be made in oral or written reports that could link participants to the research. An audio recording will be taken with the participant's agreement and not be transferred to a secondary party.

**Rights**: participation in this study is a voluntary basis. The participants have the right to declare to participate or not in this study. If they decide to participate, they have the right to withdraw from the study at any time and this will not label them for any loss of benefits, which they otherwise are entitled. They do not have to answer any question that they do not want to answer.