

https://dspace.org/

The burden of cancers associated with HIV in the South African Public Health Sector, 2004-2014

Dhokotera Tafadzwa

2019
https://hdl.handle.net/10539/34811

Downloaded from a DSpace repository. https://dspace.org/

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

WITS School of Public Health

The burden of cancers associated with HIV in the South African Public Health Sector, 2004-2014

Tafadzwa Dhokotera

A research report submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in partial fulfilment of the requirements for the degree of Master of Science in Epidemiology: Epidemiology and Biostatistics.

Johannesburg, February 2019

TAFADZWA DHOKOTERA 1662599

DECLARATION

I, Tafadzwa Gladys Dhokotera, student number 1662599, declare that this Research Report is my own work under the supervision of Dr Mazvita Sengayi and Dr Jabulani Ncayiyana. It is being submitted in partial fulfilment of the degree of Master of Science in Epidemiology and Biostatistics to the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination to any other University.



Tafadzwa Dhokotera

Johannesburg, February 2019

DEDICATION

This Master's degree is dedicated to my parents who have offered unwavering support and encouragement throughout the whole process my sister Rudorwashe Dhokotera for being my own personal cheerleader and Kuzi and Glory for giving me space to study when asked, thank you.

PRESENTATIONS FROM THIS STUDY

Dhokotera T, Bohlius J, Spoerri A, Egger M, Ncayiyana J, Olago V, Singh E, Sengayi M. Trends in Cancer Risk in HIV-infected Patients in South Africa (2004 – 2014). In: Proceedings of the 22nd International AIDS Conference; 2018 July 23-27; Amsterdam, Netherlands.

Dhokotera T, Bohlius J, Spoerri A, Egger M, Ncayiyana J, Naidu G, Olago V, Singh E, Sengayi M. HIV and Cancer Risk in Adolescents and Young Adults in South Africa, 2004-2014. In: Proceedings of the Conference On Retroviruses and Opportunistic Infections; 2019 March 4-7; Seattle, USA.

ABSTRACT

The impact of South Africa's high human immunodeficiency virus (HIV) burden on cancer risk is not fully understood, particularly in the context of antiretroviral treatment (ART). We examined national cancer trends and excess cancer risk in people living with HIV (PLHIV) compared to those who are HIV-negative. We used probabilistic record linkage to match cancer records provided by the National Cancer Registry to HIV data provided by the National Health Laboratory Service (NHLS). We also used text search of specific HIV terms from the clinical section of pathology reports to determine HIV status of cancer patients. We used logistic and Joinpoint regression models to evaluate the risk and trends in cancers in PLHIV compared to HIV-negative patients from 2004-2014. In sensitivity analysis, we used inverse probability weighting (IPW) to correct for possible selection bias. A total of 329 208 cancer cases from public sector laboratories were reported to the NCR from 2004-2014 with the HIV status known for 95 279 (28.9%) cancer cases. About 50% of all the female cancer cases (n=30 486) with a known status were HIV-positive. PLHIV were at higher risk of AIDS-defining cancers (Kaposi sarcoma [adjusted OR:134.01, 95% CI:111-162], non-Hodgkin lymphoma [adjusted OR:2.73, 95% CI:2.56-2.91] and, cervix [adjusted OR:1.7, 95% CI:1.63-1.77], conjunctival cancer [adjusted OR:21.5, 95% CI:16.3-28.4] and human papilloma virus (HPV) related cancers (including; penis [adjusted OR:2.35, 95% CI:1.85-2.99]), and vulva [adjusted OR:1.94, 95% CI:1.67-2.25] compared to HIV-negative patients. Analysis using the Inverse Probability Weighting yielded comparable results. There is need for improved awareness and screening of conjunctival cancer and HPV-associated cancers at HIV care centres. Further research and discussion is warranted on inclusive HPV vaccination in PLHIV.

Keywords: HIV, cancer risk, South Africa, epidemiology, HPV

ACKNOWLEDGEMENTS

My research supervisors' Dr Mazvita Sengayi and Dr Jabulani Ncayiyana for their support and advice throughout the research process. Ms Ntombizodwa Ndlovu for co-ordinating the research support meeting which helped me stick to timelines and try to be on schedule. All co-authors for their critical comments and advice.

National Cancer Registry, National Health Laboratory Service

National Health Laboratory Service Corporate Data Warehouse

FUNDING

This work was supported by a grant from the U.S. Civilian Research & Development Foundation (CRDF Global). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of CRDF Global.

This work was supported by the NIH administrative supplement to Existing NIH Grants and Cooperative Agreements (Parent Admin Supplement) (The South African HIV Cancer Match Study; U01AI069924-09, PI Matthias Egger, co-PI Julia Bohlius) PEPFAR supplement (PI Matthias Egger) and the Swiss National Science Foundation (The South African HIV cancer Match Study, 320030_169967, PI Julia Bohlius) .The contents are solely the responsibility of the authors and do not necessarily reflect the views of the funding bodies.

TABLE OF CONTENTS

DECLARATION i

DEDICATION ii

PRESENTATIONS FROM THIS STUDY iii

ABSTRACT iv

ACKNOWLEDGEMENTS v

TABLE OF CONTENTS vi

LIST OF FIGURES viii

LIST OF TABLES ix

LIST OF ABBREVIATIONS/ NOMENCLATURE x

1. INTRODUCTION AND LITERATURE REVIEW 2

1.1. INTRODUCTION 2
1.2. BACKGROUND 2
1.3. LITERATURE REVIEW 3
1.3.1. HIV-related cancers 3
1.3.2. Cancer trends in PLHIV 4
1.3.3. Risk factors for HIV-related cancers 6
1.3.4. Cancer risk in PLHIV in South Africa 7
1.3.5. Record Linkage 7
1.4. PROBLEM STATEMENT 8
1.5. JUSTIFICATION 9
1.6. AIM 9
1.7. OBJECTIVES 9
1.8. REFERENCES 10

2. SUBMISSABLE MANUSCRIPT 14

2.1. INTRODUCTION 15
2.2. RESEARCH IN CONTEXT 16
2.2.1. Evidence before this study 16
2.2.2. Added value of this study 16
2.2.3. Implications of all the available evidence 17
2.3. METHODS 17

2.3.1. Study setting and design 17
2.3.2. Study population, variables and data sources 17
2.3.3. Data management 19
2.3.4. Data analysis 19
2.3.5. Sensitivity analysis 20
2.3.6. Ethical considerations 21
2.4. RESULTS 21
2.5. DISCUSSION 30
2.2. REFERENCES 35
3. EXTENDED METHODS AND RESULTS 38
4. OVERALL CONCLUSIONS AND RECOMMENDATIONS 42
REFERENCES 45
APPENDICES 47
Plagiarism form 48
Ethics certificate 49
Turnitin report 50
Journal guidelines 51
Data management and data analysis do files 58

LIST OF FIGURES

Figure2.1: Percentage contribution of ADCs and NADCs to the total cancer burden amongst PLHIV in South Africa, 2004-2014 23
Figure2.2: Trends in cancer risk for selected cancers amongst PLHIV in the South African public health sector, 2004-2014 27
Figure 2.3: Trend in Attributable risk fractions amongst PLHIV in the South African public health sector, 2004-2014 29
Figure 3.1: Trends in cancer incidence amongst PLHIV in the South African Public Health sector, 2004-2014 39

LIST OF TABLES

Table 2.1: Characteristics and distribution of public sector cancer cases by HIV status, 2004-2014 22
Table 2.2 Odds ratios for cancer in PLHIV by complete case analysis and by weighted analysis 25
Table 3.1: South African mid-year population estimates and the estimated HIV prevalence by gender, 2004-2014 39
Table3.2: Comparisons of Age Standardized cancer incidence rates by HIV status in the South African public health sector, 2004-2014 40

LIST OF ABBREVIATIONS/ NOMENCLATURE

ADC – AIDS Defining Cancer

APC – Annual Percentage Change

ARF – Attributable Risk Fraction

ART–Anti-retroviral Treatment

BCAH – Burden of Cancer Associated with HIV

CDW – Corporate Data Warehouse

EBV – Epstein Barr Virus

HIV–Human Immunodeficiency Virus

HPV – Human Papilloma Virus

ICD-0 -3 – International Classification of Diseases for Oncology 3rd edition

IPW – Inverse Probability Weighting

KS – Kaposi's Sarcoma

NADC- Non-AIDS Defining Cancer

NCR–National Cancer Registry

NHL – Non-Hodgkin's Lymphoma

NHLS–National Health Laboratory Service

PLHIV–People Living with Human Immunodeficiency Virus

SAM – South African HIV-cancer Match study

1. INTRODUCTION AND LITERATURE REVIEW

1.1.INTRODUCTION

This review will focus on the burden of both Human Immunodeficiency virus (HIV) and cancer globally and locally in South Africa. It will also discuss the concept of record linkage and present literature where this method has been used in epidemiological studies. A review of published HIV associated cancer research will be presented particularly research that focuses on comparisons between HIV positive and negative individuals. To finish off the chapter, the aims and objectives of the study will be listed.

1.2.BACKGROUND

In Sub-Saharan Africa, HIV is of greatest concern with 25.6 million people currently living with the virus and two-thirds of new infections in the world coming from this region (1). In South Africa, as of 2015, approximately 12.7% of the population was living with HIV (2). There has been a decline in HIV and Acquired Immune Deficiency Syndrome (AIDS) related deaths from the year 2007 attributed to an increase in availability of antiretroviral treatment (ART) which became accessible in South Africa from 2004 (3). As a result, people with HIV are living longer than they used to before the ART era.

However, with the increase in longevity amongst people living with HIV (PLHIV), their cancer risk has also increased especially for those malignancies associated with age. About 25% of cancer cases in low and middle-income countries are attributable to infectious causes and in 2012, 31.3% of the cancers in sub-Saharan Africa were attributable to infections (4,5). HIV infection has always been associated with cancer. A sentinel observation of increased Kaposi Sarcoma cases amongst men who had sex with men led to the discovery of HIV and also to the cancer being included as part of the AIDS case definition at the beginning of the AIDS pandemic (6). Later on, HIV was pronounced a risk factor for cancer by International Agency for Research on Cancer in 1996 (7). This then necessitates the need to investigate the risk, incidence and the burden of cancer in PLHIV compared to HIV negative individuals.

The South African HIV Cancer Match (SAM) study is a probabilistic record linkage study of a national HIV cohort. It is created from National Health Laboratory Service (NHLS) HIV laboratory data (CD4 counts, viral load, HIV tests), linked to the National Cancer Registry (NCR) data, in order to study cancer risk in HIV positive people. The HIV dataset consists of 49 million records for the entire country. The creation of the cohort has been based on names, surname, date of birth, gender, location, national ID number, hospital folder number and laboratory number. Included in the cohort are all individuals with at least one positive HIV ELISA test, a rapid HIV test, a HIV DNA PCR result (in infants); a CD4 cell count or a viral load measurement.

In the primary SAM study, the use of deterministic and probabilistic record linkage techniques was employed to find records belonging to the same individual, with follow-up times being created from the first date of a positive HIV test to the last HIV related laboratory measurement. This pre-processed NHLS HIV data was then linked to the pathology-based NCR data to determine the incidence and risk factors of cancer in PLHIV. The current study, Burden of Cancers Associated with HIV (BCAH) study is nested within the SAM study. The BCAH study aimed to determine the proportion of cancers resulting from HIV as well as the additional risk that PLHIV have compared to HIV negative people or the general South African population. The same underlying principles of record linkage were used to link NHLS HIV data and the NCR pathology-based cancer data.

1.3.LITERATURE REVIEW

1.3.1. HIV-related cancers

The cancers associated with HIV are broadly classified into two groups, the AIDS defining cancers (ADCs) and non-AIDS defining cancers (NADCs) respectively. The ADCs consist of Kaposi sarcoma (KS), Non-Hodgkin's Lymphoma (NHL) and Invasive Cervical Cancer (ICC). The term AIDS defining was first used in the early days of the epidemic to help in diagnosis as the presence of these cancers was suggestive of HIV infection. Conversely, the NADCs

consist of all the other cancers that are more frequent in PLHIV compared to those who are not infected. These malignancies include Hodgkin's lymphoma, lung cancer, pharyngeal cancer and anal cancer. Some researchers further classify the NADCs into infection related and non-infection related malignancies. (8,9).

1.3.2. Cancer trends in PLHIV

Different trends in cancer incidence and risk have been observed depending on the geographical location of the study. In developed countries, significant decreases in the incidence and risk of ADCs have been noted, especially for KS and NHL (10–12). This has largely been attributed to increased and efficient ART distribution resulting in improved immunity. Although most studies have reported NHL as showing a decreasing trend in the ART era, Burkitt's lymphoma, an NHL sub-type has not demonstrated a similar trend.

As the ADCs have been decreasing in incidence and risk, the burden of NADCs has been on the rise in resource rich countries (9,12–15). Increasing trends for anal, skin, liver, oesophageal and lung cancers have been reported especially in comparison with the pre-ART era (16,17). Part of the reason for this increased risk of the named NADCs have been lifestyle related factors that are higher in PLHIV such as smoking (lung and oropharyngeal cancers) and high alcohol intake (liver and oropharyngeal cancers) (10,18). Also elevated amongst PLHIV in the ART era are infection related cancers. Studies in the United States (US) have shown that 40-70% of all cancers in PLHIV are attributable to infectious causes (9,19). Of special interest is HPV (anogenital cancers, and oropharyngeal cancers) and the hepatitis viruses (liver cancer). However, the occurrence of the NADCs has not been consistently high for all cancers amongst PLHIV. There has been no increase in the incidence or prevalence of malignancies such as colon, prostate and breast cancer with some suggesting that this could be as a result of the immune system not being part of the development of these malignancies (18). Moreover,

compared to HIV negative individuals, HIV positive individuals have a lower risk for colon, prostate and breast cancer (20).

On the other hand, developing countries have largely experienced more of the ADCs and other infection related cancers as well as eye cancer (8,21,22). Following the introduction of ART in Africa, substantial decreases in ADCs, particularly KS were noted (8,23–25). However, after the initial decline, in KS incidence and risk, some studies have observed that KS risk is still high for PLHIV(8). Some reports have associated the initial drop in incidence as a reflection of a decrease in HIV prevalence rather than the introduction of ART(23). Other researchers argue that most HIV patients present to the health care centres with a high burden of disease and initiation of ART with advanced illness is not as protective to cancer risk. Opposite to what has been observed in resource rich counties, NHL has not decreased significantly especially Burkitt's lymphoma (8,24). This may be as a result of poor distribution of ART and co-infection with Epstein Barr Virus (EBV) although some reports argue that EBV is unlikely to be the cause as the background risk is extremely high as well. Cervical cancer on the other hand has been on the rise even with the introduction and expansion of ART (24,26–28). This has been attributed to age, high HPV co-infection amongst women with HIV, and the lack of an association between immunosuppression and cervical cancer (8,18,27,29–31). Studies have found that low CD4 counts are not associated with increased cervical cancer risk. Others have reported that, the presence of HIV itself is thought to impede the clearance of HPV from the body thereby further increasing the risk of cervical cancer and other HPV related cancers(31). Overall, NADCs are still lower in incidence and risk in resource limited settings like Africa. However, infection related cancers are now on the rise in the ART era , mostly HPV-related malignancies such as anogenital cancers other than cervix (8). In addition to this, other cancers such as eye and squamous cell carcinoma of the skin have also been associated albeit with limited evidence to HPV(23). Unlike the reports in the developed countries, liver and lung

cancer have not been reported as high risk amongst PLHIV in resource limited countries. Possible reasons have included a lower prevalence of lifestyle related factors such as smoking and increased alcohol intake, which results in increased cancer incidence in resource-rich settings. In addition, under reporting is also a possible reason why the incidence recorded in resource-limited countries is lower than that the incidence in developed countries (32).

1.3.3. Risk factors for HIV-related cancers

The occurrence of ADCs and NADCs in PLHIV is dependent on different risk factors. The risk factors associated with cancers in PLHIV have included immunosuppression due to HIV infection and an increased prevalence in PLHIV of the usual cancer risk factors such as increased survival, smoking and other lifestyle related factors (18,22,33). In addition, there is a growing aging HIV population on ART with a suboptimal immune function. The treatment (ART), does not repair the immune function fully, particularly cell-mediated immunity which is responsible for tumour surveillance in cancer (34). The immune system plays a vital role in the prevention of cancer. This is achieved by the cell-mediated immune system which recognizes and eliminates tumour cells that express tumour specific antigens, clearing them from the system (35–37). The CD4 cells are a part of this cell-mediated immunity therefore, the process is compromised in PLHIV making them at higher risk for cancer. Moreover, continuing immune activation due to delayed treatment is possible in this population as the expansion of ART has not been consistent. This continued immune activation results in the development of non-AIDS related events such as cardiovascular disease and other malignancies (38,39). However, this has not been true for all cancer. Some cancers like cervical cancer are still increasing in risk even with improved immunity and other cancers like prostrate and breast cancer have maintained lower risk amongst PLHIV. With ART availability, PLHIV are also at risk of developing age related cancers as they now live longer (9,11). Another major risk for cancer development in PLHIV is the poor control of co-infections such as Human Herpes virus-8, Human papilloma virus as well as Hepatitis B and C (9,18,22). These viruses

are known oncogenic viruses on their own and therefore pose an additional risk for the HIV infected individuals. Research suggests that HIV impedes the clearance of HPV(28,40).

1.3.4. Cancer risk in PLHIV in South Africa

Prior to the introduction of ART, elevated cancer risk was observed for all ADCs as well as some NADCs such as SCC skin, and anogenital cancer of organs other than cervix (41). Studies in the ART era have shown decreases in incidence and improved survival in KS amongst PLHIV (42). A cross sectional study conducted at a tertiary hospital in Johannesburg also reported a high prevalence (34%) of HIV amongst newly diagnosed adult cancer patients (43). High HPV co-infection in women living with HIV has also been reported as a possible reason as to why cervical cancer risk is higher amongst those who are HIV infected (27).

1.3.5. Record Linkage

Record linkage is defined as the process of combining information from two separate sources or from the same source that belong to the same individual (44). It involves the use of a master file which is the main data set and a file of interest which has the additional information (44). There are many techniques that are used when it comes to record linkage. These include deterministic record linkage, probabilistic record linkage as well as machine learning. Deterministic record linkage works on the basis that there is a known identifying variable(s) that is found in both data sets being linked (44). As a result, the only outcomes that can be obtained from this method are a match and no match. Deterministic record linkage does not take into account the degree of disagreement between the datasets, which is when the probabilistic method comes in. The comparison in probabilistic record linkage leaves room for judgement of whether the records are of the same patient or not. It calculates to what extent the records disagree or partially agree. The machine learning approach to record linkage is also able to discern the degree of a match. In addition to this, it includes the use of a training set

where records that are known to be a match are used to train the algorithm on what a match is and what a non-match is(44).

The feasibility of record linkage approaches has been proven in a number of HIV and cancer studies particularly in the developed world (8,9,13,17,19,21,45). The success of record linkage studies has largely relied on well-defined HIV populations and high quality cancer registries. However, there have been limitations as well. The sample size for some of the studies have been relatively small, socio-economic factors have influenced access to cancer diagnosis and assessment of risk factors has been limited (18,21,22). This would probably lead to possible imprecise estimates of cancers and their burden in PLHIV. In the South African context feasibility of record linkage has been proven. However, it has been mainly done in centralized areas (43,46).

1.4.PROBLEM STATEMENT

Although studies on cancer in people living with HIV (PLHIV) in the ART era have been done, most studies have been in the developed countries where most of the infected people are white men and the main mode of transmission is homosexual and ART has been rolled out since 1996 (9,11,13). In South Africa, the HIV burden is highest in black women, transmission is mainly heterosexual and ART roll out only began in 2004 (2). Studies on cancer risk in PLHIV done in the ART era in South Africa have mostly included HIV cohorts, case series and case control methodologies. Most of these studies have lacked comparisons between individuals who are HIV positive and those who are HIV negative. The study populations have been small and specific to particular demographics which limits the external validity of the research results.

1.5.JUSTIFICATION

On its own, HIV has proven to be a high burden. Many strategies have been implemented to decrease incident cases and to manage the prevalent cases thus increasing longevity. With this increase, there is a possibility of double burden of disease amongst PLHIV as their cancer risk also increases. The cancer burden is unknown and the additional risk of cancer in PLHIV in the South African context is yet to be determined. As a result, linking the HIV data to the cancer registry data would allow for assessment of the burden and additional risk that PLHIV possess an approach that has been used in other studies and proven feasible as mentioned above. By highlighting the cancers that mostly affect PLHIV, the study would help inform public health policies on impact of HIV on national cancer burden and aid in effective cancer resource allocation as well as development of cancer prevention strategies among PLHIV.

1.6.AIM

The overall aim of the study is to estimate the burden of laboratory-diagnosed cancer attributable to HIV in the South African public sector, 2004 - 2014. This will highlight the impact of the HIV epidemic on overall cancer burden, and the potential of HIV control as a cancer prevention measure in South Africa.

1.7.OBJECTIVES

1. To estimate the proportion of cancers reported to the NCR that are associated with HIV.
2. To estimate standardized incidence ratios (SIR) for cancer in HIV positive people compared to HIV negative/unknown with 95% confidence intervals
3. To describe 10-year trends (2004 – 2014) in cancer incidence and spectrum by HIV status in the South African public sector

1.8.REFERENCES

1 World Health Organization. HIV/AIDS - WHO Fact sheet. 2015. https://www.who.int/news-room/fact-sheets/detail/hiv-aids (accessed July 25, 2017).

2 Statistics South Africa. Mid-year population estimates. www.statssa.gov.za/publications/./P03022014.pdf 2015; : 1–18.

3 Shisana O, Rhele T, Simbayi LC, et al. South African National HIV Prevalence, Incidence and Behaviour Survey, 2012. 2014.

4 World Health Organization. Cancer Fact Sheet N297. https://www.who.int/news-room/fact-sheets/detail/cancer. (accessed July 25, 2017).

5 Plummer M, de Martel C, Vignat J, Ferlay J, Bray F, Franceschi S. Global burden of cancers attributable to infections in 2012: a synthetic analysis. Lancet Glob Heal. 2016;4:e609–16.

6 Gopal S, Achenbach CJ, Yanik EL, et al. Moving Forward in HIV-Associated Cancer. J Clin Oncol. 2017;32:876–80.

7 World Health Organization. IARC monographs on the evaluation of carcinogenic risks to humans : Volume 67 Human immunodeficiency viruses and human T-cell lymphotropic viruses. Lyon, 1996 http://monographs.iarc.fr/ENG/Monographs/vol67/mono67.pdf. (accessed July 25, 2017)

8 Dryden-Peterson S, Medhin H, Kebabonye-Pusoentsi M, et al. Cancer incidence following expansion of HIV treatment in Botswana. PLoS One. 2015;10:1–13.

9 Silverberg MJ, Chao C, A Leyden W, et al. Hiv Infection and the Risk of Cancers With and Without a Known Infectious Cause. Aids. 2010;23:2337–45.

10 Engels EA, Biggar RJ, Hall HI, et al. Cancer risk in people infected with human immunodeficiency virus in the United States. Int J Cancer. 2008;123:187–94.

11 Shiels MS, Pfeiffer RM, Gail MH, et al. Cancer burden in the HIV-infected population in the United States. J Natl Cancer Inst. 2011;103:753–62.

12 Crum-Cianflone N, Hullsiek KH, Marconi V, et al. Trends in the incidence of cancers among HIV-infected persons and the impact of antiretroviral therapy: A 20-year cohort study. AIDS. 2009;23:41–50.

13 Dal Maso L, Polesel J, Serraino D, et al. Pattern of cancer risk in persons with AIDS in Italy in the HAART era. Br J Cancer. 2009;100:840–7.

14 De Martel C, Shiels MS, Franceschi S, et al. Cancers attributable to infections among adults with HIV in the United States. AIDS. 2015;29:2173–81.

15 Bedimo RJ, McGinnis KA, Dunlap M, Rodriguez-Barradas MC, Justice AC. Incidence of non-AIDS-defining malignancies in HIV-infected versus noninfected patients in the HAART Era: Impact of immunosuppression. J Acquir Immune Defic Syndr.

2009;52:203–8.

16 Bohlius J, Maxwell N, Spoerri A, et al. Incidence of AIDS-defining and Other Cancers in HIV-positive Children in South Africa. Pediatr Infect Dis J. 2016;35:e164–70.

17 Hernández-Ramírez RU, Shiels MS, Dubrow R, Engels EA. Cancer risk in HIV-infected people in the USA from 1996 to 2012: a population-based, registry-linkage study. Lancet HIV. 2017;4:e495–504.

18 Albini L., Calabresi A., Gotti D., et al. Burden of non-AIDS-defining and non-virus-related cancers among HIV-infected patients in the combined antiretroviral therapy era. AIDS Res Hum Retroviruses. 2013;29:1097–104.

19 de Martel C, Shiels MS, Franceschi S, et al. Cancers attributable to infections among adults with HIV in the United States. AIDS. 2015;23:2173–81.

20 Coghill AE, Engels EA, Schymura MJ, Mahale P, Shiels MS. Risk of Breast, Prostate, and Colorectal Cancer Diagnoses Among HIV-Infected Individuals in the United States. J Natl Cancer Inst. 2018;110:959–66.

21 Mbulaiteye SM, Katabira ET, Wabinga H, et al. Spectrum of cancers among HIV-infected persons in Africa: The Uganda AIDS-Cancer Registry Match Study. Int J Cancer. 2006;118:985–90.

22 Godbole S V., Nandy K, Gauniyal M, et al. HIV and cancer registry linkage identifies a substantial burden of cancers in persons with HIV in India. Medicine. 2016;95:e4850.

23 Chokunonga E, Borok MZ, Chirenje ZM, Nyakabau AM, Parkin DM. Trends in the incidence of cancer in the black population of Harare, Zimbabwe 1991-2010. Int J Cancer. 2013;133:721–9.

24 Jaquet A, Odutola M, Ekouevi DK, et al. Cancer and HIV infection in referral hospitals from four West African countries. Cancer Epidemiol. 2015;39:1060–5.

25 Bohlius J, Valeri F, Maskew M, et al. Kaposi's Sarcoma in HIV-infected patients in South Africa: Multicohort study in the antiretroviral therapy era. Int J cancer. 2014;135:2644–52.

26 Bateman AC, Katundu K, Mwanahamuntu MH, et al. The burden of cervical pre-cancer and cancer in HIV positive women in Zambia: a modeling study. BMC Cancer. 2015;15:541.

27 Adler DH, Wallace M, Bennie T, et al. Cervical dysplasia and high-risk human papillomavirus infections among HIV-infected and HIV-uninfected adolescent females in South Africa. Infect Dis Obstet Gynecol. 2014;2014:498048.

28 Holmes RS, Hawes SE, Touré P, et al. HIV infection as a risk factor for cervical cancer and cervical intraepithelial neoplasia in Senegal. Cancer Epidemiol Biomarkers Prev, 2009;18:2442–6.

29 Rohner E, Sengayi M, Goeieman B, et al. Cervical cancer risk and impact of Pap-based screening in HIV-positive women on antiretroviral therapy in Johannesburg, South Africa. Int J Cancer. 2017;141:488–96.

30 Matovelo D, Magoma M, Rambau P, Massinde A, Masalu N. HIV serostatus and tumor differentiation among patients with cervical cancer at Bugando Medical Centre. BMC Res Notes. 2012;5:406.

31 Adjorlolo-Johnson G, Unger ER, Boni-Ouattara E, et al. Assessing the relationship between HIV infection and cervical cancer in Côte d'Ivoire: a case-control study. BMC Infect Dis. 2010;10:242.

32 Singh E, Underwood JM, Nattey C, Babb C, Sengayi M, Kellett P. South African National Cancer Registry: Effect of withheld data from private health systems on cancer incidence estimates. South African Med J. 2015;105:107.

33 Grulich AE, Wan X, Law MG, Coates M, Kaldor JM. Risk of cancer in people with AIDS. Aids. 1999;13:839–43.

34 Wilson EMP, Sereti I. Immune restoration after antiretroviral therapy: the pitfalls of hasty or incomplete repairs. Immunol Rev. 2013;254:343–54.

35 Dunn GP, Old LJ, Schreiber RD, Louis S, Burnet FM, Thomas L. The immunobiology of cancer immunosurveillance and immunoediting. Immunity. 2004;21:137–48.

36 Swann JB, Smyth MJ. Review series Immune surveillance of tumors. J Clin Invest. 2007;117:1137–46.

37 Vesely MD, Kershaw MH, Schreiber RD, Smyth MJ. Natural innate and adaptive immunity to cancer. Annu Rev Immunol. 2011;29:235–71.

38 Hsu DC, Sereti I. Serious Non-AIDS Events: Therapeutic Targets of Immune Activation and Chronic Inflammation in HIV Infection. Drugs. 2016;76:533–49.

39 Rajasuriar R, Wright E, Lewin SR. Impact of antiretroviral therapy (ART) timing on chronic immune activation/inflammation and end-organ damage. Curr Opin HIV AIDS. 2015;10:35–42.

40 Chaturvedi AK, Madeleine MM, Biggar RJ, Engels EA. Risk of human papillomavirus-associated cancers among persons with AIDS. J Natl Cancer Inst. 2009;101:1120–30.

41 Stein L, Urban MI, O'Connell D, et al. The spectrum of human immunodeficiency virus-associated cancers in a South African black population: Results from a case-control study, 1995-2004. Int J Cancer. 2008;122:2260–5.

42 Sengayi MM, Kielkowski D, Egger M, Dreosti L, Bohlius J. Survival of patients with Kaposi's sarcoma in the South African antiretroviral treatment era: A retrospective cohort study. South African Med J. 2017;107:871.

43 Sengayi M, Babb C, Egger M, Urban MI. HIV testing and burden of HIV infection in

black cancer patients in Johannesburg, South Africa: a cross-sectional study. BMC Cancer. 2015;15:144.

44 Harron K. Introduction to Data Linkage. ADRN Publ. 2016;38.

45 Grulich AE, Li Y, McDonald A, Correll PK, Law MG, Kaldor JM. Rates of non-AIDS-defining cancers in people with HIV infection before and after AIDS diagnosis. Aids. 2002;16:1155–61.

46 Sengayi M, Spoerri A, Egger M, et al. Record linkage to correct under-ascertainment of cancers in HIV cohorts: The Sinikithemba HIV clinic linkage project. Int J cancer. 2016;139:1209–16.

2. SUBMISSABLE MANUSCRIPT

The burden of cancers associated with HIV in the South African Public Health Sector, 2004-2014: a record linkage study

Tafadzwa Dhokotera[1,2], Julia Bohlius[3], Adrian Spoerri[3], Matthias Egger[3,4], Jabulani Ncayiyana[2,5], Victor Olago[1,2], Elvira Singh[1,2], Mazvita Sengayi[1,2]

[1]National Cancer Registry, National Health Laboratory Service, Johannesburg, South Africa

[2]Division of Epidemiology and Biostatistics, School of Public Health, University of the Witwatersrand, Johannesburg, South Africa

[3]Institute of Social and Preventive Medicine (ISPM), University of Bern, Switzerland

[4]Centre for Infectious Disease Epidemiology and Research (CIDER), School of Public Health and Family Medicine, University of Cape Town, South Africa

[5]Division of Epidemiology and Biostatistics, School of Public Health and Family Medicine, University of Cape Town, South Africa

2.1. INTRODUCTION

In Sub-Saharan Africa 25.6 million people currently live with the Human Immunodeficiency Virus (HIV) as of 2016.[1] In South Africa, approximately 14% of the population was living with HIV in 2017.[2] Since the introduction of antiretroviral treatment (ART) in 2004, there has been an increase in longevity amongst people living with HIV (PLHIV).[3] With this increase in longevity and the known association between cancer and HIV, the risk for cancer amongst PLHIV has increased. However, the additional risk of cancer that PLHIV in South Africa have compared to those who are HIV negative in the ART era is not fully documented.

Studies in developed countries have shown a higher burden of non-AIDS Defining Cancers (NADCs) amongst PLHIV in the ART era particularly, anal, skin, liver, and lung cancer.[4,5] Associated with this is age, race, unavailability of ART in some cases, HIV transmission route, lifestyle related factors and immunosuppression.[4,6,7] However, not all NADCs have exhibited differential rates before and after ART, for example breast, colon and prostate cancers, leading to the possibility that not all cancers are associated with immunosuppression.[8] In contrast, developing countries still have a higher burden of AIDS defining cancers (ADCs), namely Kaposi Sarcoma (KS), cervical cancer and non-Hodgkin lymphoma (NHL). This is largely due to co-infections with oncogenic viruses and possibly, poor access to ART.[9,10]

The cancer data provided by the National Cancer Registry (NCR) lacks information on HIV status amongst cancer patients as HIV status is not routinely collected. The South African HIV Cancer Match (SAM) study is a probabilistic record linkage study. It consists of a national HIV cohort created from National Health Laboratory Service (NHLS) HIV laboratory data (CD4 counts, viral load, HIV tests), linked to the NCR data, in order to study cancer risk in HIV positive people.[11] The current study is nested within the SAM study. We aimed to determine

the impact of HIV on cancer burden and the additional risk that PLHIV have compared to HIV negative people or the general South African population.

2.2.RESEARCH IN CONTEXT

2.2.1. Evidence before this study

We searched PubMed for relevant studies published between January 2008 and March 2018 using the MeSH terms, "cancer", "HIV", "risk factor", "attributable risk", and "Africa". To supplement this, we searched Google scholar using the same key terms. We also browsed the references of key articles for additional publications that centred on cancer burden, risk and trends in PLHIV given availability of ART. Key papers from the period before 2008 were also included.

In African studies, most cancers amongst PLHIV in the ART era are ADCs and other infection related cancers. Substantial decreases have been noted for KS after ART introduction although the risk still remains high. However, on the rise is cervical cancer and other HIV related cancers such as conjunctiva and squamous cell carcinoma of the skin. Studies on HIV and cancer done in South Africa have involved HIV cohorts or case control studies which do not allow for generalization to the general population. In addition, data on cancer risk in PLHIV compared to people without HIV is lacking especially in the ART era (from 2004 onwards).

2.2.2. Added value of this study

This study highlights the cancer burden in PLHIV compared to HIV negative individuals at a national level in South Africa. The risk of ADCs, virus-related NADCs (anal, penile vulva, and lip cancers) and other NADCs such as squamous cell carcinoma of the skin (SCC skin) and conjunctival cancer is elevated amongst PLHIV. While the risk of KS remains high, there was a considerable decrease in risk after the introduction of ART in 2004. The greater proportion of females also allows for an extensive analysis of female specific cancers to be done with vulvar cancer being highlighted as high risk for females living with HIV. The risk of conjunctival, cervical, anal, penile, and vulvar cancers continued to rise in the ART era. This

leads to the conclusion that Human Papilloma Virus (HPV) associated cancers (cervix, anus, penis and vulva cancers) continue to cause significant morbidity as all of these cancers other than conjunctival cancer are HPV-related.

2.2.3. Implications of all the available evidence

In the ART era, although the risk is still high, there has been a gradual decline in ADCs amongst PLHIV. This might be as a result of changes in HIV treatment policies in South Africa such as improved ART coverage and initiation of ART at higher CD4 counts. However, the increased risk of anogenital cancers mostly associated with HPV highlights the need for improved awareness and screening for HPV and associated cancers at HIV care centres. Further research and discussion is warranted on inclusive HPV vaccination in PLHIV.

2.3.METHODS

2.3.1. Study setting and design

The NHLS is the largest diagnostic pathology service in South Africa. It provides laboratory and public health services to over 80% of the South African population. This is achieved through a national network of laboratories in all the nine provinces of South Africa. The NHLS' Corporate Data Warehouse (CDW) is an electronic data repository for all public sector laboratory data. The NCR's main mandate is pathology-based cancer surveillance with both private and public laboratories legislated to report all cancer cases to the institution. This was a cross sectional study of all cancers diagnosed in public sector laboratories from 2004-2014 with HIV data being obtained from the NHLS' CDW.

2.3.2. Study population, variables and data sources

We included all records of patients diagnosed with cancer in public healthcare laboratories from 2004 – 2014. Cancer diagnosis was coded according to International Classification of Diseases for Oncology (ICD-O-3) excluding all cancer pre-cursor lesions. Since the source of our HIV data was the NHLS, which services the public sector, we excluded cancer records from the private sector. Our rationale was, if a patient accessed cancer care at a private facility,

they were more likely to access HIV care at a private facility as well.[12] From our linkage out of the 335 589 cancer records that were reported from the private sector only 1122 had a known HIV result.

An individual was considered HIV positive or negative if the HIV diagnostic test result was positive or negative respectively. If the result was indeterminate or neither positive nor negative, the HIV result was regarded as unknown. In addition, HIV monitoring tests such as HIV viral load and CD4 counts (coded CD4/ARV in NHLS dataset), were used to assume an HIV positive status. To supplement the NHLS HIV dataset, repeated text mining was done to extract more HIV results from the clinical section of pathology reports on confirmed cases of cancer reported to the NCR. By definition, text mining refers to the drawing out of important and specific information from a block of text.[13] The text mining process involved the use of key terms used to refer or infer HIV status. The key words used included, "HIV", "HIV+" "HIV positive" "AIDS", "haart", "ART", "ARV", "antiretroviral", anti-retroviral", "RVD", "RVD positive", "retroviral disease", "immune suppression", "immunosuppression", "immuno-suppression", "acquired immune-deficiency", "retroreactive", "immunocompromised", "HIV reactive", "CD4", "regimen 1 treatment", "reg 1 treatment "Retroviral disease", "RVD", "HIV", "HAART" and "ARV". From the extracted records a series of samples were taken and reviewed to refine the search terms. For those with multiple HIV results, we used the last known HIV result. Therefore, if an individual was HIV negative and then became HIV positive later we used the HIV positive result. Demographic characteristics and potential confounders such as age, gender and race were extracted from the NCR database.

2.3.3. Data management

The HIV and cancer datasets were linked using the in house CDW probabilistic record linkage algorithm. This algorithm is used to link all the laboratory records that belong to the same individual within the entire NHLS database. The linkage variables include name, surname and date of birth as the linkage variables. For records to be considered a match, the first letter of the first names should match and two components of the date of birth must also match. First names and surnames are given the same linkage weights (40% each) and the date of birth contributes 20% of the overall weight. For records with a recorded national identity number, exact matching is done and this is used to validate the linkage. Records that attain a score of 90% and above are considered a match. After linkage, duplicates were removed and private cancer records were excluded and a final sample of 329 280 records remained.

Cervical cancer, KS and NHL were classified as ADCs and the rest of the malignancies as NADCs. Although Burkitt's lymphoma is a NHL subtype, it was analysed as an individual cancer due to its relevance in the African context even before HIV.[14] The NADCs were further grouped into virus-related and virus un-related cancers. The following were classified as virus-related cancers according to the IARC Monograph Working group assessment; liver cancer (hepatitis viruses), penis, vulva, vagina, anal, oropharynx, larynx and tonsil (Human Papilloma Virus (HPV) other than cervix), and Hodgkin's lymphoma and nasopharyngeal cancer (Epstein Barr Virus (EBV)).[15] Although all the ADCs are associated with other viruses they were not included in the virus related NADCs category. For descriptive purposes, age was classified into five-year age groups (0-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, and 60+).

2.3.4. Data analysis

Using only the records with a known HIV status we determined the characteristics of cancer patients (age in five-year age groups, gender, race, cancer type (NADC or ADC) and cancer diagnosis year) by HIV status with 95% confidence intervals. To determine the additional risk

that PLHIV had of developing specific cancers as per ICD-O -3 coding, logistic regression models were fitted adjusting for age (as a continuous variable), gender (males and females), race (Asian, Black, Coloured and White), and cancer diagnosis year (modelled as a continuous variable).

We assessed trends in cancer risk for selected cancers by plotting yearly crude odds ratios using Joinpoint regression models (Joinpoint Regression Program, Version 4.6.0.0. April, 2018 Statistical Research and Applications Branch, National Cancer Institute). The program allows one to determine if the trend observed is statistically significant or not. In most cases the independent variable is the calendar year. Observed odds ratios (or other parameters such as incidence rate or counts) are joined in straight lines at each time point hence the term joinpoint. The model goes to identify at which time point a significant change in trend is observed as well as the magnitude of the change (Annual Percentage Change (APC)). Permutation tests are then used to select the final model that better describes the change in trends. To determine the contribution of HIV to the cancer burden in South Africa, we calculated Attributable Risk Fractions (ARFs) using adjusted odds ratios as demonstrated by Newton et al.[16]

2.3.5. Sensitivity analysis

Clinicians are more likely to request an HIV test if the patient is symptomatic, hence creating a selection bias. With high number of missing HIV status, inverse probability weighting (IPW) methods were used as a post-hoc sensitivity analysis to correct for possible selection bias. We created the weights using age, gender, cancer diagnosis year and cancer type similar to the method used by Dryden-Petersen et al.[9]

Analysis was done using Stata version 15 (College Station, TX: StataCorp LP). P-values of less than 0.05 were considered to be statistically significant.

2.3.6. Ethical considerations

Permission to use the routinely collected NHLS and NCR data was sought from the relevant authorities. Ethical approval to conduct the study was granted by the University of the Witwatersrand Human Research Ethics Committee [Ethics certificate numbers (SAM: M160944) and (BCAH: M171083)].

2.4.RESULTS

From 2004 until 2014, a total of 329 208 cancers were reported to the NCR by the public sector laboratories. Probabilistic record linkage identified 90 796 HIV results and through text mining of cancer pathology reports an additional 4483 HIV results were found. Of the 95 279 (28·93%) cancer patients with a known HIV status, 46951 (14·26%) were HIV positive. Amongst PLHIV, cancer proportions were highest between the ages of 25 and 49 (Table 2.1 below). In contrast, 37% (n=17 890) of all cancer patients that were HIV negative were in the over 60 age group. Across all the HIV status subgroups, the greater proportion of cancers was observed in the Black population although the magnitude was highest in PLHIV at 85% (n= 40 004). A general increase in cancer proportions was observed for all cancers irrespective of the HIV status by calendar year. Compared to the HIV negative individuals and those with an unknown status, more ADCs were observed in PLHIV. Throughout the study period, amongst the HIV positive individuals, ADCs remained constantly higher than NADCs (Figure 2.1 below).

Table 2.1: Characteristics and distribution of public sector cancer cases by HIV status, 2004-2014

	HIV POSITIVE		HIV NEGATIVE		HIV UNKNOWN	
Characteristic	n=46951	Proportion (95% CI)	n=48328	Proportion (95% CI)	n=234001	Proportion (95% CI)
Gender						
Female	30487	0.65(0.645-0.654)	30 032	0.62(0.617-0.626)	129051	0.55(0.549-0.554)
Male	16443	0.35(0.346-0.355)	18 287	0.38(0.374-0.383)	104413	0.45(0.444-0.448)
Missing	21	0(0-0.001)	9	0(0-0)	538	0(0.002-0.002)
Age						
0-14	590	0.01(0.012-0.014)	2 755	0.06(0.055-0.059)	3091	0.01(0.013-0.014)
15-19	338	0.01(0.006-0.008)	582	0.01(0.011-0.013)	1331	0.01(0.005-0.006)
20-24	1317	0.03(0.027-0.03)	742	0.02(0.014-0.016)	2407	0.01(0.01-0.011)
25-29	3915	0.08(0.081-0.086)	1 020	0.02(0.02-0.022)	4839	0.02(0.02-0.021)
30-34	6733	0.14(0.14-0.147)	1 295	0.03(0.025-0.028)	7822	0.03(0.033-0.034)
35-39	7773	0.17(0.162-0.169)	2 250	0.05(0.045-0.048)	10468	0.04(0.044-0.046)
40-44	7220	0.15(0.151-0.157)	3 452	0.07(0.069-0.074)	14051	0.06(0.059-0.061)
45-49	5812	0.12(0.121-0.127)	5 081	0.11(0.102-0.108)	17996	0.08(0.076-0.078)
50-54	4671	0.1(0.097-0.102)	6 291	0.13(0.127-0.133)	22903	0.1(0.097-0.099)
55-59	3351	0.07(0.069-0.074)	6 588	0.14(0.133-0.139)	26590	0.11(0.112-0.115)
60+	4738	0.1(0.098-0.104)	17 890	0.37(0.366-0.374)	113501	0.49(0.483-0.487)
Missing	493	0.01(0.01-0.011)	382	0.01(0.007-0.009)	9003	0.04(0.038-0.039)
Race						
Asian	333	0.01(0.006-0.008)	826	0.02(0.016-0.018)	6869	0.03(0.029-0.03)
Black	40004	0.85(0.849-0.855)	25 335	0.52(0.52-0.529)	140946	0.6(0.6-0.604)
Coloured	2620	0.06(0.054-0.058)	10 107	0.21(0.206-0.213)	28793	0.12(0.122-0.124)
White	1548	0.05(0.05-0.054)	1 641	0.22(0.212-0.219)	10115	0.2(0.2-0.204)
Missing	2446	0.03(0.031-0.035)	10 419	0.03(0.032-0.036)	47279	0.04(0.042-0.044)
Province						
Eastern Cape	3314	0.07(0.068-0.073)	4 020	0.08(0.081-0.086)	32379	0.14(0.137-0.14)
Free State	4364	0.09(0.09-0.096)	5 178	0.11(0.104-0.11)	16773	0.07(0.071-0.073)
Gauteng	20957	0.45(0.442-0.451)	15 534	0.32(0.317-0.326)	62004	0.26(0.263-0.267)
Kwazulu-Natal	1434	0.03(0.029-0.032)	712	0.01(0.014-0.016)	35268	0.15(0.149-0.152)
Limpopo	3661	0.08(0.076-0.08)	553	0.01(0.01-0.012)	15529	0.07(0.065-0.067)
Mpumalanga	3277	0.07(0.067-0.072)	593	0.01(0.011-0.013)	9390	0.04(0.039-0.041)
North West	2869	0.06(0.059-0.063)	1 526	0.03(0.03-0.033)	9700	0.04(0.041-0.042)
Northern Cape	1098	0.02(0.022-0.025)	1 396	0.03(0.027-0.03)	6561	0.03(0.027-0.029)
Western Cape	5962	0.13(0.124-0.13)	18 791	0.39(0.384-0.393)	45466	0.19(0.193-0.196)
Missing	15	0(0-0)	25	0(0-0.001)	932	0(0.004-0.004)
Type of cancer						
NADC	17604	0.37(0.371-0.379)	34 965	0.72(0.72-0.727)	172260	0.74(0.734-0.738)
ADC	27125	0.58(0.573-0.582)	10 288	0.21(0.209-0.217)	42992	0.18(0.182-0.185)
Primary site unknown	2222	0.05(0.045-0.049)	3 075	0.06(0.061-0.066)	18750	0.08(0.079-0.081)
Cancer diagnosis year						
2004	1026	0.02(0.021-0.023)	1 767	0.04(0.035-0.038)	23656	0.1(0.1-0.102)
2005	2337	0.05(0.048-0.052)	3 281	0.07(0.066-0.07)	21784	0.09(0.092-0.094)
2006	3038	0.06(0.062-0.067)	3 678	0.08(0.074-0.078)	23101	0.1(0.098-0.1)
2007	3616	0.08(0.075-0.079)	3 882	0.08(0.078-0.083)	22376	0.1(0.094-0.097)

2008	4434	0.09(0.092-0.097)	4 378	0.09(0.088-0.093)	22506	0.1(0.095-0.097)
2009	4782	0.1(0.099-0.105)	4 839	0.1(0.097-0.103)	21917	0.09(0.092-0.095)
2010	5194	0.11(0.108-0.113)	4 926	0.1(0.099-0.105)	21279	0.09(0.09-0.092)
2011	5714	0.12(0.119-0.125)	5 485	0.11(0.111-0.116)	19769	0.08(0.083-0.086)
2012	5931	0.13(0.123-0.129)	5 749	0.12(0.116-0.122)	21592	0.09(0.091-0.093)
2013	5719	0.12(0.119-0.125)	5 401	0.11(0.109-0.115)	18881	0.08(0.08-0.082)
2014	5159	0.11(0.107-0.113)	4 942	0.1(0.1-0.105)	17141	0.07(0.072-0.074)

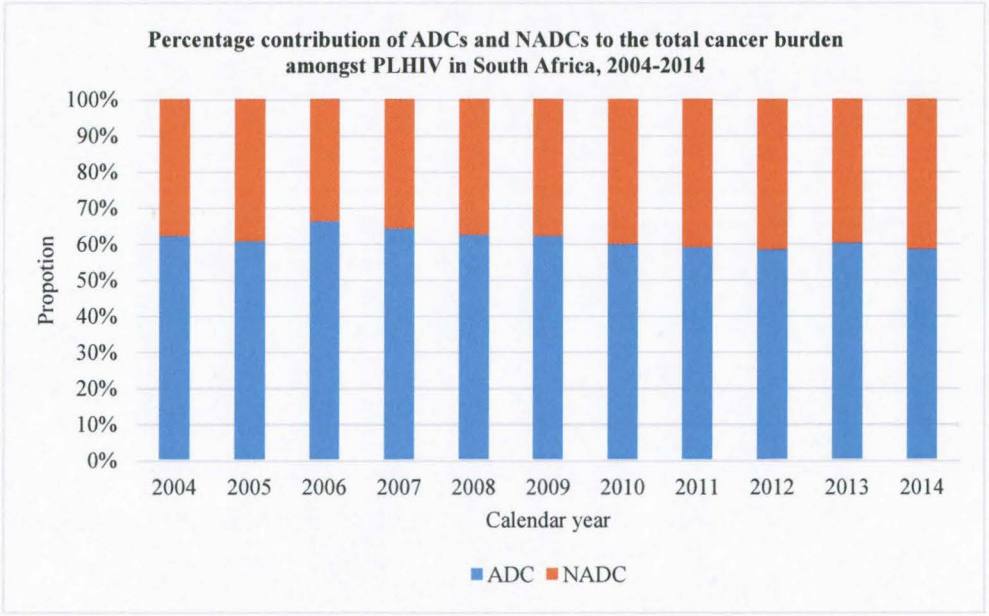


Figure2.1: [1]Percentage contribution of ADCs and NADCs to the total cancer burden amongst PLHIV in South Africa, 2004-2014

Correcting for age, gender, race, and year of cancer diagnosis, cancer risk was highest in the HIV positive population for all ADCs (Kaposi sarcoma, NHL, and cervical cancer) with an overall adjusted odds ratio of 4.5 (95% CI =4.35-4.65). The NHL subtypes Burkitt's lymphoma (adjusted OR: 6·48, 95% CI (5·21-8·07)), Diffuse large B-cell lymphoma (DLBCL) (adjusted OR 2·93 95% CI (2·67-3·22)) and Diffuse immunoblastic large B-cell lymphoma

[1] Figure one is a comparison of incident cancers by cancer type in PLHIV. Given in the graph is a percentage of the total cancers in PLHIV each year.

(DILBCL) (adjusted OR 12·1 95% CI (9·02-16·3)). Compared to HIV negative individuals, PLHIV were 0.74 times less likely to develop NADCs (adjusted OR: 0·26, 95% CI (0·25-0·26). As a group, virus-related NADCs were not significantly associated with HIV but most of the HPV-associated cancers such as anal, penile, vulva and lip and Hodgkin's lymphoma (EBV-associated), were high risk in HIV positive individuals [$p<0.0001$]. Liver cancer which is associated with hepatitis viruses, presented as low risk in PLHIV. With the exception of Squamous Cell Carcinoma (SCC) of the skin, Basal Cell Carcinoma (BCC), eye, and conjunctival cancers which were strongly associated with HIV ($p<0·0001$), non-virus related NADCs were low risk amongst PLHIV. The weighted analysis produced results that were comparable to the complete case analysis (table 2.2).

Table 2.2 Odds ratios for cancer in PLHIV by complete case analysis and by weighted analysis

	COMPLETE CASE		WEIGHTED	
Cancer	**Odds Ratio (95% CI)**	**p-value**	**Odds Ratio (95% CI)**	**p-value**
ADC	4·5 (4·35-4·65)	<0·0001	3·46 (3·33-3·59)	<0·0001
Kaposi	134 (111-161)	<0·0001	98·8 (80·9-120)	<0·0001
Cervix	1·7 (1·63-1·77)	<0·0001	1·68 (1·60-1·76)	<0·0001
NHL	2·73 (2·56-2·91)	<0·0001	2·89 (2·71-3·08)	<0·0001
Burkitt's lymphoma	6·48 (5·21-8·07)	<0·0001	7·83 (6·38-9·62)	<0·0001
Non-Hodgkins NOS	4·26 (3·4-5·34)	<0.0001	4·85 (3·88-6·07)	<0·0001
DLBCL	2·93 (2·67-3·22)	<0.0001	3·27 (3·00-3·59)	<0·0001
DILBCL	12·1 (9·02-16·3)	<0·0001	12·0 (8·91-16·1)	<0·0001
Follicular	0·65 (0·49-0·87)	0·004	0·73 (0·55-0·98)	0·038
NADC	0·26 (0·25-0·26)	<0·0001	0·36 (0·35-0·38)	<0·0001
Virus-related NADC	0·75 (0·71-0·78)	<0·0001	0·77 (0·73-0·82)	<0·0001
Anus	1·63 (1·33-2·00)	<0·0001	1·61 (1·30-1·99)	<0·0001
Hodgkin's lymphoma	1·22 (1·09-1·37)	<0·0001	1·43 (1·27-1·61)	<0·0001
Liver	0·45 (0·39-0·53)	<0·0001	0·51 (0·42-0·61)	<0·0001
Vulva	1·94 (1·67-2·25)	<0·0001	1·82 (1·58-2·10)	<0·0001
Vagina	0·83 (0·67-1·03)	0·09	0·88 (0·70-1·11)	0·29
Penis	2·35 (1·85-2·99)	<0·0001	2·06 (1·63-2·61)	<0·0001
Lip, oral cavity and Pharynx (C00-C14)	0·55 (0·50-0·59)	<0·0001	0·67 (0·61-0·73)	<0·0001
Gum	0·74 (0·41-1·36)	0·34	0·9 (0·46-1·76)	0·76
Lip	2·72 (1·71-4·32)	<0·0001	3·47 (1·95-6·16)	<0·0001
Mouth	0·56 (0·49-0·66)	<0·0001	0·63 (0·53-0·75)	<0·0001
Naso-oropharynx	0·48 (0·42-0·56)	<0·0001	0·57 (0·48-0·68)	<0·0001
Salivary gland	0·65 (0·51-0·82)	<0·0001	0·71 (0·54-0·93)	0·01
Tongue	0·51 (0·43-0·60)	<0·0001	0·57 (0·47-0·70)	<0·0001
Virus-unrelated NADC	0.30 (0.29-0.31)	<0·0001	0·43 (0·42-0·45)	<0·0001
BCC	1·27 (1·08-1·49)	<0·0001	1·39 (1·17-1·67)	<0·0001
Bladder	0·74 (0·61-0·89)	<0·0001	0·86 (0·69-1·08)	0·2
Bone	0·32 (0·26-0·40)	<0·0001	0·41 (0·30-0·54)	<0·0001
Brain	0·27 (0·22-0·34)	<0·0001	0·34 (0·26-0·43)	<0·0001
Colorectal	0·43 (0·38-0·47)	<0·0001	0·55 (0·48-0·62)	<0·0001
Eye	5·9 (5·11-6·82)	<0·0001	7·73 (6·60-9·05)	<0·0001
Conjunctiva	21·5 (16·3-28·4)	<0·0001	20·8 (15·2-28·5)	<0·0001
Haematology	0·49 (0·37-0·65)	<0·0001	0·58 (0·43-0·80)	<0·0001
Kidney	0·18 (0·15-0·21)	<0·0001	0·23 (0·19-0·29)	<0·0001
Larynx	0·56 (0·49-0·64)	<0·0001	0·62 (0·53-0·72)	<0·0001
Leukaemia	0·25 (0·23-0·28)	<0·0001	0·33 (0·30-0·37)	<0·0001
Lung	0·52 (0·48-0·57)	<0·0001	0·62 (0·56-0·68)	<0·0001
Melanoma	0·77 (0·63-0·93)	0·01	0·95 (0·76-1·20)	0·69
Mesothelioma	0·5 (0·34-0·75)	<0·0001	0·51 (0·32-0·80)	<0·0001
Myeloma	0·61 (0·52-0·71)	<0·0001	0·75 (0·63-0·89)	<0·0001
Oesophagus	0·57 (0·52-0·64)	<0·0001	0·58 (0·52-0·66)	<0·0001

Pancreas	0·43 (0·33-0·58)	<0·0001	0·6 (0·42-0·84)	<0·0001
SCC Skin	1·83 (1·64-2·04)	<0·0001	1·88 (1·67-2·11)	<0·0001
Skin	0·95 (0·77-1·16)	0·6	1·11 (0·87-1·41)	0·41
Small intestines	0·33 (0·23-0·48)	<0·0001	0·35 (0·24-0·53)	<0·0001
Stomach	0·43 (0·37-0·49)	<0·0001	0·58 (0·49-0·70)	<0·0001
Thyroid	0·75 (0·59-0·95)	0·02	1·01 (0·74-1·37)	0·97
Uterus	0·38 (0·34-0·43)	<0·0001	0·44 (0·39-0·50)	<0·0001
Breast	0·43 (0·41-0·45)	<0·0001	0·5 (0·47-0·53)	<0·0001
Placenta	0·85 (0·58-1·24)	0·39	1·05 (0·70-1·56)	0·83
Prostate	0·85 (0·76-0·95)	<0·0001	0·95 (0·84-1·08)	0·46
Testis	0·37 (0·26-0·53)	<0·0001	0·48 (0·32-0·74)	<0·0001
Poorly specified histology at any site	0·82 (0·77-0·88)	<0·0001	0·95 (0·88-1·03)	0·2

OR= Odds ratios determined using logistic regression models adjusting for age, gender, race, province and year of cancer diagnosis. NHL=non-Hodgkin's lymphoma. DBCL=Diffuse B-cell lymphoma. DIBCL=Diffuse immunoblastic B-cell lymphoma. NOS= not otherwise specified. BCC= Basal cell carcinoma. ADC= AIDS defining cancer. NADC= non-AIDS defining cancer. SCC skin= squamous cell carcinoma of the skin. Virus-related NADCs= liver cancer (hepatitis viruses), penis, vulva, vagina, anal, lip, mouth, gum, salivary gland and tonsil (Human Papilloma Virus (HPV) associated malignancies other than cervix), and Hodgkin's lymphoma and nasopharyngeal cancer (Epstein Barr Virus (EBV). The weighted analysis included inverse probability weights estimated from known HIV status age, gender, cancer diagnosis and cancer diagnosis year.

Trends in cancer risk for selected individual cancers varied, with significant increases observed for cervix, anus, vulva, conjunctiva and penis from 2004-2014 in PLHIV. Although the APC was not significant for Kaposi sarcoma, there was a substantial decrease in risk between 2004 and 2006 with no changes observed thereafter. Prior to 2011, there was no significant difference in risk for anal, vulva and penile cancers between those who were HIV negative and PLHIV but significant increases were observed after 2011. For Burkitt's lymphoma, and NHL, whilst the risk was higher in PLHIV, there was relatively no change over the study period. Although insignificant, the trend line for Hodgkin's lymphoma was suggestive of an increase in cancer risk.

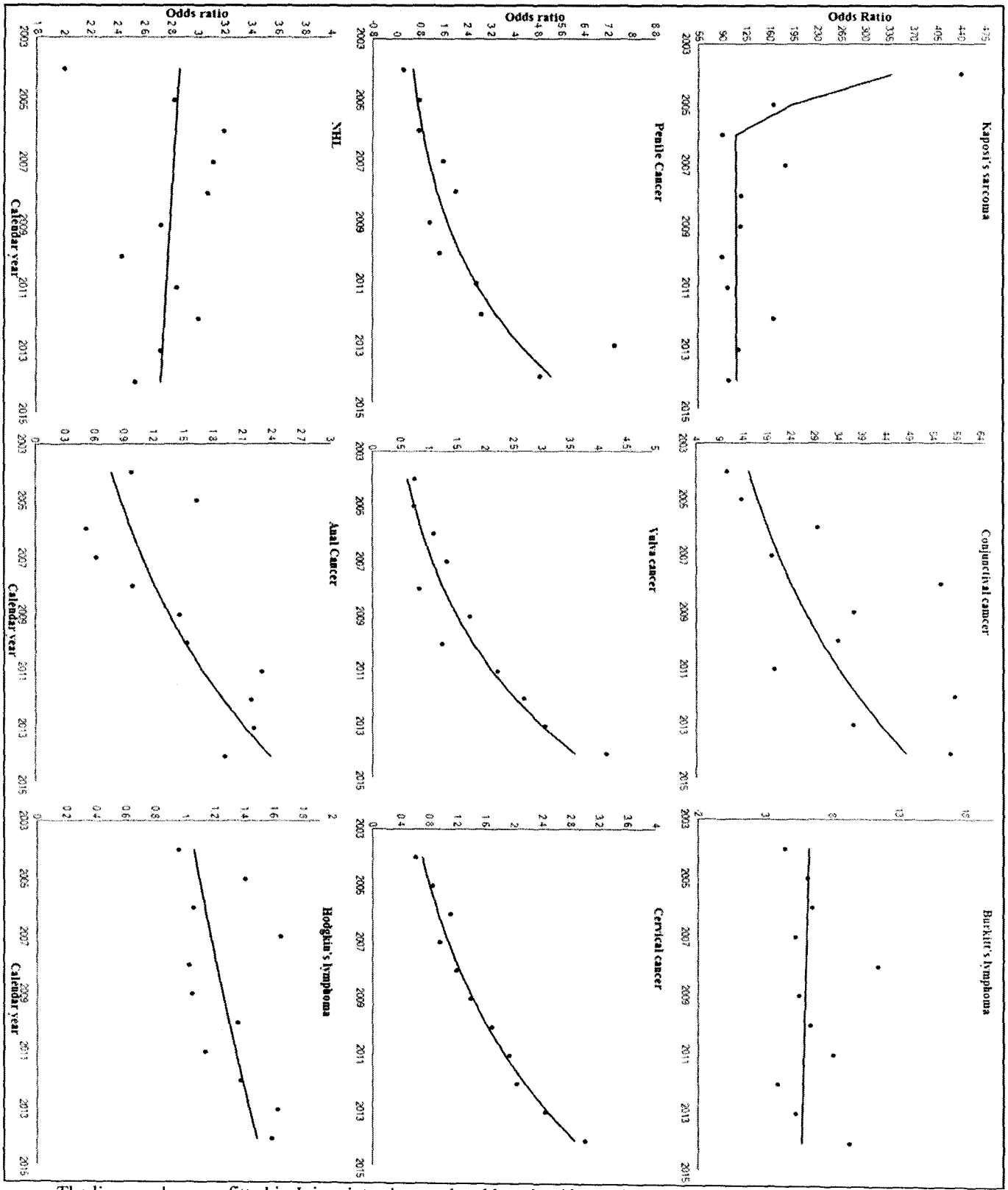


Figure2.2: Trends in cancer risk for selected cancers amongst PLHIV in the South African public health sector, 2004-2014

The line graphs were fitted in Joinpoint using crude odds ratios (dots). The annual percentage change in odds ratios was significant (p-value <0.05) for all cancers selected for in-depth analysis of trends (Figure 3) except for Kaposi sarcoma, Burkitt's lymphoma, NHL and Hodgkin's lymphoma.

There was no shift in burden between ADCs and NADCs observed amongst incident cancers (Figure 1). Using weighted estimates of the odds ratio, 41% of all ADCs reported between 2004 and 2014 were attributable to HIV. The contribution of HIV on ADCs increased by 22% within the study period (Figure 2). No particular contribution by HIV towards NADCs was noted as prior to 2012, given the negative ARFs, HIV did not seem to contribute to the burden of virus related NADCs amongst PLHIV in the public sector. However, the proportion of virus-related ADCs attributable to HIV is beginning to rise.

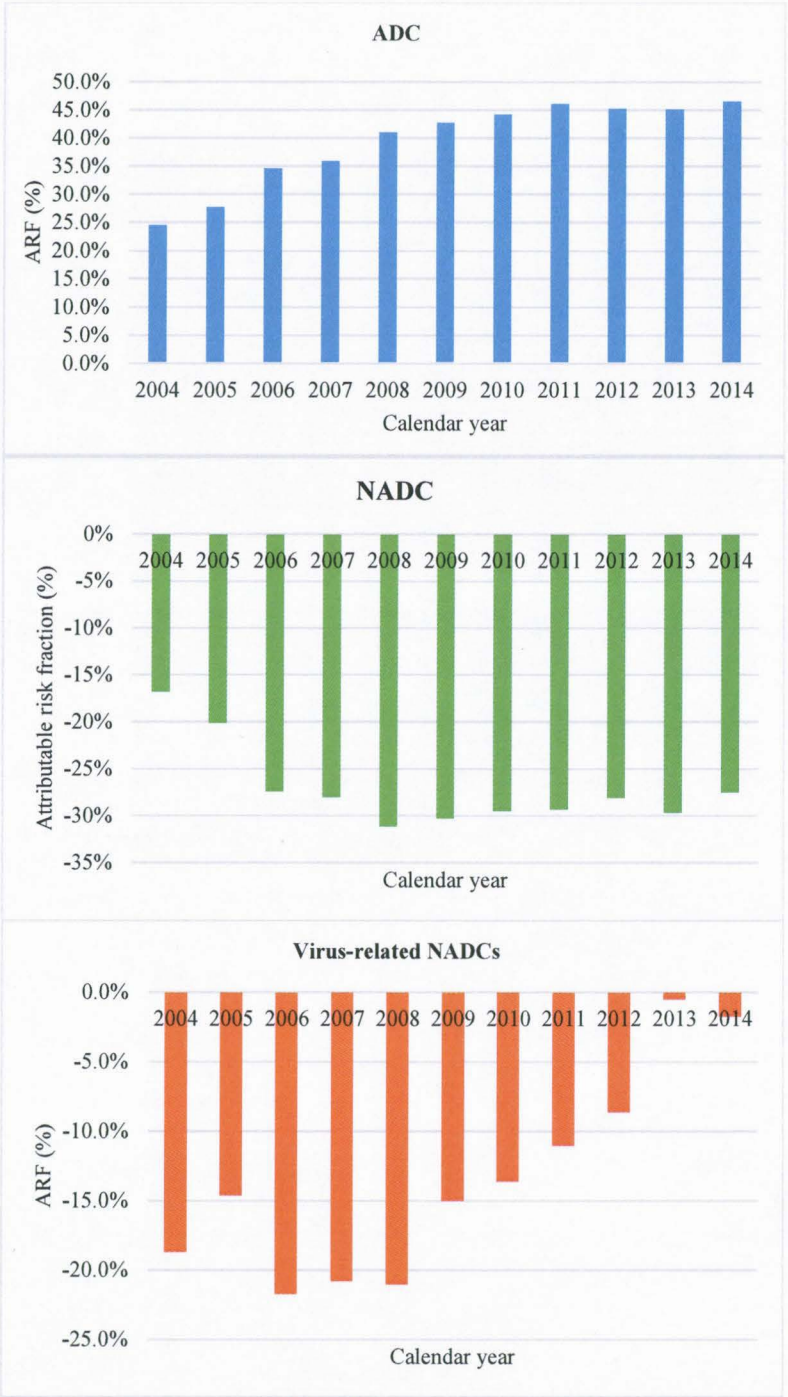


Figure 2.3: Trend in Attributable risk fractions amongst PLHIV in the South African public health sector, 2004-2014 Using adjusted odds ratios adjusting for age, gender, race, year of cancer diagnosis, and Province. ADC= AIDS defining cancer (includes Kaposi sarcoma, non-Hodgkin's lymphoma and cervical cancer). NADC= non-AIDS defining cancer (all cancers that are not ADCs or virus-related). Virus-related NADCs= liver cancer (hepatitis viruses), penis, vulva, vagina, anal, lip, mouth, gum, salivary gland and tonsil (Human Papilloma Virus (HPV) associated malignancies other than cervix), and Hodgkin's lymphoma and nasopharyngeal cancer (Epstein Barr Virus (EBV).

2.5.DISCUSSION

Over the period 2004-2014 (ART era), the risk of all ADCs and some virus-related NADCs was higher amongst HIV positive individuals compared to those who were HIV negative. Highest risk was observed for KS [adjusted OR: 134, 95% CI 111-161], conjunctival cancer [adjusted OR: 21.5, 95% CI 16.3-28.4], DILBCL [adjusted OR 12·1 95% CI 9.02-16.3] and Burkitt's lymphoma [adjusted OR: 6.48, 95% CI 5.21-8.07]. Amongst the virus-related NADCs, HPV-associated cancers such as lip, anal, penile and vulva cancer had the strongest associations with HIV. Squamous cell carcinoma of the skin, BCC and conjunctival cancer were the only virus-unrelated NADCs for which PLHIV were at a higher risk than those who were HIV negative. Over time, amongst PLHIV a significant upward trend in risk was observed for cancer of the conjunctiva, and anogenital cancers other than cervix.

The spectrum of cancers observed in this study was comparable to what has been observed in other African countries. Similar to a case control study by Stein et al conducted before ART was available, the risk of KS, cervical cancer, NHL, anogenital cancers other than cervix and SCC skin was elevated amongst PLHIV in our present study.[17] For KS, the risk was higher compared to the one reported by Stein et al (adjusted OR: 50·4, 95% CI 34·2-74·3) In both our study and the Stein pre-ART study, the relative risk estimates (odds ratios) were adjusted for age, gender, race and year of diagnosis which allowed for comparability. Possible explanations for the higher risk in our study is that, until 2016 when the universal test and treat policy was adopted in South Africa, treatment initiation was dependent on CD4 counts.[18] This led to a high proportion of immunosuppressed individuals with low CD4 counts, a risk factor for ADCs.[18,19] As a result, the risk of KS remained elevated even after ART introduction. Moreover, it is possible that the pick-up rate of KS at HIV clinics improved with the expansion of ART and improvements in HIV treatment policies in South Africa hence the greater strength of association observed. Despite the high risk reported for KS in our study, it was lower than

reported in other studies particularly those done in the developed countries.[4,20] In South Africa, the prevalence of Human Herpes Virus (HHV8) was high even before the HIV era, therefore creating a high KS background risk.[21] In addition to this, clinical diagnosis of KS is quite prevalent in the African context with no biopsies or other samples being sent to the laboratory.[22] Therefore, under-reporting of KS to the pathology-based cancer registry may have been possible.

In contrast, the risk reported in our study for NHL (adjusted OR: 2·73, 95% CI 2·56-2·91) was lower than the one reported by Stein et al. (adjusted OR: 6·1, 95% CI 4·4-8·4) which points to a possible reduction in risk of NHL after the introduction of ART.[17] There was no change noted before and after ART in overall cervical cancer risk although an upward trend was observed in the ART era. This is in line with other reports from Africa with various reasons being put forth to account for the increase in cervical cancer risk even with the introduction of ART. These include advanced disease upon ART initiation and older age. [23,24] Another theory that has been put forward is the lack of a relationship between cervical cancer risk and immunosuppression. Some studies have demonstrated that low CD4 counts do not necessarily amount to increased risk of cervical cancer and other HPV-related cancers.[24] As such, restoration of immunity with ART will not necessarily lead to a reduced risk of cervical cancer. In addition, the prevalence of HPV (a known risk factor for cervical cancer) is higher amongst women living with HIV.[24,25] Possible co-infection with HPV has also been highlighted in this study with increased risk amongst PLHIV observed for HPV associated cancers such as vulva, anus, penis, and lip.

Besides the ADCs and HPV related cancers, we observed other additional cancers as high risk in PLHIV in the ART era. Compared to HIV negative individuals, the risk of conjunctival cancer, Hodgkin's lymphoma and BCC was also higher in PLHIV in our study. Before ART, there were no reports of conjunctival cancer and BCC as being high risk amongst PLHIV in South Africa[17]. Conjunctival cancer has been reported as high risk amongst PLHIV in

Africa.[24,26] High rates of solar radiation and unproved associations with HPV have been cited as possible reasons why this cancer is common in Sub-Saharan Africa compared to Western countries.[24] Like SCC skin, we observed stronger associations between HIV and BCC. Reports have linked age and white race to higher BCC risk in PLHIV with immunosuppression and increased viral loads only being linked to SCC skin.[27,28] On the other hand PLHIV were less likely to develop virus-unrelated cancers such as breast and prostate which is in line with the literature.[4,7,20] Lower risks were also observed for lung and liver cancers in PLHIV which is contrary to other reports especially those done in resource rich areas.[4,7,8] The reason for the decreased risk in our study might have been the methods used to ascertain cancers. We used pathology records to ascertain cancer status. Compared to population-based surveillance, pathology based surveillance underestimate the number of cases occurring in any given setting. Moreover, in the resource rich countries, there is a higher prevalence of lifestyle related factors such as smoking which results in lung cancer and increased alcohol intake which results in liver cancer in HIV cohorts.[7]

In the ART era, different cancer trends have been observed with ADCs decreasing upon ART introduction in other settings.[6,8,20] In particular, KS has declined with the introduction and expansion of ART hence supporting the association between this cancer type and immunosuppression.[6,20,29] In our, study following the initial drop in KS risk after ART introduction in 2004, there has not been a significant change in risk amongst PLHIV in the ART era.[9] This is similar to what was reported in a recent study done in Botswana which demonstrated a decrease in KS risk with ART introduction but no significant change with increased roll out of ART.[9] The arguments for this are similar to the reasons why KS risk was reported as higher in our study compared to the pre-ART era which include HIV treatment policies and improved pick-up rate. The trend in NHL risk exhibited a slight but insignificant decrease over the 11-year period. Whilst some studies have shown decreasing trend in NHL in

the ART era in PLHIV others have shown stable trends even with the increased rollout of ART.[9,20] This has largely been as a result of Burkitt's lymphoma as its incidence has remained constant even in the ART era. Also showing increasing trends in the ART era were most HPV related anogenital cancers (cervix, anus, penis and vulva). Although anal cancer is on the rise, the risk reported is not as high as observed in developed countries. This is possibly due to the difference in HIV epidemiology between South Africa and developed countries. In the latter, the main mode of HIV transmission is men who have sex with men (MSM) through receptive anal sex where as in South Africa, HIV transmission is mainly heterosexual.[4,5] Co-infection with HPV is higher amongst people living with HIV with the routes of transmission being similar to HIV.[30] Both anal and cervical cancer are associated with HPV, but the different transmission routes will result in more cervical cancer in the African context and more anal cancer in developed countries.

This was the first nationwide study to compare cancer risk amongst the HIV positive and HIV negative people in the ART era. Laboratory confirmation of both cancer and HIV allowed for high specificity of HIV and cancer diagnosis. Although a greater proportion of the HIV status was unknown, the methods used to ascertain HIV status such as probabilistic record linkage and text search ensured that we extracted and matched most of the HIV records. The greater percentage of black population with HIV and cancer was reflective of the HIV epidemic in South Africa as well as patterns of access to public health services. In addition to this, the use of IPW allowed for assessment of the risk estimates given the possible selection bias due to the high proportion of missing HIV status. The conclusions from the weighted analysis (IPW) were congruent with the complete case analysis. Moreover, women were well represented with enough numbers for cancers that are common in females to be fully analysed.

Despite all these strengths, our study had limitations. Due to its laboratory-based surveillance system, the NCR underreports some cancers that are diagnosed clinically or radiologically. This might potentially result in misrepresentation of association between HIV and these cancers. Another limitation of our study was overrepresentation of the HIV positive individuals. Doctors are more likely to note down the HIV status of a patient if the patient is tested positive. In addition to this, specific cancers such as KS and other symptoms that are known to be associated with HIV are more likely to prompt a clinician to request an HIV test to be done on the patient.[31] This will result in a higher HIV testing and subsequently higher HIV prevalence compared to the general population. Therefore, with the text mining of doctors clinical notes in pathology-reports, we were more likely to pick up those that were tested positive than those that were tested negative or never tested. However, the percentage of those with a known HIV status in this study (28.9%) was similar to the HIV testing prevalence between 2013 and 2014 (26.1%) [32].Furthermore, data on other potential confounders such as lifestyle patterns (smoking, alcohol intake, diet and exercise) and other opportunistic infections was not available. Access to this information would have possibly made the results more robust. The evaluation of cancer risk in PLHIV as a function of time was not possible in this study. However, through the SAM study determination of cancer risk with a person-time denominator will be possible.

In conclusion, PLHIV have a higher risk for all ADCs and most virus-related NADCs. The risk of anogenital cancers and conjunctival cancer continues to rise in the ART era. Therefore, ART alone is inadequate in reducing cancer in PLHIV. Most of these cancers are HPV-related. Targeted public health interventions for HPV such as screening and expansion of HPV vaccination amongst PLHIV are essential in reducing the burden. To consolidate these efforts, ART expansion and availability should be strengthened. With the introduction of universal

ART treatment in 2016, further decreases in ADCs are expected provided individuals report to health care centres before the HIV disease has advanced.

2.2.REFERENCES

1 WHO. World Health Organization. HIV/AIDS - WHO Fact sheet. 2015. https://www.who.int/news-room/fact-sheets/detail/hiv-aids (accessed July 25, 2017).

2 Osterhaus A, Munster V, Fouchier R. Avian Influenza. Iowa, 2008 DOI:10.1159/000151603.

3 Shisana O, Rhele T, Simbayi LC, *et al.* South African National HIV Prevalence, Incidence and Behaviour Survey, 2012. 2014.

4 Dal Maso L, Polesel J, Serraino D, *et al.* Pattern of cancer risk in persons with AIDS in Italy in the HAART era. *Br J Cancer* 2009; **100**: 840–7.

5 Silverberg MJ, Chao C, Leyden WA, *et al.* HIV infection, immunodeficiency, viral replication, and the risk of cancer. *Cancer Epidemiol Biomarkers Prev* 2011; **20**: 2551–9.

6 Crum-Cianflone N, Hullsiek KH, Marconi V, *et al.* Trends in the incidence of cancers among HIV-infected persons and the impact of antiretroviral therapy: A 20-year cohort study. *AIDS* 2009; **23**: 41–50.

7 Albini L., Calabresi A., Gotti D., *et al.* Burden of non-AIDS-defining and non-virus-related cancers among HIV-infected patients in the combined antiretroviral therapy era. *AIDS Res Hum Retroviruses* 2013; **29**: 1097–104.

8 Bedimo RJ, McGinnis KA, Dunlap M, Rodriguez-Barradas MC, Justice AC. Incidence of non-AIDS-defining malignancies in HIV-infected versus noninfected patients in the HAART Era: Impact of immunosuppression. *J Acquir Immune Defic Syndr* 2009; **52**: 203–8.

9 Dryden-Peterson S, Medhin H, Kebabonye-Pusoentsi M, *et al.* Cancer incidence following expansion of HIV treatment in Botswana. *PLoS One* 2015; **10**: 1–13.

10 Rohner E, Valeri F, Maskew M, *et al.* Incidence Rate of Kaposi Sarcoma in HIV-Infected Patients on Antiretroviral Therapy in Southern Africa. *JAIDS J Acquir Immune Defic Syndr* 2014; **67**: 547–54.

11 Sengayi M, Chen W, Spoerri A, Singh E, Egger M. South African HIV cancer match study: A pilot study towards precision public health. *Top Antivir Med* 2018; **26**: 281s–282s.

12 Blecher M, Kollipara A, DeJager P, Zulu N. Health Financing. In: Padarath A, English R, eds. South African Health Review, 10th edn. Health Systems Trust, 2011. DOI:10.1093/heapol/czq021.

13 Harpaz R, Callahan A, Tamang S, *et al.* Text Mining for Adverse Drug Events: the Promise, Challenges, and State of the Art. *Drug Saf* 2014; **37**: 777–90.

14 Orem J, Mbidde EK, Lambert B, De Sanjose S, Weiderpass E. Burkitt's lymphoma in Africa, a review of the epidemiology and etiology. *Afr Health Sci* 2007; **7**: 166–75.

15 Bouvard V, Baan R, Straif K, *et al.* A review of human carcinogens—Part B: biological agents. *Lancet Oncol* 2009; **10**: 321–2.

16 Newson R. Attributable and unattributable risks and fractions and other scenario comparisons. *Stata J* 2013; **13**: 672–98.

17 Stein L, Urban MI, O'Connell D, *et al.* The spectrum of human immunodeficiency virus-associated cancers in a South African black population: Results from a case-control study, 1995-2004. *Int J Cancer* 2008; **122**: 2260–5.

18 Southern African HIV Clinicians Society. Guidelines for adherence to antiretroviral therapy in adolescents and young adults (expanded version): Recommendations, resources and references. Johannesburg, South Africa, 2017 http://www.sahivsoc.org/Files/Adolescent (LONG)_Guidelines_2017_WEB2.pdf (accessed April 6, 2018).

19 Sengayi MM, Kielkowski D, Egger M, Dreosti L, Bohlius J. Survival of patients with Kaposi's sarcoma in the South African antiretroviral treatment era: A retrospective cohort study. *South African Med J* 2017; **107**: 871.

20 Hernández-Ramírez RU, Shiels MS, Dubrow R, Engels EA. Cancer risk in HIV-infected people in the USA from 1996 to 2012: a population-based, registry-linkage study. *Lancet HIV* 2017; **4**: e495–504.

21 Sitas F, Newton R. Kaposi's sarcoma in South Africa. *J Natl Cancer Inst Monogr* 2001; : 1–4.

22 Amerson E, Woodruff CM, Forrestel A, *et al.* Accuracy of Clinical Suspicion and Pathologic Diagnosis of Kaposi Sarcoma in East Africa. *J Acquir Immune Defic Syndr* 2016; **71**: 295–301.

23 Shiels MS, Pfeiffer RM, Gail MH, *et al.* Cancer burden in the HIV-infected population in the United States. *J Natl Cancer Inst* 2011; **103**: 753–62.

24 Sasco AJ, Jaquet A, Boidin E, *et al.* The challenge of AIDS-related malignancies in sub-Saharan Africa. *PLoS One* 2010; **5**: e8621.

25 Adler DH, Wallace M, Bennie T, *et al.* Cervical dysplasia and high-risk human papillomavirus infections among HIV-infected and HIV-uninfected adolescent females in South Africa. *Infect Dis Obstet Gynecol* 2014; **2014**: 498048.

26 Chokunonga E, Borok MZ, Chirenje ZM, Nyakabau AM, Parkin DM. Trends in the incidence of cancer in the black population of Harare, Zimbabwe 1991-2010. *Int J Cancer* 2013; **133**: 721–9.

27 Crum-Cianflone N, Hullsiek KH, Satter E, *et al.* Cutaneous Malignancies Among HIV-Infected Persons. *Arch Intern Med* 2009; **169**: 1130.

28 Lanoy E, Costagliola D, Engels EA. Skin cancers associated with HIV infection and solid-organ transplantation among elderly adults. *Int J cancer* 2010; **126**: 1724–31.

29 Bohlius J, Valeri F, Maskew M, *et al.* Kaposi's Sarcoma in HIV-infected patients in South Africa: Multicohort study in the antiretroviral therapy era. *Int J cancer* 2014; **135**: 2644–52.

30 Dandapani S V, Eaton M, Thomas CR, Pagnini PG. HIV- positive anal cancer: an

update for the clinician. *J Gastrointest Oncol* 2010; **1**: 34–44.

31 Wong EY, Jordan WC, Malebranche DJ, *et al.* HIV testing practices among black primary care physicians in the United States. *BMC Public Health* 2013; **13**: 96.

32 Tshepo M, Lebogang S, Refilwe S, Lesego M. HIV and AIDS. In: Massyn N, Padarath A, Peer N, Day C, eds. District Health Barometer 2016/17. Health Systems Trust, 2017: 173–84.

3. EXTENDED METHODS AND RESULTS

We estimated age standardized cancer incidence rates (ASIR) by HIV status using the mid-year South African general population as provided by Statistics South Africa as the denominator. South Africa has a high burden of HIV and using the general population estimates, the cancer incidence for the HIV negative individuals are underestimated. In sensitivity analysis, the denominator was corrected for HIV status by subtracting the HIV population. The HIV population was determined from the HIV prevalence as provided by the South African HIV Prevalence survey. The survey is representative of the South African population and was initially conducted in 2002 followed by one in 2005, 2008, and 2012. Extrapolation and interpolation methods were used to estimate the HIV prevalence for the years in between. For those who were HIV positive, the denominator was the general South African HIV population as per the prevalence survey. The coverage of the NHLS's services is approximately 80%. To account for the 20% difference in the NHLS catchment population and the actual general population, 20% of the general population was removed pro-rata. The ASIR were also used to assess the trends of cancers across the 11-year period.

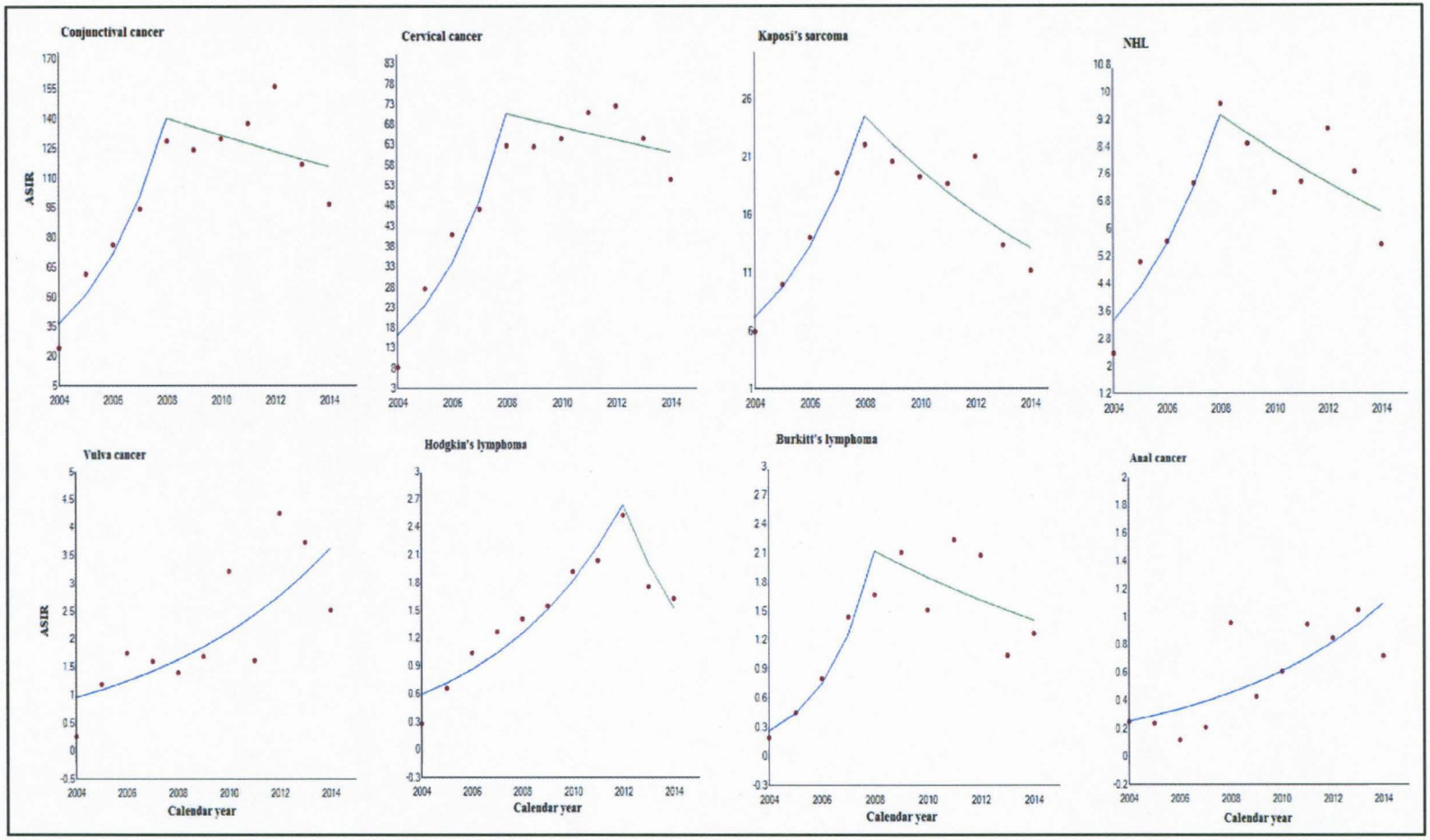


Figure 3.1: Trends in cancer incidence amongst PLHIV in the South African Public Health sector, 2004-2014
ASIR=Age Standardized Incidence rates

Table 3.1: South African mid-year population estimates and the estimated HIV prevalence by gender, 2004-2014

	Females		Males	
Calendar year	Mid-year population estimate	HIV prevalence	Mid-year population estimate	HIV prevalence
2004	24421706	13.7	23832021	11.2
2005	23817900	13.7	23070300	11.2
2006	24063300	14.2	23327600	11.1
2007	24288100	14.7	23562600	10.9
2008	25242200	15.1	23444800	10.8
2009	25451800	15.8	23868700	11.1
2010	25662300	16.5	24329000	11.5
2011	26071721	17.1	24515036	11.9
2012	23385737	17.8	21434040	12.3
2013	27158721	18.5	25823270	12.6
2014	27635944	19.2	26366008	13.0

The mid-year population estimate as per Statistics South Africa (StatsSA) estimate. The HIV prevalence is based on the South African HIV incidence, prevalence and behaviour survey. Since the survey is periodic, extrapolation and interpolation methods were used to determine the prevalence for the years in between

Table3.2: Comparisons of Age Standardized cancer incidence rates by HIV status in the South African public health sector, 2004-2014

Cancer	ASIR/ 100 000 POPULATION	
	HIV NEGATIVE	HIV POSITIVE
Overall Cancer	**78·2 (78·0-78·5)**	**103·4(102·0-104·9)**
ADC	**14·2 (14·1-14·3)**	**49·7 (48·7-50·6)**
Cervical	19·0 (18·8-19·2)	50·6 (49·1-52·2)
Kaposi Sarcoma	2·0 (1·9-2·0)	15·5 (15·0-16·0)
NHL	1·7 (1·7-1·7)	6·7 (6·4-7·0)
Burkitt's lymphoma	0·1 (0·1-0·1)	1·3 (1·2-1·4)
NADC	**64·1(63·8-64·3)**	**53·7 (52·6-54·8)**
Anal	0·3 (0·3-0·3)	0·6 (0·5-0·8)
BCC	3·4 (3·4-3·5)	0·8 (0·7-1·0)
Bladder	1·2 (1·2-1·3)	0·7 (0·5-0·8)
Bone	0·3 (0·3-0·3)	0·3 (0·2-0·3)
Brain	0·4 (0.4-0·4)	0·3 (0·2-0·4)
Breast	18·8 (18·6-19·0)	13·0 (12·3-13·7)
Colorectal	3·4 (3·3-3·5)	1·7 (1·5-1·9)
Connective Tissue	1·1 (1·0-1·1)	1·0 (0·9-1·2)
Endocrine	0·1 (0·1-0·1)	0·1 (0·0-0·1)
Eye	1·0 (0·9-1·0)	2·5 (2·3-2·7)
Gum	0·1 (0·1-0·1)	0·1 (0·0-0·1)
Haematology	0·1 (0·1-0·1)	

Hodgkin Lymphoma	0·4 (0·4-0·4)	1·4 (1·3-1·6)
Kidney	0·6 (0·6-0·6)	0·5 (0·4-0·6)
Larynx	1·3 (1·3-1·3)	1·4 (1·2-1·6)
Leukaemia	1·1 (1·0-1·1)	1·7 (1·6-1·9)
Lip	0·1 (0·1-0·2)	0·2 (0·1-0·3)
Liver	0·8 (0·7-0·8)	0·7 (0·6-0·8)
Lung	4·6 (4·5-4·6)	3·8 (3·5-4·1)
Melanoma	1·0 (1·0-1·1)	0·7 (0·6-0·8)
Mesothelioma	0·2 (0·2-0·2)	0·1 (0·1-0·2)
Mouth	1·3 (1·2-1·3)	1·1 (0·9-1·3)
Myeloma	0.5 (0·5-0·5)	1·1 (0·9-1·2)
Naso-oropharynx	0.9 (0·8-0·9)	0·8 (0·7-1·0)
Oesophagus	4.4 (4·3-4·4)	2·6 (2·3-2·8)
Ovary	0.7 (0·6-0·7)	0·6 (0·5-0·7)
Pancreas	0.3 (0·3-0·4)	0·3 (0·2-0·3)
Penis	0.5 (0·5-0·6)	1·1 (0·9-1·3)
Placenta	0·1 (0·1-0·1)	0·1 (0·1-0·2)
Prostate	15·9 (15·7-16·1)	7·8 (7·1-8·5)
Salivary Gland	0·3 (0·3-0·3)	0.3 (0.3-0·4)
SCC Skin	2·8 (2·8-2·9)	2·3 (2·1-2·5)
Skin	0·6 (0·6-0·6)	0·4 (0·3-0·5)
Small Intestines	0·2 (0·2-0·2)	0·1 (0·1-0·1)
Stomach	2·1 (2·0-2.1)	1·1 (1·0-1·3)
Testis	0·3 (0·2-0·3)	0·2 (0·1-0·3)
Thyroid	0·5 (0·4-0·5)	0·3 (0·3-0·4)
Tongue	1·0 (0·9-1·0)	0·8 (0·7-0·9)
Uterine	3·8 (3·7-3·9)	4·9 (4·3-5·4)
Vagina	0·5 (0·5-0·5)	0·9 (0·7-1·1)
Vulva	0·7 (0·7-0·7)	2·1 (1·8-2·4)

ASIR= Age Standardized Incidence Rate. NHL=non-Hodgkin's lymphoma. BCC= Basal cell carcinoma. ADC= AIDS defining cancer. NADC= non-AIDS defining cancer. SCC skin= squamous cell carcinoma of the skin.

4. OVERALL CONCLUSIONS AND RECOMMENDATIONS

In this study we sought to determine the additional cancer risk that PLHIV have compared to those who are HIV negative in the ART era. Our results demonstrated an elevated incidence of ADCs, HPV-related cancers as well as eye cancer in PLHIV in the ART era. Cancer incidence continued to rise in PLHIV after ART introduction in 2004 and peaked for most cancers between 2008-2009. Post 2008-2009, the incidence of Kaposi sarcoma decreased significantly. Though insignificant, reductions in incidence were also noted for all other cancers (Fig 4) with the exception of Hodgkin's lymphoma, vulva, and anal cancer which continued to rise. Even with this decline in cancer incidence, the likelihood of cancer for PLHIV was still elevated with high risk being observed for ADCs, HPV-related cancers, conjunctival cancer and squamous cell carcinoma of the skin.

The continued rise in cancer incidence even after introduction of ART reflects the effect of inadequate ART coverage. Before the universal test and treat policy was adopted in South Africa, treatment initiation was based on CD4 counts and/or WHO disease staging. In 2004, ART became freely available in the public sector with patients who had CD4 counts of less than 200 cell/ul or in the WHO stage IV of disease being eligible for treatment (1). Patients were also evaluated to determine if they were psychological fit to receive the treatment. In 2010 in addition to the 2004 recommendations, those who had a co-infection with TB we also automatically eligible for the free ART (2). In 2011, the criteria was then expanded to include all patients who had a CD4count of less than 350 cells/ul (2). With the continued and improved coverage and expansion of ART as well as better treatment policies across the years, decreases in cancer incidence were noted in the later years. The high risk of ADCs however, reflects that more still needs to be done in strengthening ART programmes as ADCs are associated with poor immunity. Specifically, patients on treatment should be retained in care and their CD4

counts and viral loads maintained in the acceptable ranges. This would potentially reduce the risk of cancers associated with immunosuppression.

The persistent high risk of HPV associated cancers reflects a possible high HPV prevalence amongst PLHIV, particularly in women. Research has shown that PLHIV have increased risk of persistent infection with various HPV sub-types with ART alone not being sufficient to eliminate the risk(3–5). Given the high HIV burden in South Africa, consolidation of ART programmes with HPV programmes like vaccination and screening should be considered and implemented especially health facilities that provide HIV care. With respect to screening, we recommend that clinicians practice symptom screening not only for ADCs but for all the other cancers that have been identified as high risk in this study such as conjunctival cancer and HPV-associated cancers. Primary interventions should also be considered. In South Africa, the HPV vaccine was introduced in 2014. It is only given to people who have not engaged in sexual intercourse, assuming that those who have, have already been exposed to HPV infection. The HPV vaccine is administered for free to girls between the ages of 9-13 regardless of HIV status in a school based programme. It is also offered at a cost to women less than 25 (catch up vaccination).(6,7) The vaccine protects against the oncogenic serotypes HPV 16 and 18. HPV vaccine has been shown to be safe and able to invoke an immune response amongst women living with HIV in other settings(5). Further research should be conducted on the efficacy of the HPV vaccine in PLHIV. The outcomes of this research would then help to inform policy on the feasibility of administering the HPV vaccine to PLHIV. The target group would be perinatally infected adolescents who might not have had exposure to HPV with the assumption being they are not yet sexually active. In addition to this, PLHIV who test negative for HPV should also be considered for routine vaccination to prevent potential infections. The cost effectiveness and benefits of vaccinating boys should also be considered.

Of all the ADCs and HPV related cancers, cervical cancer still remains the cancer of high public health relevance in PLHIV in South Africa. Despite the availability and increased coverage of ART, cervical cancer risk has significantly increased over the years amongst PLHIV. In addition, South Africa has had a cervical cancer screening programme since 2000. In the first edition of the policy, the recommendation was for women to get three free pap smears done in their lifetime beginning at the age of 30 with 10-year intervals in between (8). There were no special considerations for HIV positive women. However, in 2010 cervical cancer screening was included as part of the routine HIV management (9). The recommendation was for HIV positive women to be screened for cervical cancer at HIV diagnosis and every three years thereafter if the smear was normal. In the 2017 edition of the cervical cancer HIV positive women and women with other immunodeficiencies were included as a high-risk group requiring frequent screening (10). The recommendations were similar to those in the HIV management guideline and in line with the WHO guidelines This is a significant step in efforts to reduce cervical cancer cases amongst women living with HIV. To ensure that this policy is made useful, it is essential to monitor its implementation and evaluate its impact on cervical cancer incidence and risk. As further research, adaptation of the HIV treatment cascade should be done to establish a cervical cancer treatment cascade. This would help to determine the number of HIV infected women that are actually screened, followed up and treated as per the policy guidelines. It would also help in identifying gaps and challenges within the cascade and the health system for additional targeted interventions. The last recommendation is to disseminate the information from this study to clinicians and all health care workers providing services to PLHIV. In addition, it should also be made available to PLHIV. One way to do this is through disease information pamphlets. The pamphlets would include information on high risk cancers in PLHIV, common symptoms and treatment options. Furthermore, the pamphlets targeting health care workers should include information on

screening options of at risk cancers. This would help clinicians in providing better care as they would be informed on the cancers to look out for in their symptom screening as suggested above. It would also help PLHIV in being knowledgeable of cancers that they are at high risk of and ideally to timely present to health care centres.

In conclusion, although decreases in incidence have been noted for some cancers in recent years, cancer risk remains highly elevated in PLHIV. The ADCs are still the major contributors of cancer burden amongst PLHIV with increasing HPV related cancers also noted. Improved ART programmes and inclusive HPV vaccination amongst PLHIV is recommended.

REFERENCES

1. South Africa National Department of Health. South Africa Antiretroviral Treatment Guidelines. 2004. http://apps.who.int/medicinedocs/documents/s17758en/s17758en.pdf . (accessed Aug 30, 2018)

2. Meyer-Rath G, Johnson LF, Pillay Y, Blecher M, Brennan AT, Long L, et al. Changing the South African national antiretroviral therapy guidelines: The role of cost modelling. PLoS One. 2017;12(10):e0186557.

3. Kojic EM, Rana AI, Cu-Uvin S. Human papillomavirus vaccination in HIV-infected women: need for increased coverage. Expert Rev Vaccines. 2016;15:105-17

4. Adler DH, Wallace M, Bennie T, Mrubata M, Abar B, Meiring TL, et al. Cervical dysplasia and high-risk human papillomavirus infections among HIV-infected and HIV-uninfected adolescent females in South Africa. Infect Dis Obstet Gynecol 2014;2014:498048.

5. Wilson EMP, Sereti I. Immune restoration after antiretroviral therapy: the pitfalls of hasty or incomplete repairs. Immunol Rev. 2013;254(1):343–54.

6. Cohen C, Adu C, Redman K, Prabdial-Sing N, Mbenenge N, Hime P, et al. Vaccine Information. 2016. Available from: http://www.amayeza-info.co.za/. (accessed July 29, 2018)

7. Dlamini SK, Madhi SA, Muloiwa R, Von Gottberg A, Moosa M-YS, Meiring ST, et al. Guidelines for the vaccination of HIV-infected adolescents and adults in South Africa. South Afr J HIV Med. 2018;19(1):1–8.

8. National Department of Health. Cervical Cancer Screening Programme [Internet]. 2002. Available from: http://screening.iarc.fr/doc/SAcervical-cancer.pdf. (accesed August 5, 2018)

9. South African National Department of Health. Clinical guidelines for the management of HIV and AIDS in adults and adolescents. 2010. Available from: http://www.sahivsoc.org/Files/Clinical_Guidelines_for_the_Management_of_HIV_AIDS_in_Adults_Adolescents_2010.pdf. (accessed October 5, 2018)

10. National Department of Health. Cervical Cancer Prevention and Control Policy. 2017. Available from: https://www.google.co.za/search?q=The+National+Policy+on+Cervical+Cancer&rlz=1C1KAFB_enZA523ZA524&oq=The+National+Policy+on+Cervical+Cancer&aqs=chrome..69i57.1342j0j8&sourceid=chrome&ie=UTF-8. (accessed August 5, 2018)

APPENDICES

APPENDIX I

Plagiarism declaration for written work: Master of Science in Epidemiology

I TAFADZWA GLADYS DHOKOTERA (Student number: 1662599) am a postgraduate student registered

for degree / Programme MSc: EPIDEMIOLOGY AND BIOSTATISTICS in the Wits School of Public Health.

I am submitting my written work for assessment for the module:

RESEARCH REPORT (COMH7177A) (course name and code).

I hereby declare the following:

- I am aware that plagiarism is the use of someone else's work without their permission and/or without acknowledging the original source).
- I am aware that plagiarism is wrong.
- I confirm that the work submitted for assessment for the above course is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature: Dhokotera Date: 04/02/2019

R14/49 Miss Tafadzwa Gladys Dhokotera

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M171083

NAME: (Principal Investigator)	Miss Tafadzwa Gladys Dhokotera
DEPARTMENT:	Public Health National Cancer Registry, Sandringham, Johannesburg
PROJECT TITLE:	The Burden of Cancers Associated with HIV in the South African Public Health Sector, 2014-2014
DATE CONSIDERED:	Adhoc
DECISION:	Approved unconditionally
CONDITIONS:	Subs-Study under Primary Study (M160944)
SUPERVISOR:	Dr Mazvita Sengayi and Dr Jabulani Ncayiyana
APPROVED BY:	Professor P. Cleaton-Jones, Chairperson, HREC (Medical)
DATE OF APPROVAL:	01/11/2017

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary on the 3rd floor, Phillip Tobias Building, Parktown, University of the Witwatersrand. I/We fully understand the conditions under which I am/we are authorised to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit to the Committee. **I agree to submit a yearly progress report**. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. in this case. the study was initially review October and will therefore be due in the month of October each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature

Date 14/11/2017

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Document Viewer

Turnitin Originality Report

Processed on: 22-Aug-2018 4:10 PM SAST
ID: 992117640
Word Count: 12073
Submitted: 1

1662599:FINAL_DRAFT2.pdf By Tafadzwa Dhokotera

Similarity Index	Similarity by Source	
9%	Internet Sources:	7%
	Publications:	5%
	Student Papers:	N/A

include quoted include bibliography excluding matches < 7 words download print
mode: quickview (classic) report

1% match (Internet from 27-Jun-2018)
http://aorticconference.org

1% match (Internet from 16-Jan-2018)
http://www.crdfglobal.org

<1% match (Internet from 03-Nov-2017)
https://boris.unibe.ch/70964/1/Sengayi%20BMCCancer%202015.pdf

<1% match (Internet from 15-Mar-2015)
http://europepmc.org

<1% match (Internet from 28-Aug-2017)
http://qspace.library.queensu.ca

<1% match (Internet from 30-Mar-2016)
http://acrabstracts.org

<1% match (publications)
Eliane Rohner, Mazvita Sengayi, Bridgette Goeieman, Pamela Michelow, Cynthia Firnhaber, Mhairi Maskew, Julia Bohlius. "Cervical cancer risk and impact of Pap-based screening in HIV-positive women on antiretroviral therapy in Johannesburg, South Africa", International Journal of Cancer, 2017

<1% match (Internet from 14-May-2018)

<1% match (publications)
Calabresi, A, A Ferraresi, A Festa, C Scarcella, F Donato, F Vassallo, RM Limina, F Castelli, and E Quiros-Roldan. "Incidence of AIDS-defining cancers and virus-related and non-virus-related non-AIDS-defining cancers among HIV-infected patients compared with the general population in a large health district of northern Italy, 1999-2009 : Cancers in HIV-infected patients in Brescia, Italy", HIV Medicine, 2013.

<1% match (publications)
"Minister Jeff Radebe - Launch of the Planning, Monitoring and Evaluation Community of Practice for C", Africa News Service, Nov 1 2015 Issue

<1% match (Internet from 27-Mar-2016)
http://www.jembi.org

<1% match (Internet from 27-May-2018)
http://www.nicd.ac.za

<1% match (Internet from 12-Jun-2018)
http://heaids.org.za

Information for Authors

The Lancet HIV is an exclusively online journal dedicated to publishing original research that advocates change in, or illuminates, HIV clinical practice. We will publish translational, epidemiological, clinical, operational, and implementation studies. All original research judged eligible for consideration by the journal's editors will be peer-reviewed within 72 h and, if accepted, published online within 8 weeks from submission. The journal will also publish relevant commentary and correspondence. Wherever possible, figures and good quality photographs (colour or black and white) should be used to supplement and to enhance the text. We also welcome videos. Further details on the different sections of *The Lancet HIV*, and how to submit to the journal, are provided below. If you require further clarification, the journal's editorial staff will be pleased to help (email TheLancetHIV@lancet.com).

Manuscripts must be solely the work of the author(s) stated, must not have been previously published elsewhere, and must not be under consideration by another journal. *The Lancet* journals are signatories of the Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals, issued by the International Committee of Medical Journal Editors (ICMJE Recommendations) and to the Committee on Publication Ethics (COPE) code of conduct for editors. We follow COPE's guidelines.

Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals http://www.icmje.org

COPE Code of Conduct http://publicationethics.org/files/u2/New_Code.pdf

How to submit your paper

Manuscript submission

Manuscript submission to all *Lancet* journals is free. Manuscripts should be submitted online via the *The Lancet HIV*'s online submission and peer review website (known as EES) at http://ees.elsevier.com/thelancethiv.

- Simply log on to EES and follow the on-screen instructions for all submissions.
- If you have not used EES before, you will need to register first. In EES, the corresponding author is the person who enters the manuscript details and uploads the submission files.
- Inclusion of illustrations (photographs, graphs, diagrams, etc) is a prerequisite for publication. Submission of original and editable artwork files is encouraged. Digital photography files should have a resolution of at least 300 dpi and be at least 107 mm wide. . Before and after images should be taken with the same intensity, direction, and colour of light.
- In almost all cases, if you have a finished manuscript, you should submit it, rather than contacting *The Lancet HIV* to enquire whether an unseen manuscript is likely to be accepted. Unless you have been asked by the Editor to submit by email, you should use the online system for all types of submission, including Correspondence.
- If you have any technical problems or questions, please contact our dedicated customer support:

For the Americas: +1 888 8347287 (09:00 to 17:00 central standard time)
For Asia and Pacific: +81 3 55615032 (09:30 to 17:30 Japan standard time)
For Europe and rest of the world: +44 1865 843577 (08:30 to 17:00 GMT)
For Chinese-speaking customers: +86 10 85208780 (9:00 to 17:30 China standard time)
For Spanish-speaking customers: +34 932 406176 (09:00 to 17:00 GMT)
For French-speaking customers: +33 171 165608 (09:00 to 17:00 GMT)
Email: TheLancetHIV@lancet.com

Covering letter

- You should upload your covering letter at the "Enter Comments" stage of the online submission process

First submissions to *The Lancet HIV* should include:

1. Covering letter
2. Manuscript including tables and panels
3. Figures
4. Author statement form (see next section)
5. Declaration of interests and source of funding statements (see next section)
6. In-press papers—one copy of each with acceptance letters
7. Protocols and CONSORT details for randomised controlled trials (see Articles)
8. We encourage disclosure of correspondence from other journals and reviewers, if previously submitted, and we might contact relevant editors of such journals
9. Research in context panel, for all primary research Articles

- Use the covering letter to explain why your paper should be published in *The Lancet HIV* rather than elsewhere

Statements, permissions, and signatures

Authors and contributors

- Designated authors should meet all four criteria for authorship in the ICMJE Recommendations
- All authors, and all contributors (including medical writers and editors), should specify their individual contributions at the end of the text.
- *The Lancet HIV* will not publish any paper unless we have the signatures of all authors.
- We suggest you use the author statement form and upload the signed copy with your submission
- Please include written consent of any cited individual(s) noted in acknowledgments or personal communications.

ICMJE Recommendations http://www.icmje.org

Author statement form www.thelancet.com/for-authors/forms#author-sigs

Forms and signatures

For Comments and Correspondence, we require you to upload your forms at submission. For original research (Articles), we will request these forms after peer review. The following signed statements are required:

- Authors' contributions
- Conflicts of interest statements (ICJME forms)
- Statements of role, if any, of medical writer or editor
- Acknowledgments—written consent of cited individual
- Personal communications—written consent of cited individual

- Use of copyright-protected material—signed permission statements from author and publisher

These statements can be scanned and submitted electronically with your submission. Please note that *The Lancet* journals will accept hand-signed and electronic (typewritten) signatures.

Declaration of interests

A conflict of interest exists when professional judgement concerning a primary interest (such as patients' welfare or validity of research) may be influenced by a secondary interest (such as financial gain). Financial relationships are easily identifiable, but conflicts can also occur because of personal relationships or rivalries, academic competition, or intellectual beliefs. A conflict can be actual or potential, and full disclosure to the Editor is the safest course. Failure to disclose conflicts might lead to publication of a correction or even to retraction. All submissions to *The Lancet HIV* must include disclosure of all relationships that could be viewed as presenting a potential or actual conflict of interest (see *Lancet* 2001; **358**: 854–56 and *Lancet* 2003; **361**: 8–9). The Editor may use such information as a basis for editorial decisions, and will publish such disclosures. Agreements between authors and study sponsors that interfere with authors' access to all of a study's data, or that interfere with their ability to analyse and interpret the data and to prepare and publish manuscripts independently, may represent conflicts of interest, and should be avoided.

- At the end of the text, under a subheading "Declaration of interests", all authors must disclose any financial and personal relationships with other people or organisations that could inappropriately influence (bias) their work. Examples of financial conflicts include employment, consultancies, stock ownership, honoraria, paid expert testimony, patents or patent applications, and travel grants, all within 3 years of beginning the work submitted. If there are no conflicts of interest, authors should state that none exist.
- All authors are required to provide a Conflict of Interest Statement and should complete a standard form, which is available at http://www.thelancet.com/for-authors/forms#icmje-coi. The form has been modified by the ICMJE following consultation with authors and editors. Further information is available in a joint ICMJE statement published on July 1, 2010. For more information see *Lancet* 2009; **374**: 1395–96.
- For Comments, *The Lancet HIV* will not publish if an author, within the past 3 years, and with a relevant company or competitor, has any stocks or shares, equity, a contract of employment, or a named position on a company board; or has been asked by any organisation other than *The Lancet HIV* to write, be named on, or to submit the paper (see *Lancet* 2004; **363**: 2–3).

ICMJE COI form
http://www.thelancet.com/for-authors/forms#icmje-coi
Joint ICMJE statement
http://www.thelancet.com/for-authors/forms#icmje-statement

Patient Consent form
http://www.thelancet.com/pb/assets/raw/Lancet/authors/lancet-consent-form.pdf

Role of the funding source

- All sources of funding should be declared as an acknowledgment at the end of the text.
- At the end of the Methods section, under a subheading "Role of the funding source", authors must describe the role of the study sponsor(s), if any, in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.
- If there is no Methods section, the role of the funding source should be stated as an acknowledgment. If the funding source had no such involvement, the authors should state this.
- The corresponding author should confirm that he or she had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Role of medical writer or editor

- If a medical writer or editor was involved in the creation of your manuscript, we need a signed statement from the corresponding author to include their name and information about funding of this person.
- This information should be added to the Acknowledgments or Contributors section.
- We require signed statements from any medical writers or editors declaring that they have given permission to be named as an author, as a contributor, or in the Acknowledgments section.

Patient and other consents

- Appropriate written consents, permissions, and releases must be obtained where you wish to include any case details, personal information, or images of patients or other individuals in *The Lancet* journals in order to comply with all applicable laws and regulations concerning privacy and/or security of personal information. Studies on patients or volunteers need approval from an ethics committee and informed consent from participants. These should be documented in your paper.
- Do not use "blackout" bars or similar devices to anonymise patients in clinical images: if you have taken consent appropriately masking is not needed.
- Since the consent form needs to comply with the relevant legal requirements of your particular jurisdiction, we do not provide sample forms; this is your responsibility. Your affiliated institution should be able to provide an appropriate form.
- For the purposes of publishing in *The Lancet* journals, a consent, permission, or release should include, without limitation, publication in all formats (including print, electronic, and websites), in sublicensed and reprinted versions (including translations), and in other works and products.
- To respect your patients' and any other individuals' privacy, please do not send signed forms to *The Lancet HIV*. Please instead complete the patient consent section of the Author statements while retaining copies of the signed forms in the event they should be needed.
- If consent, permission, or release is made subject to any conditions, *The Lancet HIV* must be made aware in writing of all such conditions before publication.
- For more information about our policy, please visit https://www.elsevier.com/about/our-business/policies/patient-consent.

Types of article and manuscript requirements

Please ensure that anything you submit to *The Lancet HIV* follows the guidelines provided for each article type. For instruction on how to format the text of your paper, including tables, figures, panels, and references, please see our **Formatting guidelines**.

Red section (Articles)

Articles

- *The Lancet HIV* prioritises reports of original research that are likely to change clinical practice or thinking
- We invite submission of all clinical trials, whether phase 1, 2, 3, or 4. For phase 1 trials, we consider those of a novel substance for a novel indication, if there is a strong or unexpected beneficial or adverse response, or a novel mechanism of action.
- We require the registration of all interventional trials, whether early or late phase, in a primary register that participates in WHO's International Clinical Trial Registry Platform (see *Lancet* 2007; **369**: 1909–11) or in ClinicalTrials.gov, in accord with ICMJE recommendations. We also encourage full public disclosure of the minimum 20-item trial registration dataset at the time of registration and before recruitment of the first participant (see *Lancet* 2006; **367**: 1631–35). The registry must be independent of for-profit interest.
- Reports of trials must conform to CONSORT 2010 guidelines and should be submitted with their protocols.
- All reports of randomised trials should include sections entitled Randomisation and masking and Outcomes, within the Methods section. Please refer to *The Lancet*'s formatting guidelines for randomised trials.
- Cluster-randomised trials must be reported according to CONSORT extended guidelines.
- Randomised trials that report harms must be described according to extended CONSORT guidelines.
- Studies of diagnostic accuracy must be reported according to STARD guidelines.
- Observational studies (cohort, case-control, or cross-sectional designs) must be reported according to the STROBE statement, and should be submitted with their protocols.
- We encourage the registration of all observational studies on a WHO-compliant registry (see *Lancet* 2010; **375**: 348).
- Genetic association studies must be reported according to STREGA guidelines.
- Systematic reviews and meta-analyses must be reported according to PRISMA guidelines. Please refer to *The Lancet*'s formatting guidelines for systematic reviews and meta-analyses.
- Reports of studies of global health estimates should be reported according to the GATHER statement (see *Lancet* 2016; published online June 28. http://dx.doi.org/10.1016/S0140-6736(16)30388-9)
- To find reporting guidelines see: http://www.equator-network.org.
- We encourage authors to share any additional data, preferably translated into English, that would facilitate the replication or further analysis of their work—eg, the raw numbers underlying their analysis or the code for any modelling. Mendeley Data is a secure online repository for research data, permitting archiving of any file type and assigning a permanent and unique digital object identifier (DOI) so that the files can be easily referenced. If authors wish to share their supporting data, and have not already made alternative arrangements, a Mendeley DOI can be referred to in a section entitled "Data sharing" at the end of the Methods section, ahead of "Role of the funding source". If authors have already deposited their data in another repository, or have made other arrangements for data to be shared (eg, by means of an adjudication process or contacting the authors), they should use this section to elaborate.

All Articles should, as relevant

- Be up to 3500 words (4500 for randomised controlled trials) with 30 references (the word count is for the manuscript text only).
- Include an abstract (semistructured summary), with five paragraphs (Background, Methods, Findings, Interpretation, and Funding), not exceeding 250 words. Our electronic submission system will ask you to copy and paste this section at the "Submit Abstract" stage.
- For randomised trials, the abstract should adhere to CONSORT extensions: abstracts (see *Lancet* 2008; **371**: 281–83).
- When reporting Kaplan-Meier survival data, at each timepoint, authors must include numbers at risk, and are encouraged to include the number of censored patients.
- For intervention studies, the abstract should include the primary outcome expressed as the difference between groups with a confidence interval on that difference (absolute differences are more useful than relative ones). Secondary outcomes can be included as long as they are clearly marked as secondary and all such outcomes are reported.
- Use the SI system of units and the recommended international non-proprietary name (rINN) for drug names. Ensure that the dose, route, and frequency of administration of any drug you mention are correct.
- Use gene names approved by the Human Gene Organisation. Novel gene sequences should be deposited in a public database (GenBank, EMBL, or DDBJ), and the accession number provided. Authors of microarray papers should include in their submission the information recommended by the MIAME guidelines. Authors should also submit their experimental details to one of the publicly available databases: ArrayExpress or GEO.
- Include any necessary additional data as part of your EES submission.
- All accepted Articles should include a link to the full study protocol published on the authors' institutional website (see *Lancet* 2009; **373**: 992 and *Lancet* 2010; **375**: 348).
- We encourage researchers to enrol women and ethnic groups into clinical trials of all phases, and to plan to analyse data by sex and by race
- For all study types, we encourage correct use of the terms sex (when reporting biological factors) and gender (when reporting identity, psychosocial, or cultural factors). Where possible, report the sex and/or gender of study participants, and describe the methods used to determine sex and gender. Separate reporting of data by demographic variables, such as age and sex, facilitates pooling of data for subgroups across studies and should be routine, unless inappropriate. Discuss the influence or association of variables, such as sex and/or gender, on your findings, where appropriate, and the limitations of the data.

Putting research into context

- All research papers (including systematic reviews/meta-analyses) submitted to any journal in *The Lancet* family must include a panel

WHO's International Clinical Trial Registry Platform
http://www.who.int/ictrp/network/trds/en/index.html

Clinical trials
http://clinicaltrials.gov

ICMJE recommendations
http://icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html

CONSORT 2010 guidelines
http://www.consort-statement.org/consort-2010

CONSORT extended guidelines
http://www.consort-statement.org/extensions/extensions/

STARD guidelines
http://www.stard-statement.org/

STROBE statement
http://www.strobe-statement.org/

STREGA guidelines
http://www.equator-network.org/reporting-guidelines/strobe-strega/

PRISMA guidelines
http://www.prisma-statement.org/

Formatting guidelines
http://www.thelancet.com/for-authors/forms#meta-analysis

GATHER statement
http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(16)30388-9/fulltext

To find reporting guidelines, see
http://www.equator-network.org

MENDELEY data
https://data.mendeley.com

Human Gene Organisation
http://www.genenames.org/

MIAME guidelines
http://fged.org/projects/miame/

Array and GEO
http://www.ebi.ac.uk/microarray-as/ae/
http://www.ncbi.nlm.nih.gov/geo

putting their research into context with previous work in the format outlined below (see *Lancet* 2014; **384**: 2176–77, for the original rationale). This panel should not contain references. Editors will use this information at the first assessment stage and peer reviewers will be specifically asked to check the content and accuracy.

- The Discussion section should contain a full description and discussion of the context. Authors are also invited to either report their own, up-to-date systematic review or cite a recent systematic review of other trials, putting their trial into context of the review.

For The Lancet journals' policy on corrections of errors see **www.thelancet.com/for-authors/forms#correction**

Research in context

Evidence before this study

This section should include a description of all the evidence that the authors considered before undertaking this study. Authors should briefly state: the sources (databases, journal or book reference lists, etc) searched; the criteria used to include or exclude studies (including the exact start and end dates of the search), which should not be limited to English language publications; the search terms used; the quality (risk of bias) of that evidence; and the pooled estimate derived from meta-analysis of the evidence, if appropriate.

Added value of this study

Authors should describe here how their findings add value to the existing evidence.

Implications of all the available evidence

Authors should state the implications for practice or policy and future research of their study combined with existing evidence.

Research in context panels should not contain references; key studies mentioned here should be referenced in the main text.

Formatting guidelines for revised manuscripts

Guidelines on format for text and figures can be found at **http://www.thelancet.com/for-authors/forms#artwork**

Formatting guidelines for randomised trials

http://www.thelancet.com/for-authors/forms#rct

Blue section (Comment, Correspondence)

Editorial

- Editorials are the voice of *The Lancet HIV*, and are written in-house by the journal's editorial-writing team and signed "*The Lancet HIV*".

Comment

- This section contains Comments that accompany papers published in *The Lancet HIV* or on issues of wide-reaching concern in HIV. Most are commissioned, but unsolicited Comments (no more than 750 words, 10 references, and one figure, panel, or small table) are also welcome. Comments may be peer reviewed.
- The place to respond to something we have published is in our **Correspondence** section.
- See **Conflicts of Interest** guidelines for comments.

Correspondence

- Letters can be written in response to previous content published in *The Lancet HIV*.
- Letters for publication must reach us within 4 weeks of publication of the original item and should be no longer than 400 words.
- Letters of general interest, unlinked to items published in the journal, can be up to 400 words long.
- Correspondence is not usually peer reviewed, but we might invite replies from the authors of the original publication, or pass on letters to these authors.
- Only one table or figure is permitted, and there should be no more than five references and five authors.
- All accepted letters are edited. Proofs will be sent to authors before publication.

Corrections

- Any substantial error in any article published in *The Lancet HIV* should be corrected as soon as possible. Blame is not apportioned; the important thing is to set the record straight.
- The *Lancet* journals have a policy for types of errors that we do and do not correct. We will always correct any error affecting a non-proprietory drug name, dose, or unit, any numerical error in the results, or any factual error in the interpretation of results.

Formatting guidelines

Language

- Manuscripts should be submitted in English. Authors writing in Chinese, Portuguese, or Spanish may wish to use the Webshop (http://webshop.elsevier.com/languageservices) to provide an English translation of their manuscript for submission.

Title page

- A brief title, author name(s), preferred degree (one only), affiliation(s), and full address(es) of the authors must be included. The name and address of the corresponding author should be separately and clearly indicated with email and telephone details.

Formatting of text

- Type a single space at the end of each sentence
- Do not use bold face for emphasis within text
- Use a comma before the final "and" or "or" in a list of items
- Type decimal points midline (ie, 23·4, not 23.4). To create a midline decimal on a PC hold down ALT key and type 0183 on the number pad, on a Mac ALT shift 9
- Numbers one to ten are written out in words unless they are used as a unit of measurement, except in figures and tables.
- Use single hard-returns to separate paragraphs. Do not use tabs or indents to start a paragraph.
- Do not use the automated features of your software, such as hyphenation, endnotes, headers, or footers (especially for references). Please use page numbering.

References

- Cite references in the text sequentially in the Vancouver numbering style, as a superscripted number after any punctuation mark. For example "...as reported by Saito and colleagues.[15]"
- Two references are cited separated by a comma, with no space. Three or more consecutive references are given as a range with an en rule. To create an en rule on a PC hold down CTRL key and minus sign on the number pad, on a Mac ALT hyphen.

- References in tables, figures, and panels should be in numerical order according to where the item is cited in the text.
- Here is an example for a journal reference (note the use of tab, bold, italic, and the en rule):
  "15[tab]Saito N, Ebara S, Ohotsuka K, Kumeta J, Takaoka K. Natural history of scoliosis in spastic cerebral palsy. *Lancet* 1998; **351:** 1687–[en rule]92."
- Give any subpart to the title of the article.
- If there are six authors or fewer, give all six in the form:
  surname space initials comma
- If there are seven or more give the first three in the same way, followed by et al.
- For a book, give any editors and the publisher, the city of publication, and year of publication.
- For a chapter or section of a book, also give the authors and title of the section, and the page numbers.
- For online material, please cite the URL, together with the date you accessed the website.
- Online journal articles can be cited using the DOI number.
- Do not put references in the Summary.

Data

- SI units are required.
- Numbers in text and tables should always be provided if % is shown.
- Means should be accompanied by SDs, and medians by IQR.
- Exact p values should be provided, unless p<0·0001.

Guidelines for supplementary material

All material should be submitted as one PDF (with numbered pages) with the paper and will be peer reviewed. Material will be published at the discretion of *The Lancet* journals' editors. All material should be provided in English.

Text

- Main heading for the web extra material should be in 12 point Times New Roman font **BOLD.**
- Text should be in 10 point Times New Roman font, single spaced.
- Headings should be in 10 point **BOLD.**

Tables

- Main table heading should be in 10 point Times New Roman font **BOLD.**
- Legends should be in 10 point, single spaced.
- Tables should be in 8 point Times New Roman font, single spaced.
- Headings within tables should be in 8 point **BOLD.**

Drug names

- Recommended international non-proprietary name (rINN) is required
- We encourage use of neuroscience-based nomenclature for psychotropic drugs

References

- Vancouver style as above.
- Numbered in order of mention in appendix and numbered separately from references in the full paper

Figures

- All images must have a minimum resolution of 300 dpi, width 107 mm.
- Main figure heading should be in 10 point Times New Roman font **BOLD.**
- Legends should be in 10 point, single spaced.

Audio/video material

- The paper to which the audio or video clip relates should be mentioned in the recording.
- Audio clip and video files should be accompanied with brief text explaining the content of the audio, names of interviewers and interviewees, date of recording, and place of recording if relevant.
- Written consent from all parties must be obtained (see also the above section on **Patient and other consents**).

Audio

- Audio material submitted as an mp3 file, no larger than 50 Mb
- Your paper may be selected for a podcast. If so, the Web Editor will contact you to arrange a pre-recorded interview to discuss your paper. For more information, see Audio.

Audio
http://www.thelancet.com/audio

Video

- Video material should be submitted in .mp4 format with aspect ratio of 16:9, and be no larger than 50 Mb
- We welcome your videos and invite you to submit any video material (reports, interviews, scans, imaging) for consideration in the online journal. Please ensure that all those featured in the video have given permission for publication (see also the previous section on **Patient and other consents**)
- All video files can be submitted alongside your article in EES

Disclosure of results before publication

- Presentation of data at a scientific meeting, as a poster, abstract, orally, on a CD, or as an abstract on the web, or on a pre-print server does not conflict with submission to *The Lancet HIV*.
- As a member journal of the International Committee for Medical Journal Editors, *The Lancet HIV* does not regard results that are posted in the same clinical trials registry in which primary registration resides as prior publication, if the results are presented in the form of a brief structured abstract or table (<500 words; see *Lancet* 2007; **369:** 1909–11). However, presentation of results in other circumstances (eg, investors' meetings) is discouraged and could jeopardise consideration of the manuscript.

Online publication

- *The Lancet HIV* publishes papers online as they become ready. You will be informed at least a week in advance of the Online publication date

Drug names
For more on neuroscience-based nomenclature see http://www.thelancet.com/pdfs/journals/lanpsy/PIIS2215-0366(17)30098-6.pdf

How *The Lancet HIV* handles your paper

Acknowledgment

- Receipt of your paper will be acknowledged by an email

containing a reference number, which should be used in all future communications.

Peer review

- Every Article published in *The Lancet HIV* has been peer reviewed. Occasional contributions (eg, Comments) are accepted without peer review.
- On submission to *The Lancet HIV*, your report will first be read by one or more of the journal's staff of physicians and scientists. This is an important feature of our selection process and many papers will be turned away on the basis of in-house assessment alone. That decision will be communicated quickly.
- Research papers selected for peer review are sent to at least three reviewers. You will receive notification of which editor is handling the peer review of your paper.

Decision

- Submissions that survive in-house assessment and peer review might be referred back to authors for revision. This is an invitation to present the best possible paper for further scrutiny by the journal; it is not an acceptance.
- Authors should give priority to such revisions; the journal will reciprocate by making a final decision quickly.
- Two copies of the revised version should be sent back, one of which should be highlighted to show where changes have been made. Detailed responses to reviewers' comments, in a covering letter, are also necessary.

The Lancet journals and other Elsevier journals

- If your paper is rejected by *The Lancet HIV*, we might judge it suitable to pass to other editors in the *Lancet*-group for consideration or to editors of other relevant journals within Elsevier's portfolio.

Appeals

- Sometimes editors make mistakes. When we do, we like to hear about them. If an author believes that an editor has made an error in declining a paper, we welcome an appeal. In your appeal letter, which should be sent to thelancetHIV@lancet.com, please state why you think the decision is mistaken, and set out your specific responses to any peer reviewers' comments if those seem to have been the main cause of rejection.
- At least two editors will decide whether to invite a revised manuscript and whether re-review, or otherwise, is indicated.

Proofs

- *The Lancet* journals employ highly skilled Assistant Editors, and it is likely that your paper will be substantially edited after acceptance to ensure that it is accurate, clear, and understandable to a wide readership.
- All figures will be redrawn into *The Lancet* style by our in-house illustrators.
- You will receive a proof from an Assistant Editor. That proof should be corrected and returned within 48 h.

Editorial research

- We are keen to better understand and improve editorial conduct, decision making, and issues related to peer review. Therefore, we occasionally take part in or do editorial research. Your submitted paper might be used in such research. If you do not want your paper entered into such a study, please let us know in your covering letter. Your decision to take part or not will have no effect on the editorial decision on your paper.

Open access and funding

Open access

- *The Lancet* journals are committed to support authors in making their research publicly and freely available. The editors encourage all authors to post their peer-reviewed, accepted article on their personal or institutional websites any time after publication in print or online. Your document should indicate the article's citation and a link to the published article on *The Lancet* website.
- For submissions of research articles from April 1, 2013, funded by Arthritis Research UK, Austrian Science Fund, British Heart Foundation, Cancer Research UK, UK Chief Scientist Office, UK Department of Health UK, UK Department of International Development (DFID), Dunhill Medical Trust, Motor Neuron Disease Association, Parkinson's UK, one of the UK Research Councils, Telethon Italy, or Wellcome Trust; for submissions from Jan 1, 2016, funded by WHO (including International Agency for Research on Cancer [IARC]); for submissions from April 1, 2016, funded by Bill & Melinda Gates Foundation; for submissions from May 1, 2016, funded by Breast Cancer Now or Bloodwise; and for submissions from July 1, 2016, funded by Worldwide Cancer Research, we offer either a "gold" open access choice or a "green" open access solution.
- For the gold open access solution, we offer a choice of creative commons licences (CC BY or CC BY-NC-ND) after payment of an article processing charge of US$5000.
- For the green open access solution, authors can deposit the final accepted version of their paper in any repository they choose 6 months after publication. Additionally, for authors who choose the green open access solution we will also make the published paper free to access on our websites 6 months after publication. See below for copyright and reuse information.
- These options will not be applied retrospectively.

NIH Public Access Policy

- To allow authors to comply with the National Institutes of Health (NIH) Public Access Policy, we will deposit accepted articles (final peer-reviewed but unedited version) reporting research that is directly funded by NIH to PubMed Central no later than 12 months after publication. For authors who are NIH employees (but not for those with just NIH funding), any peer-reviewed accepted article of any type will be deposited by us in PubMed Central in its unedited format no later than 12 months after publication. All Howard Hughes Medical Institute scientists who are authors of accepted Articles can also have their accepted manuscripts deposited in PubMed Central if they complete the relevant section of the Journal Publishing Agreement.

Click here for Elsevier's agreements with funding bodies.

Copyright and reuse

Authors will be asked to sign a transfer of copyright agreement, which recognises the common interest that both journal and author(s) have in the protection of copyright. We accept that some authors (eg, government employees in some countries) are unable to transfer copyright. However, such policies do not provide anyone other than *The Lancet* journals the right to make in any form facsimile copies of the version printed.

Gold open access articles are published under Creative Commons licensing, which enables authors to retain copyright while allowing others to copy, distribute, and make some uses of their work, provided full credit is given to them as originators. Authors will be offered a choice of two licences (CC BY or CC BY-NC-ND) depending on whether or not they wish to allow commercial reuse of their work and whether or not they wish to allow others to alter their work in the course of its reuse. Authors will be asked to sign an exclusive licence (or non-exclusive licence for government employees) to permit our publisher, Elsevier, to publish the work.

For Creative Commons licensing see http://creativecommons.org/licenses/.

PubMed Central

For content funded by the UK Medical Research Council or Wellcome Trust, we will deposit the final version of the manuscript in PubMed Central on publication.

What happens after publication?

Press release

Press releases are issued by *The Lancet* journals' press office for selected content published in our journals. You will be advised in advance if your paper has been selected for press release. *The Lancet* journals' media relations team will contact you with detailed instructions about the embargo for your paper, and will provide a draft press release for your comments ahead of the publication date. If your institution would like to issue a press release for your paper, please inform pressoffice@lancet.com.

Press release
http://www.thelancet.com/press-room

Author interview

Your paper may be selected for a podcast. If so, the Web Editor will contact you to arrange a pre-recorded interview to discuss your paper. For more information, see **Audio.**

Data storage

Authors may be required to provide the raw data for research papers when they are under review and up to 10 years after publication in *The Lancet HIV*.

Responsible sharing

The Lancet supports responsible sharing. We recognise that authors want to share their papers and we encourage this. Find out how you can share your paper here

For Creative Commons licensing see http://creativecommons.org/licenses/

Responsible sharing
http://www.elsevier.com/sharing-articles

```
1
2                                    (R)
3   /__    /   ____/   /   ____/
4     /  /   /___/   /   /___/
5
6   Data cleaning and dat management
7   User: Dhokotera T
8   Project: Burden of Cancer associated with HIV
9
10  ********************************************************************************
11
12  cd"C:\Projects\BCAH\STATA\Data"
13  use ncr_demographics.dta, clear
14  count
15  unique unique_patient_id
16  destring unique_patient_id,replace
17  destring unique_patient_id, force replace
18
19
20  ///merging demographic data to hiv data using unique number///
21  mmerge unique_patient_id using ncr_hiv_extract.dta,  umatch(unique_patient_id)
22  sort unique_patient_id,stable
23  duplicates tag unique_patient_id, gen(duptag)
24  drop dup
25  rename registered_date hiv_diagy
26  rename episode_no labno
27
28  ////deterministic linkage using laboratory (episode) numbers///
29  mmerge labno using NCR2004_2014.dta,  umatch(labno)
30  rename _merge merge
31  save cancer_nhlshiv
32
33  ****textmining data******
34  import delimited "C:\Projects\BCAH\STATA\Data\EXCEL DATA\hiv_pos (2).csv",clear
35  save text_mine_hivpos
36
37  import delimited "C:\Projects\BCAH\STATA\Data\EXCEL DATA\hiv_neg.csv", clear
38  save text_mine_hivneg
39  append using text_mine_hivpos
40  gen hiv=1 if hiv_status=="positive"
41  replace hiv=0 if hiv_status=="negative"
42  drop hiv_status
43  rename lab_no labno
44  save hiv_text_mining2
45
46  ****linking cancer data to textmining data***
47  use NCR2004_2014.dta, clear
48  mmerge labno using hiv_text_mining2.dta,  umatch(labno)
49  rename hiv hiv_status
50  keep if hiv_status!=.
51  drop _merge is_duplicated gender notes report_type v1
52
53  append using cancer_nhlshiv
54
55  sort unique_patient_id
56  duplicates tag unique_patient_id, gen(duptag)
57  tab duptag
58  codebook regno
59  save ncr_cancer_hiv_merge, replace
60  br unique_patient_id Rep_cat result_text
61
62  ********************************************************************************
    ********************************************************************************
63
64
65  ///DATA CLEANING///
66  use ncr_cancer_hiv_merge,clear
67  br unique_patient_id patient_firstname patient_surname Surname Firstname regno
    unique_patient_id labno test_method_name result_value result_text result_conformed
68  sort unique_patient_id,stable
69  unique regno
70
71  duplicates report unique_patient_id regno result_text result_value result_conformed
    result_value
72  duplicates drop unique_patient_id  regno result_text result_value result_conformed
```

```
    result_value,force
73  count
74  unique regno
75
76  keep mdesc morph_group race gender prov hosp facility_name facility_code calc_pop tdesc
    Surname Firstname sndname hiv_status nhls_priv unique_patient_id IMP_POP sex age hiv_dia
     aggp agp regno test_method_name result_value result_text result_conformed cd4_count
    cd4_percent elisa_result_type diagdate Rep_cat diagy province_name
77
78  codebook cd4_count cd4_percent elisa_result_type
79
80  ///ASSIGN HIV STATUS USING TEST METHODS AND HIV RESULTS///
81  ****hiv results***
82  replace hiv_status=0 if substr(result_text,1,3)=="NEG"
83  replace hiv_status=0 if result_text=="NON-REACTIVE"
84  replace hiv_status=0 if result_text=="NR"
85  replace hiv_status=0 if result_text=="N"
86  replace hiv_status=1 if result_text=="REACTIVE"
87  replace hiv_status=1 if substr(result_text,1,3)=="POS"
88  replace hiv_status=1 if result_text=="P"
89  replace hiv_status=1 if result_text=="RE"
90  replace hiv_status=0 if result_conformed =="NEGATIVE"
91  replace hiv_status=1 if substr(result_conformed,1,4) =="REAC"
92  replace hiv_status=0 if substr(result_conformed,1,5) =="NON-R"
93  replace hiv_status=1 if result_conformed =="POSITIVE"
94  replace hiv_status=1 if result_conformed =="REACTIVE  "
95
96
97
98  tab hiv_status
99  tab hiv_status, mi
.00 codebook result_text,detail
.01 codebook result_text,tab(39)
.02
.03 ***test methods***
.04 br test_method_name result_value result_text result_conformed elisa_result_type if
    hiv_status==.
.05 tab test_method_name
.06 replace hiv_status=1 if test_method_name =="HIV VIRAL LOAD"
.07 replace hiv_status=1 if test_method_name =="CD4 ARV "
.08 replace hiv_status=1 if test_method_name =="CD4 ARV FACSCALIBUR"
.09
.10 destring cd4_count, replace force
.11 replace hiv_status=1 if cd4_count!=.
.12
.13 replace hiv_status=2 if result_text=="LOW POSITIVE" & hiv_status<=1
.14 replace hiv_status=2 if result_text=="HIVLP" & hiv_status<=1
.15 replace hiv_status=2 if result_text=="WR" & hiv_status<=1
.16 replace hiv_status=2 if result_text=="WEAK" & hiv_status<=1
.17 replace hiv_status=2 if result_text=="EQUIVOCAL" & hiv_status<=1
.18 replace hiv_status=2 if result_text=="INDETERMINATE" & hiv_status<=1
.19
.20
.21 tab hiv_status, mi
.22 br test_method_name result_value result_text result_conformed elisa_result_type if
    hiv_status==.
.23
.24
.25 codebook test_method_name
.26 codebook test_method_name,tab(15)
.27
.28 br test_method_name result_value result_text result_conformed elisa_result_type if
    hiv_status==.
.29 duplicates report regno hiv_status Rep_cat
.30 duplicates drop regno hiv_status Rep_cat, force
.31 count
.32 unique regno
.33
.34 gsort regno hiv_status
.35 by regno : generate dup=cond(_N==1,0,_n)
.36 br if dup>1
.37 drop if dup>=2 & hiv_status==.
.38 drop dup
.39
.40
```

```
gsort regno hiv_status
by regno : generate dup=cond(_N==1,0,_n)
br regno hiv_status Rep_cat dup
drop if dup==1 & hiv_status==0
drop dup
codebook regno

gsort regno hiv_status
by regno : generate dup=cond(_N==1,0,_n)
br if dup>1
drop if dup>=2 & hiv_status==2
codebook regno



tab hiv_status
tab hiv_status, mi

drop cd4_percent elisa_result_type test_method_name result_value result_text
result_conformed cd4_count
use ncr_hiv_public_private, clear

drop if nhls_priv=="P"
save ncr_hiv_public, replace



**********DATA  CLEANING OF MERGED HIV CANCER DATA**********
use ncr_hiv_public, clear
gen hiv_diag_date = date(hiv_diagy, "DMY")
format hiv_diag_date %td

replace province_name="GAUTENG" if province_name=="CITY OF JOHANNESBURG METRO"
replace province_name=prov if province_name==""
tab province_name, mi
replace province_name="EASTERN CAPE" if province_name=="EC"
replace province_name="FREE STATE" if province_name=="FS"
replace province_name="GAUTENG" if province_name=="GAU"
replace province_name="KWAZULU-NATAL" if province_name=="KZN"
replace province_name="LIMPOPO" if province_name=="LIM"
replace province_name="MPUMALANGA" if province_name=="MPU"
replace province_name="NORTHERN CAPE" if province_name=="NC"
replace province_name="" if province_name=="NOT APPLICABLE"
replace province_name="" if province_name=="UNKNOWN"
replace province_name="WESTERN CAPE" if province_name=="WC"
tab province_name, mi

save, replace
keep hosp province_name
gsort hosp -province_name
br
by hosp : generate dup=cond(_N==1,0,_n)
drop if dup>=2
drop if hosp==""
rename province_name PROVINCE
save hosp_province, replace


*********merge by hospital to assign province*******
use ncr_hiv_public, clear
mmerge hosp using hosp_province.dta

replace province_name=PROVINCE if province_name==""
tab province_name, mi

encode province_name, g(Province)
labelbook Province
tab Province, mi
drop province_name

********CHANGING DATA FROM STRING FORMAT//LABELLING DATA //CATEGORISING DATA*********
replace sex=gender if sex==""
drop gender
encode sex, g(gender)
drop sex
```

```
tab gender, mi
replace gender=. if gender==3
br if gender==.

tab gender, mi

replace calc_pop=race if calc_pop==""
drop race
encode calc_pop, g(race)
tab race, mi
replace race=. if race==4
br Surname Firstname sndname calc_pop if race==.
drop calc_pop

gen age_group=.
tostring age_group, replace
destring age, replace
replace age_group= "0-14" if age<=14
replace age_group= "15-19" if age>14 & age<=19
replace age_group= "20-24" if age>19 & age<=24
replace age_group= "25-29" if age>24 & age<=29
replace age_group= "30-34" if age>29 & age<=34
replace age_group= "35-39" if age>34 & age<=39
replace age_group= "40-44" if age>39 & age<=44
replace age_group= "45-49" if age>44 & age<=49
replace age_group= "50-54" if age>49 & age<=54
replace age_group= "55-59" if age>54 & age<=59
replace age_group= "60+" if age>=60 & age!=.
encode age_group, gen(age_cat)
tab age_cat, mi

drop age_group



label define pop 1"ASIAN" 2"BLACK" 3"COLOURED" 5"Unknown" 6"WHITE"
label values race pop
tab race
replace race=. if race==5
label define sex 1"FEMALE" 2"MALE" 3"UNKNOWN"
label values gender sex

label variable gender "GENDER"
label variable age_cat "AGE CATEGORIES"
label variable hiv_status "HIV STATUS"
label variable diagy "YEAR OF CANCER DIAGOSIS"
label variable age "AGE"
label variable diagy "CANCER DIAGNOSIS DATE"
label variable hiv_diag_date "HIV DIAGNOSIS DATE"
drop dup  Surname Firstname sndname



****assigning cancers as AIDS defining and NON-AIDS DEFINING****
encode Rep_cat, g(cancer_diagnosis)
drop Rep_cat

****GENERATING INDIVIDUAL CANCER VARIABLES***
gen anal_cancer=1 if cancer_diagnosis==1
replace anal_cancer=0 if cancer_diagnosis!=1

gen BCC=1 if cancer_diagnosis==2
replace BCC =0 if cancer_diagnosis!=2

gen bladder_cancer=1 if cancer_diagnosis==3
replace bladder_cancer=0 if cancer_diagnosis!=3

gen bone_cancer=1 if cancer_diagnosis==4
replace bone_cancer=0 if cancer_diagnosis!=4

gen brain_cancer=1 if cancer_diagnosis==5
replace brain_cancer=0 if cancer_diagnosis!=5

gen breast_cancer=1 if cancer_diagnosis==6
replace breast_cancer=0 if cancer_diagnosis!=6
```

```
290
291   gen burkitts=1 if cancer_diagnosis==7
292   replace burkitts=0 if cancer_diagnosis!=7
293
294   gen cervical_cancer=1 if cancer_diagnosis==8
295   replace cervical_cancer=0 if cancer_diagnosis!=8
296
297   gen colorectal_cancer=1 if cancer_diagnosis==9
298   replace colorectal_cancer=0 if cancer_diagnosis!=9
299
300   gen connective_tissue=1 if cancer_diagnosis==10
301   replace connective_tissue=0 if cancer_diagnosis==10
302
303   gen endocrine_cancer=1 if cancer_diagnosis==11
304   replace endocrine_cancer=0 if cancer_diagnosis==11
305
306   gen eye_cancer=1 if cancer_diagnosis==12
307   replace eye_cancer=0 if cancer_diagnosis!=12
308
309   gen Conjuctiva=1 if tdesc=="Conjunctiva"
310   replace Conjuctiva=0 if Conjuctiva==.
311
312   gen gum_cancer=1 if cancer_diagnosis==13
313   replace gum_cancer=0 if cancer_diagnosis!=13
314
315   gen haematology=1 if cancer_diagnosis==14
316   replace haematology=0 if cancer_diagnosis!=14
317
318   gen hodgkin_lymphoma=1 if cancer_diagnosis==15
319   replace hodgkin_lymphoma=0 if cancer_diagnosis!=15
320
321   gen ill_defined=1 if cancer_diagnosis==16
322   replace ill_defined=0 if cancer_diagnosis!=16
323
324   gen Kaposi=1 if cancer_diagnosis==17
325   replace Kaposi=0 if cancer_diagnosis!=17
326
327   gen kidney=1 if cancer_diagnosis==18
328   replace kidney=0 if cancer_diagnosis!=18
329
330   gen larynx=1 if cancer_diagnosis==19
331   replace larynx=0 if cancer_diagnosis!=19
332
333   gen leukaemia=1 if cancer_diagnosis==20
334   replace leukaemia=0 if cancer_diagnosis!=20
335
336   gen lip=1 if cancer_diagnosis==21
337   replace lip=0 if cancer_diagnosis!=21
338
339   gen liver=1 if cancer_diagnosis==22
340   replace liver=0 if cancer_diagnosis!=22
341
342   gen lung_cancer=1 if cancer_diagnosis==23
343   replace lung_cancer=0 if cancer_diagnosis!=23
344
345   gen melanoma=1 if cancer_diagnosis==24
346   replace melanoma=0 if cancer_diagnosis!=24
347
348   gen mesothilioma=1 if cancer_diagnosis==25
349   replace mesothilioma=0 if cancer_diagnosis!=25
350
351   gen mouth=1 if cancer_diagnosis==26
352   replace mouth=0 if cancer_diagnosis!=26
353
354   gen myeloma=1 if cancer_diagnosis==27
355   replace myeloma=0 if cancer_diagnosis!=27
356
357   gen naso_oropharynx=1 if cancer_diagnosis==28
358   replace naso_oropharynx=0 if cancer_diagnosis!=28
359
360   gen NHL=1 if cancer_diagnosis==29
361   replace NHL=0 if cancer_diagnosis!=29
362   replace NHL =1 if burkitts==1
363
364   gen oesophageal_cancer=1 if cancer_diagnosis==30
```

```
replace oesophageal_cancer=0 if cancer_diagnosis!=30

gen others=1 if cancer_diagnosis==31
replace others=0 if cancer_diagnosis!=31

gen ovary=1 if cancer_diagnosis==32
replace ovary=0 if cancer_diagnosis!=32

gen pancreas=1 if cancer_diagnosis==33
replace pancreas=0 if cancer_diagnosis!=33

gen penis=1 if cancer_diagnosis==34
replace penis=0 if cancer_diagnosis!=34

gen placenta=1 if cancer_diagnosis==35
replace placenta=0 if cancer_diagnosis!=35

gen primarysite_unknown=1 if cancer_diagnosis==36
replace primarysite_unknown=0 if cancer_diagnosis!=36

gen prostate_cancer=1 if cancer_diagnosis==37
replace prostate_cancer=0 if cancer_diagnosis!=37

gen scc_skin=1 if cancer_diagnosis==38
replace scc_skin=0 if cancer_diagnosis!=38

gen salivary_gland=1 if cancer_diagnosis==39
replace salivary_gland=0 if cancer_diagnosis!=39

gen skin=1 if cancer_diagnosis==40
replace skin=0 if cancer_diagnosis!=40

gen merkelcellcarcinoma=1 if mdesc=="Merkel cell carcinoma"
replace merkelcellcarcinoma=0 if merkelcellcarcinoma==.

gen small_intestines=1 if cancer_diagnosis==41
replace small_intestines=0 if cancer_diagnosis!=41

gen stomach=1 if cancer_diagnosis==42
replace stomach=0 if cancer_diagnosis!=42

gen testis=1 if cancer_diagnosis==43
replace testis=0 if cancer_diagnosis!=43

gen thyroid=1 if cancer_diagnosis==44
replace  thyroid=0 if cancer_diagnosis!=44

gen tongue=1 if cancer_diagnosis==45
replace tongue=0 if cancer_diagnosis!=45

gen uteral_cancer=1 if cancer_diagnosis==46
replace uteral_cancer=0 if cancer_diagnosis!=46

gen vagina=1 if cancer_diagnosis==47
replace vagina=0 if cancer_diagnosis!=47


gen vulva=1 if cancer_diagnosis==48
replace vulva=0 if cancer_diagnosis!=48

gen viral_cancers=1 if cancer_diagnosis==1
replace viral_cancers=1 if cancer_diagnosis==22
replace viral_cancers=1 if cancer_diagnosis==47
replace viral_cancers=1 if cancer_diagnosis==48
replace viral_cancers=1 if cancer_diagnosis==34
replace viral_cancers=1 if cancer_diagnosis==15
replace viral_cancers=0 if viral_cancers==.

gen cancer_type=.
replace cancer_type= 1 if Kaposi==1
replace cancer_type=1 if NHL==1
replace cancer_type= 1 if cervical_cancer==1
replace cancer_type=0 if cancer_type==. & cancer_diagnosis!=36
```

```
gen non_viral=1 if viral_cancers==0
replace non_viral=0 if non_viral==.

gen adc=1 if cancer_type==1
replace adc=0 if adc==.
gen nadc=1 if cancer_type==0
replace nadc=0 if nadc==.
gen site_unknown=1 if cancer_type==.
replace site_unknown=0 if site_unknown==.


label variable cancer_type "Type of cancer"
label define adc_vs_nadc 0"NADC"  1"ADC"
****remember missing is primary site unknown****
label values cancer_type adc_vs_nadc

label variable tdesc "TOPOGRAPHY DESCRIPTION"
label variable mdesc "MORPHOLOGY DESCRIPTION"
label variable cancer_diagnosis "CANCER DIAGNOSIS"
label variable diagdate "CANCER DIAGNOSIS DATE"
label variable  viral_cancers "VIRUS RELATED CANCERS"
label variable  non_viral "NON-VIRUS RELATED CANCERS"
label variable  adc "AIDS DEFINING CANCER"
label variable  nadc "non-AIDS DEFINING CANCER"
label variable  site_unknown "POORLY SPECIFIED HISTOLOGY SITE"
label variable  non_viral "NON-VIRUS RELATED CANCERS"

save analysis_data, replace


********** GENERATING THE GENERAL POPULATION ESTIMATES CORRECTED FOR HIV //using both ipolation and extrapolation methods*****
**********************to be used for calculating standardized rates******************
import delimited "C:\Projects\BCAH\STATA\Data\EXCEL DATA\prevalence 2008-20012 males.csv", clear
rename hiv_prevalence hiv_prevalence_males
bysort age_cat: ipolate hiv_prevalence_males year, g(hiv_ipo) epolate
gen hiv_prev=hiv_prevalence_males
replace hiv_prevalence_males = hiv_ipo
drop hiv_ipo
encode age_cat, g(age)
drop age_cat
rename age age_cat
generate gender=.
replace gender=2
save hiv_prevalence_males, replace

clear
import delimited "C:\Projects\BCAH\STATA\Data\EXCEL DATA\prevalence 2008-20012 females.csv", clear
rename hiv_prevalence hiv_prevalence_females
bysort age_cat: ipolate hiv_prevalence_females year, g(hiv_ipo) epolate
gen hiv_prev=hiv_prevalence_females
replace hiv_prevalence_females= hiv_ipo
drop hiv_ipo
encode age_cat, g(age)
drop age_cat
rename age age_cat
generate gender=.
replace gender=1
save hiv_prevalence_females, replace

append using hiv_prevalence_males
rename hiv_prevalence_females hiv_prevalance
replace hiv_prevalance= hiv_prevalence_males if hiv_prevalance==.
drop hiv_prevalence_males

label define sex 1 "FEMALE" 2 "MALE"
label values gender sex
save hiv_prevalence, replace



import delimited "C:\Projects\BCAH\STATA\Data\EXCEL DATA\STATS SA POPULATION ESTIMATES.csv", clear
```

```
encode age, g(age_cat)
drop age
label define sex 1 "FEMALE" 2 "MALE"
label values gender sex
drop v5 v6
sort gender
sort year
sort year age_cat gender
****adjusting for the 20% difference between NHLS catchment and population estimates****
merge m:1 age_cat using percentage_adjustment
gen percent_pop= pop*(percentage/100)
rename pop general_pop
rename percent_pop pop
keep if _merge==3
drop _merge
///add prevalence data fom the hiv prevalence study///
merge 1:1 year age_cat gender using hiv_prevalence
drop if _merge==1
drop if _merge==2
drop _merge


gen hiv_neg_prev=.
replace hiv_neg=100- hiv_prevalance

gen correct_pop=round(pop*(hiv_neg_prev/100))
gen hiv_pop=round(pop*(hiv_prevalance/100))
egen age_yr_gen = group(year gender age_cat), label
save population_data, replace

*****PREPARING DATA FOR MERGING WITH POPULATION ESTIMATES*********
use analysis_data,clear

drop  adc nadc viral_cancers non_viral Conjuctiva merkelcellcarcinoma anal_cancer BCC
bladder_cancer bone_cancer brain_cancer breast_cancer burkitts cervical_cancer
colorectal_cancer connective_tissue endocrine_cancer eye_cancer gum_cancer haematology
hodgkin_lymphoma ill_defined Kaposi kidney larynx leukaemia lip liver lung_cancer melano
 mesothilioma mouth myeloma naso_oropharynx NHL oesophageal_cancer others ovary pancreas
penis placenta primarysite_unknown prostate_cancer salivary_gland scc_skin skin
small_intestines stomach testis thyroid tongue uteral_cancer vulva vagina

egen age_yr_gen = group( diagy gender age_cat ), label
gen adc=.
replace adc= cancer_type if cancer_type==1
gen nadc=.
replace nadc= cancer_type if cancer_type==0
gen primarysiteunknown=1 if cancer_type==.
gen positive=.
replace positive= hiv_status if hiv_status==1
gen negative=.
replace negative = hiv_status if hiv_status==0

gen anal_cancer=1 if cancer_diagnosis==1
gen BCC=1 if cancer_diagnosis==2
gen bladder_cancer=1 if cancer_diagnosis==3
gen bone_cancer=1 if cancer_diagnosis==4
gen brain_cancer=1 if cancer_diagnosis==5
gen breast_cancer=1 if cancer_diagnosis==6
gen burkitts=1 if cancer_diagnosis==7
gen cervical_cancer=1 if cancer_diagnosis==8
gen colorectal_cancer=1 if cancer_diagnosis==9
gen connective_tissue=1 if cancer_diagnosis==10
gen endocrine_cancer=1 if cancer_diagnosis==11
gen eye_cancer=1 if cancer_diagnosis==12
gen gum_cancer=1 if cancer_diagnosis==13
gen haematology=1 if cancer_diagnosis==14
gen hodgkin_lymphoma=1 if cancer_diagnosis==15
gen ill_defined=1 if cancer_diagnosis==16
gen Kaposi=1 if cancer_diagnosis==17
gen kidney=1 if cancer_diagnosis==18
gen larynx=1 if cancer_diagnosis==19
gen leukaemia=1 if cancer_diagnosis==20
gen lip=1 if cancer_diagnosis==21
gen liver=1 if cancer_diagnosis==22
gen lung_cancer=1 if cancer_diagnosis==23
```

```
gen melanoma=1 if cancer_diagnosis==24
gen mesothilioma=1 if cancer_diagnosis==25
gen mouth=1 if cancer_diagnosis==26
gen myeloma=1 if cancer_diagnosis==27
gen naso_oropharynx=1 if cancer_diagnosis==28
gen NHL=1 if cancer_diagnosis==29
gen oesophageal_cancer=1 if cancer_diagnosis==30
gen others=1 if cancer_diagnosis==31
gen ovary=1 if cancer_diagnosis==32
gen pancreas=1 if cancer_diagnosis==33
gen penis=1 if cancer_diagnosis==34
gen placenta=1 if cancer_diagnosis==35
gen primarysite_unknown=1 if cancer_diagnosis==36
gen prostate_cancer=1 if cancer_diagnosis==37
gen scc_skin=1 if cancer_diagnosis==38
gen salivary_gland=1 if cancer_diagnosis==39
gen skin=1 if cancer_diagnosis==40
gen small_intestines=1 if cancer_diagnosis==41
gen stomach=1 if cancer_diagnosis==42
gen testis=1 if cancer_diagnosis==43
gen thyroid=1 if cancer_diagnosis==44
gen tongue=1 if cancer_diagnosis==45
gen uteral_cancer=1 if cancer_diagnosis==46
gen vagina=1 if cancer_diagnosis==47
gen vulva=1 if cancer_diagnosis==48
gen merkelcellcarcinoma=1 if mdesc=="Merkel cell carcinoma"
gen Conjuctiva=1 if tdesc=="Conjunctiva"
gen viral_cancers=1 if cancer_diagnosis==1
replace viral_cancers=1 if cancer_diagnosis==22
replace viral_cancers=1 if cancer_diagnosis==47
replace viral_cancers=1 if cancer_diagnosis==48
replace viral_cancers=1 if cancer_diagnosis==34
replace viral_cancers=1 if cancer_diagnosis==15
gen non_viral=1 if cancer_type==nadc & viral_cancers==.


save precollapse_ASIR, replace

drop if hiv_status==.
drop if age_cat==.
collapse (count) non_viral viral_cancers merkelcellcarcinoma anal_cancer BCC
bladder_cancer bone_cancer brain_cancer breast_cancer burkitts cervical_cancer
colorectal_cancer connective_tissue endocrine_cancer eye_cancer Conjuctiva gum_cancer
haematology hodgkin_lymphoma ill_defined Kaposi kidney larynx leukaemia lip liver
lung_cancer melanoma mesothilioma mouth myeloma naso_oropharynx NHL oesophageal_cancer
others ovary pancreas penis placenta primarysite_unknown prostate_cancer salivary_gland
scc_skin skin small_intestines stomach testis thyroid tongue uteral_cancer vulva vagina
positive negative adc nadc, by( age_yr_gen hiv_status age_cat gender)
**********MERGING DATA TO GENERAL POPULATION ESTIMATES******
merge m:1 age_yr_gen using population_data
drop _merge

************************GENERATING THE POPULATION VARIABLE**********************

rename pop gen_pop
gen pop=.
replace pop=hiv_pop if hiv_status==1
replace pop=correct if hiv_status==0
codebook pop
br if pop<=0
replace pop=5241 if pop==-5241
save ASIR_COMPLETECASE, replace


**********COMBINIG HIV- + UNKNOWN ASSUMING THEY ARE BOTH NEGATIVE**********
use precollapse_ASIR, clear
drop if age==.
replace hiv_status=0 if hiv_status==.
collapse (count) non_viral viral_cancers merkelcellcarcinoma anal_cancer BCC
bladder_cancer bone_cancer brain_cancer breast_cancer burkitts cervical_cancer
colorectal_cancer connective_tissue endocrine_cancer eye_cancer Conjuctiva gum_cancer
haematology hodgkin_lymphoma ill_defined Kaposi kidney larynx leukaemia lip liver
lung_cancer melanoma mesothilioma mouth myeloma naso_oropharynx NHL oesophageal_cancer
others ovary pancreas penis placenta primarysite_unknown prostate_cancer salivary_gland
scc_skin skin small_intestines stomach testis thyroid tongue uteral_cancer vulva vagina
```

```
positive negative adc nadc, by( age_yr_gen hiv_status age_cat gender)

**********MERGING DATA TO GENERAL POPULATION ESTIMATES******
merge m:1 age_yr_gen using population_data
drop _merge

************************GENERATING THE POPULATION VARIABLE**********************

rename pop gen_pop
gen pop=.
replace pop=hiv_pop if hiv_status==1
replace pop=correct if hiv_status==0
codebook pop
br if pop<=0
replace pop=5241 if pop==-5241
save ASIR, replace
```

```
 1
 2                                  (R)
 3   /__    /    ____/   /    ____/
 4     /  /   /___/   /   /___/
 5
 6   Data Analysis
 7   User: Dhokotera T
 8   Project: Burden of Cancer associated with HIV
 9
10
11   ********************************************************************************
12
13   cd"C:\Projects\BCAH\STATA\Data"
14   use analysis_data, clear
15
16   ****demographic characteristics summary for all public sector records*****************
17   tab age_cat, mi
18   tab gender, mi
19   tab race, mi
20   tab Province, mi
21   tab hiv_status, mi
22   tab cancer_type, mi
23   tab diagy, mi
24
25   ***********converting the indeterminate result to unknown****
26   replace hiv_status=. if hiv_status==2
27   tab hiv_status, mi
28
29
30   *******OBJECTIVE 1: DEMOGRAPHIC CHARACTERISTICS BY HIV STATUS******************
31   tab gender hiv_status, mi col chi2
32   tab age_cat hiv_status, mi col chi2
33   tab race hiv_status, mi col chi2
34   tab Province hiv_status, mi col chi2
35   tab cancer_type hiv_status, mi col chi2
36   tab diagy hiv_status, mi col chi2
37
38
39
40   sort cancer_type
41   by cancer_type: tab diagy hiv_status
42
43   tab hiv_status cancer_type if hiv_status==1
44   proportion cancer_type, over(hiv_status)
45
46   replace hiv_status=2 if hiv_status==.
47   foreach var in gender age_cat race Province cancer_type diagy {
48   proportion `var', missing over(hiv_status)
49   est store `var'
50   estout `var' using proportions.txt, c("b(fmt(2)) ci(fmt(3))")   append
51   }
52
53   ////percentage contribution of NADC vs ADC total cancer burden across the years///
54   tab diagy cancer_type, if hiv_status==1
55
56
57
58   ******OBJECTIVE 2: ESTIMATE ASIR and SIR************************
59   ********calculating overall cancer incidence in HIV+ and in HIV-***********
60   use ASIR, clear
61
62   ///negative population value///
63
64   drop if age_cat==.
65   gen overall_cancer=pos
66   replace overall_cancer=neg if overall_cancer==0
67   distrate overall_cancer pop using standard_pop, stand(age_cat) mult(100000) by(hiv_statu
68   distrate overall_cancer pop using standard_pop, stand(age_cat) by( year hiv_status )
     saving(overall, replace) format(%8.2f) mult(100000)
69   use overall.dta, clear
70   outsheet using overallcomplete.csv , comma
71
72    **********TREND IN CANCER GROUPS BY HIV_STATUS///COMPLETE CASE ANALYSIS**************
73    use ASIR_COMPLETECASE,clear
74
```

```
75   foreach var in nadc adc viral non_viral {
76   use ASIR_COMPLETECASE,clear
77   distrate `var' pop using standard_pop, stand(age_cat) by( year hiv_status) saving(`var',
     replace) format(%8.2f) mult(100000)
78   use `var'.dta, clear
79   keep if hiv_status==1
80   outsheet using `var'3.csv, comma
81   }
82
83   *********SPECIFIC CANCER TRENDS************************
84   foreach var in cervical_cancer uteral_cancer vulva vagina {
85   use ASIR, clear
86   keep if gender==1 & hiv_status==1
87   distrate `var' pop using standard_pop, stand(age_cat) by( year) saving(`var', replace)
     format(%8.2f) mult(100000)
88   use `var'.dta, clear
89   keep if hiv_status==1
90   outsheet using `var'.csv , comma
91   }
92
93   foreach var in penis{
94   use ASIRtrends, clear
95   drop if gender==2
96   distrate `var' pop using standard_pop, stand(age_cat) by( year hiv_status ) saving(`var'
     replace) format(%8.2f) mult(100000)
97   use `var'.dta, clear
98   keep if hiv_status==1
99   outsheet using `var'.csv , comma
.00  }
.01
.02  foreach var in  anal_cancer burkitts eye_cancer Conjuctiva hodgkin_lymphoma Kaposi NHL
     oesophageal_cancer penis scc_skin prostate_cancer uteral_cancer vulva vagina {
.03  use ASIRtrends, clear
.04  distrate `var' pop using standard_pop, stand(age_cat) by( year hiv_status ) saving(`var'
     format(%8.2f) mult(100000)
.05  use `var'.dta
.06  keep if hiv_status==1
.07  outsheet using `var'.csv , comma
.08  }
.09
.10
.11
.12
.13  **************COMPUTING INDIVIDUAL SIR*************
.14  foreach var in penis prostate_cancer testis{
.15  use ASIR, clear
.16  drop if gender==1
.17  distrate `var' pop using standard_pop, stand(age_cat) by(hiv_status ) saving(`var',
     replace) format(%8.2f) mult(100000)
.18  }
.19
.20  foreach var in cervical_cancer uteral breast vulva vagina placenta{
.21  use ASIR, clear
.22  drop if gender==2
.23  distrate `var' pop using standard_pop, stand(age_cat) by(hiv_status ) saving(`var',
     replace) format(%8.2f) mult(100000)
.24  }
.25
.26  foreach var in merkelcellcarcinoma anal BCC bladder bone brain burkitts colorectal
     Conjuctiva connective_tissue endocrine eye gum{
.27  use ASIR, clear
.28  distrate `var' pop using standard_pop, stand(age_cat) by(hiv_status) saving(`var',replac
      format(%8.2f) mult(100000) list( `var' rateadj lb_gam ub_gam  se_gam srr lb_srr ub_srr)
.29  }
.30
.31  foreach var in haematology hodgkin_lymphoma ill_defined Kaposi kidney larynx leukaemia
     leukaemia lip liver lung melanoma mesothilioma mouth myeloma naso_oropharynx NHL
     oesophageal others {
.32  use ASIR, clear
.33  distrate `var' pop using standard_pop, stand(age_cat) by(hiv_status) saving(`var',replac
      format(%8.2f) mult(100000) list( `var' rateadj lb_gam ub_gam se_gam srr lb_srr ub_srr)
.34  }
.35
.36  foreach var in  pancreas primarysite_unknown salivary_gland scc_skin skin small_intestir
      stomach testis thyroid tongue{
```

```
use ASIR, clear
distrate `var' pop using standard_pop, stand(age_cat) by(hiv_status) saving(`var',replac
 format(%8.2f) mult(100000) list( `var' rateadj lb_gam ub_gam se_gam srr lb_srr ub_srr)
}


 penis.dta prostate_cancer.dta testis.dta


**********SENSITIVITY ANALYSIS COMBINING HIV NEG AND UNKNOWN**********


use ASIR, clear

gen overall_cancer=pos
replace overall_cancer=neg if overall_cancer==0
distrate overall_cancer pop using standard_pop, stand(age_cat) mult(100000) by(hiv_statu
distrate overall_cancer pop using standard_pop, stand(age_cat) by( year hiv_status )
saving(overall, replace) format(%8.2f) mult(100000)
use overall.dta, clear
outsheet using overallcomplete.csv , comma


 **********TREND IN CANCER GROUPS BY HIV STATUS**********
 use ASIR,clear

foreach var in nadc adc viral non_viral {
use ASIR,clear
distrate `var' pop using standard_pop, stand(age_cat) by( year hiv_status) saving(`var',
replace) format(%8.2f) mult(100000)

use `var'.dta, clear
keep if hiv_status==1
outsheet using `var'.csv, comma
}

*********SPECIFIC CANCER TRENDS**********
foreach var in breast cervical_cancer uteral_cancer vulva vagina {
use ASIR, clear
keep if gender==1
distrate `var' pop using standard_pop, stand(age_cat) by( year hiv_status ) saving(trend
 append) format(%8.2f) mult(100000)
use `var'.dta, clear
keep if hiv_status==1
outsheet using `var'.csv , comma
}

foreach var in penis{
use ASIR, clear
drop if gender==2
distrate `var' pop using standard_pop, stand(age_cat) by( year hiv_status ) saving(`var'
replace) format(%8.2f) mult(100000)
use `var'.dta, clear
keep if hiv_status==1
outsheet using `var'.csv , comma
}

foreach var in  anal_cancer burkitts eye_cancer Conjuctiva hodgkin_lymphoma Kaposi NHL
oesophageal_cancer penis scc_skin prostate_cancer uteral_cancer vulva vagina {
use ASIR, clear
distrate `var' pop using standard_pop, stand(age_cat) by( year hiv_status ) saving(`var'
format(%8.2f) mult(100000)
use `var'.dta
keep if hiv_status==1
outsheet using `var'.csv , comma
}



**********COMPUTING INDIVIDUAL SIR**********
foreach var in penis prostate_cancer testis{
use ASIR, clear
drop if gender==2
distrate `var' pop using standard_pop, stand(age_cat) by( hiv_status ) saving(`var',
```

```
    replace) format(%8.2f) mult(100000)
205 use `var'.dta, clear
206 keep if hiv_status==1
207 outsheet using `var'.csv , comma
208 }
209
210 foreach var in cervical_cancer uteral breast vulva vagina placenta{
211 use ASIR, clear
212 drop if gender==2
213 distrate `var' pop using standard_pop, stand(age_cat) by( year hiv_status ) saving(`var'
    replace) format(%8.2f) mult(100000)
214 use `var'.dta, clear
215 keep if hiv_status==1
216 outsheet using `var'.csv , comma
217 }
218
219
220 foreach var in merkelcellcarcinoma anal BCC bladder bone brain burkitts colorectal
    Conjuctiva connective_tissue endocrine eye gum{
221 use ASIR, clear
222 distrate `var' pop using standard_pop, stand(age_cat) by(hiv_status) saving(`var',replac
     format(%8.2f) mult(100000) list( `var' rateadj lb_gam ub_gam  se_gam srr lb_srr ub_srr)
223 }
224
225 foreach var in haematology hodgkin_lymphoma ill_defined Kaposi kidney larynx leukaemia
    leukaemia lip liver lung melanoma mesothilioma mouth myeloma naso_oropharynx NHL
    oesophageal others {
226 use ASIR, clear
227 distrate `var' pop using standard_pop, stand(age_cat) by(hiv_status) saving(`var',replac
     format(%8.2f) mult(100000) list( `var' rateadj lb_gam ub_gam se_gam srr lb_srr ub_srr)
228 }
229
230 foreach var in  pancreas primarysite_unknown salivary_gland scc_skin skin small_intestin
     stomach testis thyroid tongue{
231 use ASIR, clear
232 distrate `var' pop using standard_pop, stand(age_cat) by(hiv_status) saving(`var',replac
     format(%8.2f) mult(100000) list( `var' rateadj lb_gam ub_gam se_gam srr lb_srr ub_srr)
233 }
234
235
236 ******LOGISTIC REGRESSION*******************
237
238 ///COMPLETE CASE ANALYSIS///
239
240 use analysis_data.dta, clear
241 drop if hiv_status==.
242
243 foreach var in anal_cancer BCC bladder_cancer bone_cancer brain_cancer  burkitts
    colorectal_cancer  eye_cancer gum_cancer haematology hodgkin_lymphoma Kaposi kidney lary
     leukaemia lip liver  lung_cancer melanoma mesothilioma mouth myeloma naso_oropharynx NH
    oesophageal_cancer pancreas salivary_gland scc_skin skin small_intestines stomach  thyro
     tongue {
244 logistic `var' i.hiv_status age i.race  i.gender  diagy
245 est store `var'
246 estout `var' using OR.txt, c("b(fmt(2)) ci(fmt(2)) p(fmt(2)) se(fmt(2))") stats(N) eform
    varlabel(1.hiv_status `var') keep (1.hiv_status )  append
247 }
248
249 foreach var in cervical_cancer uteral breast vulva vagina placenta {
250 logistic `var' i.hiv_status age  i.race diagy if gender==1
251 est store `var'
252 estout `var' using OR.txt, c("b(fmt(2)) ci(fmt(2)) p(fmt(2)) se(fmt(2))") stats(N) eform
    varlabel(1.hiv_status `var') keep (1.hiv_status )  append
253 }
254
255 foreach var in  penis prostate_cancer testis  {
256 logistic `var' i.hiv_status age  i.race  diagy if gender==2
257 est store `var'
258 estout `var' using OR.txt, c("b(fmt(2)) ci(fmt(2)) p(fmt(2)) se(fmt(2))") stats(N) eform
    varlabel(1.hiv_status `var') keep (1.hiv_status )  append
259 }
260
261 foreach var in Conjuctiva nadc adc site_unknown viral_cancers non_viral {
262 logistic `var' i.hiv_status age  i.race  i.gender diagy
263 est store `var'
```

```
estout `var' using OR.txt, c("b(fmt(2)) ci(fmt(2)) p(fmt(2)) se(fmt(2))") stats(N) eform
varlabel(1.hiv_status `var') keep (1.hiv_status )   append
}

gen Oral_cavity=1 if cancer_diagnosis==21
replace Oral_cavity=1 if cancer_diagnosis==26
replace Oral_cavity=1 if cancer_diagnosis==28
replace Oral_cavity=1 if cancer_diagnosis==45
replace Oral_cavity=1 if cancer_diagnosis==13
replace Oral_cavity=1 if cancer_diagnosis==19
replace Oral_cavity=1 if cancer_diagnosis==39
replace Oral_cavity=0 if Oral_cavity==.

replace Leiomyosarcoma =0 if mdesc!="Leiomyosarcoma, NOS"
replace viral_cancers =1 if cancer_diagnosis==26
replace viral_cancers =1 if cancer_diagnosis==28
replace viral_cancers =1 if cancer_diagnosis==45
replace viral_cancers =1 if cancer_diagnosis==13
replace viral_cancers =1 if cancer_diagnosis==19
replace viral_cancers =1 if cancer_diagnosis==39


gen Oral_cavity=1 if cancer_diagnosis==

logistic Oral_cavity i.hiv_status age  i.race  i.gender diagy [pweight=iprob]

logistic viral_cancers i.hiv_status age  i.race  i.gender diagy [pweight=iprob]

*******SENSITIVITY ANALYSIS*************
******OPTION 1: with ipw estimates*****

****inverse probability weighting
*-----------------------------
use analysis_data.dta, clear

gen obs =1
replace obs = 0 if hiv_status==.
*Use logistic regression to find a model for obs
logit obs i.age_cat i.gender i.cancer_diagnosis i.diagy

predict fitprob


*Using the weights to adjust for missing observations
*-----------------------------------------------------
gen iprob= 1/fitprob
*To get an idea of variation in the weights
hist iprob

/*To weight the analysis in stata is simple, we just include the command [pweight=iprob]
in theregression:*/
foreach var in Conjuctiva anal_cancer BCC bladder_cancer bone_cancer brain_cancer
burkitts colorectal_cancer  eye_cancer gum_cancer haematology hodgkin_lymphoma Kaposi
kidney larynx leukaemia lip liver  lung_cancer melanoma mesothilioma mouth myeloma
naso_oropharynx NHL oesophageal_cancer pancreas salivary_gland scc_skin skin
small_intestines stomach  thyroid tongue {
logistic `var'  i.hiv_status age  i.race  diagy [pweight=iprob]
est store `var'
estout `var' using IPW.txt, c("b(fmt(2)) ci(fmt(2)) p(fmt(2)) se(fmt(2))") stats(N) efor
  varlabel(1.hiv_status `var') keep (1.hiv_status )  append
}

foreach var in cervical_cancer uteral breast vulva vagina placenta {
logistic `var'  i.hiv_status age  i.race diagy if gender==1 [pweight=iprob]
est store `var'
estout `var' using IPW.txt, c("b(fmt(2)) ci(fmt(2)) p(fmt(2)) se(fmt(2))") stats(N) efor
  varlabel(1.hiv_status `var') keep (1.hiv_status )  append
}

foreach var in  penis prostate_cancer testis  {
logistic `var'  i.hiv_status age  i.race  diagy if gender==2 [pweight=iprob]
est store `var'
estout `var' using IPW.txt, c("b(fmt(2)) ci(fmt(2)) p(fmt(2)) se(fmt(2))") stats(N) efor
  varlabel(1.hiv_status `var') keep (1.hiv_status )  append
}
```

```
330
331
332   foreach var in nadc adc site_unknown viral_cancers non_viral oropharyngeal {
333   logistic `var'  i.hiv_status age  i.race  i.gender diagy [pweight=iprob]
334   est store `var'
335   estout `var' using IPW.txt, c("b(fmt(2)) ci(fmt(2)) p(fmt(2)) se(fmt(2))") stats(N) efor
        varlabel(1.hiv_status `var') keep (1.hiv_status ) append
336   }
337
338
339   **********OPTION 2: HIVneg vs HIVneg + unknown**********************
340   use analysis_data, clear
341   tab hiv_status, mi
342   replace hiv_status=0 if hiv_status==.
343
344
345   foreach var in Conjuctivs anal_cancer BCC bladder_cancer bone_cancer brain_cancer
      burkitts colorectal_cancer  eye_cancer gum_cancer haematology hodgkin_lymphoma Kaposi
      kidney larynx leukaemia lip liver  lung_cancer melanoma mesothilioma mouth myeloma
      naso_oropharynx NHL oesophageal_cancer pancreas salivary_gland scc_skin skin
      small_intestines stomach  thyroid tongue {
346   logistic `var'  i.hiv_status age i.race i.gender diagy
347   est store `var'
348   estout `var' using neg_unknown.txt, c("b(fmt(2)) ci(fmt(2)) p(fmt(2)) se(fmt(2))") stats
      ) eform  varlabel(1.hiv_status `var') keep (1.hiv_status )  append
349   }
350
351
352   foreach var in cervical_cancer uteral breast vulva vagina placenta {
353   logistic `var'  i.hiv_status age  i.race diagy if gender==1
354   est store `var'
355   estout `var' using neg_unknown.txt, c("b(fmt(2)) ci(fmt(2)) p(fmt(2)) se(fmt(2))") stats
      ) eform  varlabel(1.hiv_status `var') keep (1.hiv_status )  append
356   }
357
358   foreach var in  penis prostate_cancer testis  {
359   logistic `var'  i.hiv_status age i.race diagy if gender==2
360   est store `var'
361   estout `var' using neg_unknown.txt, c("b(fmt(2)) ci(fmt(2)) p(fmt(2)) se(fmt(2))") stats
      ) eform  varlabel(1.hiv_status `var') keep (1.hiv_status )  append
362   }
363
364
365   foreach var in nadc adc site_unknown viral_cancers non_viral {
366   logistic `var'  i.hiv_status age i.race i.gender  diagy
367   est store `var'
368   estout `var' using neg_unknown.txt, c("b(fmt(2)) ci(fmt(2)) p(fmt(2)) se(fmt(2))") stats
      ) eform  varlabel(1.hiv_status `var') keep (1.hiv_status )  append
369   }
370
371
372
373
374
375   **********************OBJECTIVE 3a: TRENDS OF CANCER RISK IN PLHIV using Odds
      ratios****************************
376
377   use analysis_data, clear
378
379   gen year2004=1 if diagy==2004
380    replace year2004=0 if diagy!=2004 & year2004==.
381
382     gen year2005=1 if diagy==2005
383    replace year2005=0 if diagy!=2005 & year2005==.
384
385     gen year2006=1 if diagy==2006
386    replace year2006=0 if diagy!=2006 & year2006==.
387
388     gen year2007=1 if diagy==2007
389    replace year2007=0 if diagy!=2007 & year2007==.
390
391    gen year2008=1 if diagy==2008
392    replace year2008=0 if diagy!=2008 & year2008==.
393
394     gen year2009=1 if diagy==2009
```

```
 replace year2009=0 if diagy!=2009 & year2009==.

 gen year2010=1 if diagy==2010
 replace year2010=0 if diagy!=2010 & year2010==.

  gen year2011=1 if diagy==2011
 replace year2011=0 if diagy!=2011 & year2011==.

 gen year2012=1 if diagy==2012
 replace year2012=0 if diagy!=2012 & year2012==.

  gen year2013=1 if diagy==2013
 replace year2013=0 if diagy!=2013 & year2013==.

   gen year2014=1 if diagy==2014
 replace year2014=0 if diagy!=2014 & year2014==.

foreach variable in anal_cancer burkitts  eye_cancer Conjuctiva hodgkin_lymphoma Kaposi NHL oesophageal_cancer penis uteral_cancer vagina vulva{
foreach var in year2004 year2005 year2006 year2007 year2008 year2009 year2010 year2011 year2012 year2013 year2014 {
logistic `variable' i.hiv_status if `var'==1
est store `var'
estout `var' using ORtrends.txt, c("b(fmt(2)) ci(fmt(2)) p(fmt(2)) se(fmt(2))") eform varlabel(1.hiv_status `var') keep (1.hiv_status )  append
}
}

foreach variable in uteral_cancer vagina vulva{
foreach var in year2004 year2005 year2006 year2007 year2008 year2009 year2010 year2011 year2012 year2013 year2014 {
logistic `variable' i.hiv_status if `var'==1 & gender==1
est store `var'
estout `var' using ORtrends.txt, c("b(fmt(2)) ci(fmt(2)) p(fmt(2)) se(fmt(2))") eform varlabel(1.hiv_status `var') keep (1.hiv_status )  append
}
}

foreach variable in  penis {
foreach var in year2004 year2005 year2006 year2007 year2008 year2009 year2010 year2011 year2012 year2013 year2014 {
logistic `variable' i.hiv_status if `var'==1 & gender==2
est store `var'
estout `var' using ORtrends.txt, c("b(fmt(2)) ci(fmt(2)) p(fmt(2)) se(fmt(2))") eform varlabel(1.hiv_status `var') keep (1.hiv_status )  append
}
}

foreach variable in adc viral non_viral nadc{
foreach var in year2004 year2005 year2006 year2007 year2008 year2009 year2010 year2011 year2012 year2013 year2014 {
quietly logistic `variable' i.hiv_status age i.race i.gender if `var'==1
est store `var'
punaf , at(hiv_status=0) eform
}
}
```