

https://dspace.org/

Investigating the association between cardiovascular diseases and cancer in South African adults from the GEMS database

Ramadhar, Anishka

2021
https://hdl.handle.net/10539/32164

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

Investigating the Association between Cardiovascular Diseases and Cancer in South African Adults from the GEMS Database

By
Anishka Ramadhar
2244064



Supervisor: Dr Mazvita Muchengeti
MB ChB (UZ), MSc (Wits) PhD (Bern)
Senior Epidemiologist – National Cancer Registry

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 in the field of Epidemiology and Biostatistics.

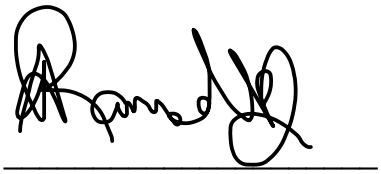
28 May 2021

Declaration

I declare that this research is my own work. It is being submitted in partial fulfilment of the requirements for the degree of Master of Science in Epidemiology in the field of Epidemiology and Biostatistics at the University of Witwatersrand, Johannesburg. This research has not been submitted previously for any degree or examination to any other institution.

Ethics clearance certificate number: M1911142

Name of candidate: Anishka Ramadhar



(Signature of candidate)

Date: 28 May 2021

Dedication:

I dedicate my Masters degree to my Nani (maternal grandmother) for her unwavering belief in me, raising me, loving me unconditionally and constantly reminding me to study hard and work hard in order to make my Nana (maternal grandfather) proud. I love you my Warrior Princess and I thank you for all that you are and all that you have done for me. This… is in your line …

Abstract

Background

The incidence of non-communicable diseases is rising globally and in South Africa, the epidemiologic transition shows a shift from communicable diseases to non-communicable diseases. Cardiovascular disease and cancer are the NCDs with the highest burden and prevalence in South Africa, and these conditions share common risk factors including tobacco smoking, inflammation, type 2 diabetes and obesity. Extensive data on CVD and cancer exist in South Africa but has not been previously explored for an association, nor has the association been explored in South African medically insured patients. In this study we explore the association between CVD exposure and cancer outcome in South African adult patients insured by the Government Employees Medical Scheme (GEMS) from 2014 to 2018 inclusive.

Methods

In this retrospective cohort study, routinely collected secondary data received from GEMS were analysed for an association between CVD and cancer. The prevalence of CVD, cancer, type 2diabetes, COPD, obesity and the CVD and cancer subtypes was calculated in this study population. Univariate and multivariate logistic regression was performed to investigate the association between CVD and cancer and the CVD subtypes and cancer, odds ratios were analysed to indicate which NCD conditions may increase the likelihood of cancer development. Stepwise regression was performed to select variables for a parsimonious model. Time-to-event analysis was conducted to measure the time to cancer occurrence and cancer incidence in patients with and without CVD. Final flexible Generalised Linear Models with cubic splines, were performed to indicate which NCD conditions are risk factors for cancer development.

Results

The study analysis comprised 1 851 615 medical records representing 722 934 unique beneficiaries. The mean age of the patients was 52.68 years with majority of the patients being female (64.48%). In this study population, 9.17% of patients had cancer and 79.31% of patients had CVD, 23.07% of all patients had T2D, 13.53% of patients in the study population had both CVD and cancer. The top 6 cancer subtypes in

descending order are breast cancer, skin neoplasms, prostate cancer, colon cancer, cervical cancer, and lung cancer. The top 6 CVD subtypes in descending order are hypertension, hyperlipidaemia, VET, angina, heart failure and ischaemia. The univariate analysis indicated that age, gender, COPD, heart failure, ischaemia, arrhythmia, atherosclerosis and hyperlipidaemia are all statistically significant risk factors for the development of cancer. The multivariate analysis indicates that age and COPD are statistically significant risk factors for cancer. The interaction between T2D and HT, angina, HF, ischaemia, arrhythmia, cardiomyopathy, and atherosclerosis respectively, increase the risk for cancer development. The interaction between HT and VET, angina, HF, ischaemia, arrhythmia, cardiomyopathy, and atherosclerosis respectively, increase the risk for cancer development.

Discussion

Collectively CVD is not a risk factor for the development of cancer, however, in this study population, CVD subtypes such as hyperlipidaemia and ischaemia increase the risk for cancer development. Interactions between CVD subtypes with T2D and HT, increase the risk for cancer. Although COPD only comprised 1.34% of this study population, regression analysis indicated that COPD is a risk factor for cancer and may be responsible for an 8-fold increase in the likelihood of developing lung cancer. Early treatment and controlling the onset and progression of the CVD subtypes and T2D may delay the onset and progression of cancer. Awareness around the risks and dangers of tobacco smoking may decrease the incidence and severity of COPD and lung cancer, as it may be independently associated with both conditions and may be a confounder. Further data collection and research is required to be conducted on CVD subtypes and their association with cancer and the common risk factors between CVD and cancer.

Acknowledgement

My sincere gratitude to my supervisor Dr Mazvita Muchengeti for her guidance, support, and perfection.

My appreciation to the GEMS team, Dr Guni Goolab and Spencer de Jongh (technical manager) for allowing me access to their database.

My gratitude to all my lecturers at the Faculty of Health Sciences at Wits for imparting their knowledge to make this degree possible.

My special and sincere gratitude to Saj, my human, for your assistance, patience, tolerance, threats to call Sanjay, unconditional support, belief that I in fact am not, “in over my head”, and for holding down this fort whenever I am away, my love and appreciation for you is beyond explanation…. and Nacho and Tinky, for all the love and being the best stress relievers.

A special thanks to Sanjay, my uncle, my teacher, my advisor, and well-wisher, you have been on this journey with me from the inception of my idea to begin my MSc, to my protocol editing and now as I come to the close of this chapter, you are loved and valued.

Special thanks for my aunt Rashmee, and my Mum Shama, for their food and saving me time from preparing tasty dishes. Your time and love are appreciated. To my uncles Jack and Muni, your constant support is what gets me through these challenges.

Thanks to the rest of my family and friends for their belief and support that I will complete this degree, to Avitesh for giving up your Sunday over lockdown to attempt to run my software, to Arishta, Seema and Sanushka for accommodating me for a few days during lectures, to Vee and Meshu for assistance with transport, and lastly to my Employer, Novartis, for affording me this opportunity.

Table of Contents

Declaration.....2

Dedication.....3

Abstract.....4

Acknowledgements.....6

Table of contents 7

List of figures 9

List of tables 9

Nomenclature.....10

CHAPTER 1.....10

INTRODUCTION.....10

1.1 Background.....11

1.1.1 Non-communicable diseases (NCDs).....11

1.1.2 Global prevalence of NCDs.....11

1.1.3 Non-communicable disease (NCD) risk factors.....11

1.1.4 Health status of South Africa and Sub-Saharan Africa 12

1.1.5 The NCD burden 12

1.1.6 Epidemiology of NCDs.....12

1.1.7 Epidemiological transition.....13

1.1.8 Disease pattern shifts.....14

1.1.9The South African health situation.....14

1.1.10 Disease trends in South Africa.....15

1.2 Literature Review.....16

1.2.1 CVD burden and risk factors.....16

1.2.2 Cancer burden and risk factors.....17

1.2.3 Associations between CVD and cancer.....19

1.3 Problem Statement 21

1.4 Justification.....21

1.5 Research question, aim and objectives.....22

1.5.1 Research question.....,22

1.5.2 Aim 22

1.5.3 Objectives.....22

CHAPTER 2.....24

METHODOLOGY 24

2.1 Study design……24
2.2 Study setting……24
2.3 Study population……24
2.4 Data source……24
2.5 Variables……25
2.6 Data management……25
2.7 Data analysis……27
2.7.1 Prevalence……27
2.7.2 Logistic regression: univariate and multivariate……28
2.7.3 Stepwise Regression, Survival analysis and Generalised Linear Models……28
2.8 Ethical considerations……30
CHAPTER 3……30
RESULTS……31
3.1 Flowchart……31
3.2 Prevalence of CVD and cancer……32
3.3 Associations between CVD and cancer………33
3.4 Time-to-event analysis……37
3.5 Establishing cancer risk factors ……40
CHAPTER 4……41
DISCUSSION AND CONCLUSION……41
4.1 Principal findings……41
4.2 Strengths and weaknesses……42
4.3 Strengths and weaknesses in relation to other studies……43
4.4 Meaning of the study: possible mechanisms and implications for clinicians or policymakers……47
4.5 Unanswered questions and future research……49
4.6 Overall conclusions and recommendations……49
REFERENCES……51
APPENDICES……54
Appendix 1: Plagiarism Declaration Form……54
Appendix 2: Ethics clearance certificate……55

List of figures

Figure 1: Time to cancer occurrence from the first claim in the total study population…35
Figure 2: Time to CVD occurrence from the first claim in the total study population……36
Figure 3: Time to cancer occurrence in patients with CVD……36
Figure 4: Time to cancer occurrence in patients without CVD……..37

List of tables

Table 1: Descriptive Analysis……....31
Table 2: Prevalence……..31
Table 3: Regression analysis ……32
Table 4: Top 6 cancer with top 6 CVD including Age……..32
Table 5: Stepwise regression……33
Table 6: Final model ……34
Table 7: 2x2 table showing cancer patients with CVD……35
Table 8: Cancer incidence rate (per 100 000) from time to event analysis populations…37
Table 9: Generalised Linear Model Regression – Risk factors for Cancer outcome……..38

Nomenclature (acronyms)

NCD – non-communicable disease
CVD – cardiovascular disease
WHO – World Health Organisation
T2D – type 2 diabetes
SSA – sub-Saharan Africa
CRD – chronic respiratory disease
CKD – chronic kidney disease
COPD – chronic obstructive pulmonary disease
HT - hypertension
GBD – global burden of disease
DALY – disability adjusted life years
YLL – years of life lost
YLD – years of life lived with disability
ET – epidemiologic transition
LMICs – low-and-middle-income-countries
TB - Tuberculosis
OR – odds ratio
GLM – Generalised Linear Models
KM – Kaplan Meier
IARC – International Agency for Research on Cancer
IHD – Ischaemic heart disease
GEMS – Government Employees Medical Scheme
ICD10 – International Classification of Diseases and Related Health Problems (10th Revision)

Chapter 1: Introduction

In this introductory chapter, the contextual background and literature review for the study are presented.

1.1Background

1.1.1 Non-communicable diseases (NCDs)

1.1.2 Global prevalence of NCDs

1.1.3 Non-communicable disease (NCD) risk factors

1.1.4 Health status of South Africa and Sub-Saharan Africa

1.1.5 The NCD burden

1.1.6 Epidemiology of NCDs

1.1.7 Epidemiological transition

1.1.8 Disease pattern shifts

1.1.9The South African health situation

1.1.10 Disease trends in South Africa

1.2 Literature Review

1.2.1 CVD burden and risk factors

1.2.2 Cancer burden and risk factors

1.2.3 Associations between CVD and cancer

1.1 Background

1.1.1 Non-communicable diseases (NCDs)

The World Health Organisation (WHO) predicts that over the following 10 years, there will be a 17% increase in deaths caused by non-communicable diseases (NCDs) (1). The increase in NCDs is transpiring at a faster rate in low to middle income compared to high-income countries. NCDs are affecting individuals at younger ages and are causing premature death to individuals in these populations (2). In September 2018, the WHO amended the main diseases in the NCD category from 4 to 5. This amendment includes cardiovascular diseases (CVD), Type 2 diabetes (T2D), chronic respiratory diseases (CRD), cancers and the latest inclusion being mental health diseases (3).

1.1.2 Global Prevalence of NCDs

Globally, NCDs are a serious health concern, having a 40% prevalence in emerging countries and a 10% prevalence in developed countries in the 30-year to 70-year age group (4). Lessened productivity levels of individuals who have NCDs as well as the costly treatment creates a heavy economic burden in countries which have a high prevalence of NCDs. Lifestyle induced NCDs render the highest rate of premature mortality among patients (4).

1.1.3 Non-communicable disease (NCD) risk factors

NCDs arise from multiple risk factors such as an unhealthy diet, tobacco use, alcohol abuse, raised blood pressure; cholesterol, T2D, physical inactivity and substance abuse (5). Aside from these risk factors, socio-economic situations create an environment in which NCDs may thrive (5). Long working hours, minimal wage, unhealthy working environments, and high unemployment rates allow for anxiety and depression which in turn increase the risk of CVD and respiratory diseases (6). In health systems, NCDs are responsible for a large amount of the budget because of high costs involved in diagnosis, treatment, management, and education of the NCDs. Untimely death and weight of disability resulting from NCDs in the most constructive years of a patient's life is indicative of the distress of individuals, families, and the larger public (7).

1.1.4 Health status of South Africa and sub-Saharan Africa

South Africa falls within the sub-Saharan African (SSA) region of the African continent. The treatment of NCDs in the African countries creates a financial burden on the low-income households that have no medical insurance (2). The health status of populations in Sub-Saharan Africa (SSA) has shown notable improvement in the last 30 years but the increase in NCDs jeopardises the progress of the health status and may hinder the socio-economic advances in SSA. This indicates that NCDs are a public health problem as well as an economic issue (2).

1.1.5 The NCD burden

The prevalence of NCD is rising within South Africa, the African continent and internationally. Seventy-two percent of global deaths are from NCDs and 80% of these deaths are due to cancer, cardiovascular disease (CVD), diabetes and respiratory conditions (8). Globally in 2017, 4% (1.7 million) of NCD deaths occurred in people below the age of 30, 38% (15.2 million) in people between the ages of 30 and 70 years, and 58% (23.6 million) in the elderly above the age of 70. Eighty-five percent of the deaths caused by NCDs occur in emerging countries (9). Males and females both had a higher possibility of dying before they reached the age of 70 from a NCD than from communicable, nutritional, maternal, and perinatal diseases collectively (8). The low income and middle-income countries especially in sub-Saharan Africa had the highest risk of NCD deaths (8). According to the WHO 2016 report, over 75% of NCD deaths (31.5 million) were attributed to low- and middle-income countries with approximately 46% of the deaths from these countries occurring before 70 years of age (10).

1.1.6 Epidemiology of NCDs

The epidemiology of NCDs in SSA is poorly described due to the absence of reputable vital statistic systems and credible data at a population-level for most countries within this region. Existing research from countries within SSA shows an increase in T2D, CVDs, cancers, mental health disorders, substance use disorders, chronic respiratory disease (CRD) and chronic kidney disease (CKD) (6,7). Surveillance results for NCD risk factors in SSA imply that majority of adults are vulnerable to one or more risk factors for NCDs. These risk factors include, but are not limited to, obesity, hypertension (HT), tobacco use, unhealthy diets, physical inactivity, or harmful alcohol use (7). In 2016 NCDs were responsible for 34% of the total deaths in Africa. The

Global Burden of Disease (GBD) study of 2017 measures population level disease burden and uses the measurement of disability adjusted life-years (DALY) (9). DALYs account for the accumulation of years of life lost (YLL) owing to untimely mortality as well as years of life lived with disability (YLDs). The DALY measurement therefore incorporates the non-fatal burden and fatal burden of disease (7,9). In SSA, across all age groups, there was an increase in total DALYs due to NCDs by 67·0% between 1990 (90·6 million) and 2017 (151·3 million). An aging population and increased population size explain most of the NCD DALY increase. The 2017 NCD age-standardised DALY rate (21 757.7 DALYs) was nearly the same as communicable, nutritional, neonatal and maternal diseases (26 491.6 DALYs), with CVDs causing 15.1% of the total NCD burden (7).

In South Africa (SA), 51% of the deaths in 2018 were from NCDs, compared to 39% in 2010 (1,11). Globally and in South Africa, CVDs carry the greatest burden, incidence, and prevalence from all the NCDs (3). Following CVDs, the NCDs that carry a large burden are T2D, respiratory conditions and cancer (12). The WHO report of 2018 for South Africa shows that 19% of the NCD deaths were claimed by CVDs, 10% by cancer, 7% by T2D and 4% by CRD. The highest risk factors in South Africa are physical inactivity, obesity, raised blood pressure, and tobacco use (13).

1.1.7 Epidemiological transition:

Epidemiological transition (ET) is a term used to describe the proportion and prevalence shift of NCDs and communicable diseases within a population. This shift indicates a noticeable increase in the number of NCD patients in a population that previously had a higher prevalence of communicable (infectious) diseases (14). Demographic movement patterns comprising primarily of economic advancement and an aging population drive the ET from communicable diseases to NCDs in low- and middle-income countries (LMICs) (15). The lifestyle-induced risk factors further propel the increase of NCDs, fuelled in turn by changes in global proliferation such as socio-cultural and economic adjustments, increasing planes of disposable income, fast economic growth, and rapid urbanisation (14,15). The mounting NCD burden resulting from the ET is a threat to individual and population level health as well as national and global healthcare systems and socioeconomic advancement (15).

1.1.8 Disease pattern shifts:

Since the beginning of the 21st century, SSA disease prevalence entered the first phase of the ET (14). Communicable diseases claimed the most lives in LMICs in the African and Asian continent (9). Diseases such as HIV, malaria, dengue, cholera, influenza, and tuberculosis (TB) are some of the infectious diseases, which, without treatment, caused death in a short period of time to numerous people (20). The low-and-middle income regions are now experiencing a shift in disease patterns where NCD mortality is increasing, especially in Africa. Patients either have a combination of communicable diseases and NCDs or are developing NCDs due to lifestyle changes (16). In many SSA countries, as communicable diseases decrease and antiretroviral therapy (ART) access increases, life expectancy increases and is accompanied with the problem of rising NCDs (7). SSA does not have the most accurate means of assessing for disease prevalence resulting in meagre measures of morbidity data and findings, therefore ET evidence is derived from mortality data confined to hospitals and post-mortem reports (17).

1.1.9 The South African health situation:

South Africa has socio-economic inequality since the nation comprises of populations with the poorest living conditions as well as those with the wealthiest living conditions in the world (16). Development in the country is inconsistent as the high-income areas are continually developed and poor areas face endless delays in development (18). This results in South Africa having a wide range of patients that suffer from infectious diseases as well as NCDs (16).

South Africa has a large population of HIV positive patients, with public health funding prioritising the distribution of free antiretroviral treatment (ART) across the country (16). The use of ART has prolonged the lifespan of HIV infected patients and these patients often die from NCDs rather than HIV. South African HIV-positive patients on ART show a high prevalence of rising visceral lipid layer, hyperglycaemia, and dyslipidaemia, and are therefore predisposed to T2D, hypertension and CVD (16,17). Increasing age shows a strong association with increasing comorbidities. In the above 40 age group, there is an overlap between NCDs and communicable diseases, with HIV, T2D and HT being the most common comorbidity combination (18).

1.1.10 Disease trends in South Africa:

Disease patterns in South Africa have shifted from a high communicable disease prevalence to an increase in NCD incidence and prevalence (19). Western lifestyle, health risk factors and the socio-economic vulnerability of South Africans, have increased their susceptibility to NCDs (19). Incomplete and interrupted trends in health, mortality and disease are indicative of the continuous uncertainty and fragility of the SSA countries and its people. SSA countries are lagging the furthest globally, in health system enhancement and longevity (20). Rural South Africa is facing a prolonged epidemiological transition with the interaction of HIV and ART, NCD risk factors and the intricate behavioural and social changes (19). Early and untimely mortality in South Africa can be attributed to NCDs, which jeopardises the progression and the socio-economic development of the country (16).

1.2 Literature Review

1.2.1 CVD burden and risk factors:

CVD, incorporating heart disease and stroke is the foremost cause of death and disabilities globally (31% of total global deaths and 45% of NCD deaths) (21,22). The appearance of CVDs originates from communicable diseases (such as rheumatic heart disease, HIV, pericarditis induced from TB and viral cardiomyopathies) and NCDs such as stroke, myocardial infarction (MI), peripheral artery disease and HT (23). NCD based CVDs have increased in all urban and semi-urban communities because of ET resulting from an increase in HT, T2D and obesity (23). HT is the leading risk factor for death from CVD, responsible for 13% of deaths globally (26).

Africa has a population of more than 1.3 billion people and contributes largely to the global CVD burden (14). In SSA in the year 2013, CVDs claimed the lives of approximately 1 million people, equating to 5.5% of the global deaths caused by CVD and 11.3% of the total deaths in Africa (14). CVD mortality contributed 38% of deaths caused by NCDs in Africa, this is indicative of the increasing problem of both CVDs and NCDs (14). In the past decade CVD deaths have increased by more than 50% in SSA resulting in increased DALYs. Ischemic heart disease (IHD) (5% of all mortality, 40% of CVDs mortality), stroke (4% of all mortality, 34% of CVD mortality) and hypertension (6–8% of all mortality, 19–24% of CVD mortality) are the common causes of CVD death in SSA (24).

Developing economies in SSA give rise to a change in income-related diets which allow for the increased consumption of energy-rich foods. This heightened consumption thereby increases the risk factors contributing to CVDs which boosts the overall CVD burden. All demographic groups across all socio-economic classes within SSA show a high prevalence of CVD risk factors resulting in an increased burden of CVD in the region (22). Among SSA adults, the prevalence of HT is 30%, T2D is 7.1% and dyslipidaemia is 25% (22).

In South Africa, after TB and T2D, CVD is the leading cause of death responsible for 1 in 6 deaths (25). Daily in South Africa, 215 people die from heart disease or strokes. Hourly 5 people experience heart attacks, 10 people have strokes and from these events, 10 people die (26).

Two out of 3 women and 1 in 3 men are overweight or obese in South Africa. Eighty percent of CVD is caused by unhealthy lifestyles, such as diets low in fruit and vegetable and the lack of physical activity. One in 4 adults have high LDL-cholesterol, 1 in 10 adults are diagnosed with diabetes, and 1 in 5 adults have impaired blood glucose control which over time, leads to co-morbidities such as CVD (26).

The major risk factors for CVD are HT, hypercholesterolemia, and hyperglycaemia, whilst the emerging risk factors are insulin resistance, atherogenic dyslipidaemia, proinflammatory and prothrombotic states (27). Building blocks to these risk factors are tobacco use, unhealthy diet, sedentary lifestyles, obesity, substance abuse and anxiety and depression from unemployment or being over-worked (28). South Africans across all demographic groups and socio-economic classes present with increased rates of microvascular and macrovascular risk factors leading to CVD. Atherosclerotic CVD (ASCVD) followed by congestive heart failure (CHF) and T2D are responsible for the high burden of DALYs (28).

1.2.2. Cancer burden and risk factors:

Globally, the cancer burden and corresponding demand on health systems and economies is increasing with the most weight on countries with vulnerable populations. Growing populations give rise to increasing cancer cases annually where LMICs experience the sharpest rise in cases. Life expectancy improvements further allow additional time for the clinical manifestation of cancer (29).

In 2016, there were 17.2 million cancer cases and 8.9 million cancer deaths worldwide. Globally, cancer cases increased by 28% between 2006 and 2016, prostate cancer showed the greatest incidence (1.4 million cases) for men and in females, breast cancer was responsible for claiming the most lives (535 000 deaths) and showed the greatest incidence (1.7 million cases) (30). Globally, the overall principal source of cancer deaths was caused by tracheal, bronchial, and lung cancer (1.2 million deaths) (30).

The International Agency for Research on Cancer (IARC) classified 11 infectious agents as group 1 carcinogens. The 4 leading pathogens - Helicobacter Pylori, Human papillomavirus (HPV), hepatitis B virus (HBV), and hepatitis C virus (HCV) – are

responsible for more than 90% of global infection-related cancers (31). In 2018, 13% of the global cancer cases were caused by an infectious agent (31). Liver cancer is the 5th and 9th most common cancer globally in males and females, respectively and is the second leading cause of global cancer deaths (32). Globally, cancer of the cervix is ranked at number 4 for both mortality and incidence and accounts for 6.6% of female cancers. Cervical cancer is responsible for 90% of the deaths in women in LMICs with SSA countries bearing the greatest burden (33).

The African population is increasing, and individuals are living to older ages. Many countries in Africa are part of the emerging market economies and show an increase in the prevalence of cancer risk factors related to the evolving economy (34). These risk factors combined with the aging and growing population, have resulted in an increase in cancer incidence (34). The African countries struggle with access to healthcare facilities, and prompt health diagnosis and treatment, resulting in poor cancer prognosis and high mortality rates. The 2018 statistics reflected these healthcare difficulties in Africa, as the proportion of cancer deaths (7.3%) was greater than the proportion of cancer incidence (5.8%) (35). The Global Cancer Incidence, Mortality and Prevalence (GLOBOCAN) predicted that by 2030; the cancer deaths will increase by 70% globally and the cancer burden in Africa will include 1.4 million new cases and 1 million cancer mortalities (36).

After IHD, cancer deaths claim the most lives globally among NCDs. In Africa, infectious agents and diseases are the foremost cause of cancer (31). In 2018, approximately 25% of cancers in Africa resulted from infections and 28.7% in SSA (31,32). In SSA females, more than a third of cancers result from infections. Human papillomaviruses (HPV) are responsible for causing 15.4% of the cancers in SSA and 25.4% of cancers in SSA females (leading to cervical, oral, anal, penile, vaginal, and vulvar cancer) (31). Kaposi's sarcoma-associated herpesvirus (KSHV) in second place claims 4.1% of cancers in SSA. HBV and HCV cause 2.7% of cancers in SSA which is the same for H. Pylori. Hepatocellular carcinoma (HCC) is the leading cause of cancer deaths in Africa driven by high rates of HVB and HVC (32). When a cancer patient presents with multiple infections, the case is attributed to only one of those infectious agents (37).

Yearly, approximately 115 000 South Africans are diagnosed with cancer, with a survival rate of 60% (38). The most common types of cancers among South African males are prostate, colorectal, lung and Kaposi Sarcoma. Among women, the most common types of cancers are breast, cervical, colorectal and cancer of the uterus (39). In a 2001 global study comparing the risks of multiple types of cancers, from 7 million cancer deaths, 35% were due to 9 risk factors which can be regulated and potentially eliminated (39). The NCR report of 2014 indicates that breast cancer is the most common cancer in South Africa (21.8% of all female cancers). This is consistent with global statistics where breast cancer is the most common in females, and after lung cancer, breast cancer is the second most common cancer overall (40, 41). After breast cancer, cervical cancer is the 2nd leading cause of cancer in South African females (33).

All cancer risk assessment models agree on common cancer risk factors which include age, history of familial cancer, ancestry, race, ethnicity, gender, body mass index (BMI), hormonal factors, diet, sedentary lifestyle, infectious agents, alcohol and tobacco use, exposure to hazardous chemicals and radiation (34,40,41). The leading risk factors for low, medium, and high-income countries were smoking, alcohol use, infections and low fruit and vegetable intake. Obesity, smoking, overweight and alcohol abuse were the most common causes of cancer in high-income countries (42).

1.2.3. Associations between CVD and cancer:

Age, obesity, diabetes, smoking, sedentary lifestyles, familial history, genetic predisposition, and inflammatory foods (processed meat, alcohol, high sugar, and wheat containing food) are risk factors for both CVD and cancer (34,43,44). The above risk factors allow for the exploration of an association between CVDs and cancer. Since CVD and cancer have numerous risk factors in common, it is plausible that CVD and cancer accompany each other. These common risk factors spur the idea that there may be a direct effect or association between CVD and cancer. The consistency and repeated occurrence of the same risk factors in CVD and cancer propose that these conditions have more commonalities than was previously accepted (44). Clonal haematopoiesis and alterations in connective tissue, diseased tissue, immune cells, and vasculature may contribute to both CVD and cancer (45). Numerous pathophysiological pathways which include increased oxidation, activated neurohormonal systems and a dysregulated immune system are shared between CVD

and cancer. (45) General bodily inflammation and tissue specific inflammation resulting from pro-inflammatory cytokines produced by tumour or cardiac tissue may be indicative of a direct relationship between CVD and cancer (45, 46). Dietary patterns comprising high dairy fat, red and processed meat, refined sugar, and fried products are positively associated with atherosclerotic disease, T2D and cancer. Diet is therefore a significant exposure common to both CVD and cancer (47).

Excess body weight is a vital player in the NCD scene and is associated with marked reduction in life expectancy. In 2016, globally 40% of adults were carrying excess body weight (approximately 2 billion) attributable to an estimated 4 million deaths (46,47). Visceral fat and intra-organ fat accrual create an environment for atherosclerosis, insulin resistant T2D and tumour development. The atherosclerosis and insulin resistant T2D lead to eventual CVD and tumour development related to intra-organ fat may lead to carcinogenesis of the breast tissue, liver, pancreas, gall bladder and stomach and other organs both in males and females (48). Excess adipose tissue causes insulin resistance and hyperinsulinemia which increase triglycerides that form a basis for CVD and favour tumour development via insulin-like growth factor binding proteins which form a basis for cancer development (48). A lower incidence of diabetes, hyperlipidaemia, cancer, and CVDs were observed in subjects that had undergone malabsorptive bariatric surgery (biliopancreatic diversion and biliointestinal bypass) which further supports excess body weight being a common and apparent risk factor for CVD and cancer (49).

The initiation and development of malignant tumours and atherosclerosis stem largely from an inflammatory process, which is validated by the reduction in CVD events and cancer incidents caused by the blockage of pro-inflammatory cytokines (50). Clonal haematopoiesis being an age-related risk factor for CVD and blood-related malignancies provide another common link between these two disease categories (49,50). The strong age dependence of clonal haematopoiesis strengthens the link of age with CVD and cancer (50). Mutations involved in the transition of epithelial cells to mesenchymal cells are a driver of tumour developments in carcinogenesis as well as arterial disease associated with atherosclerotic events. This transition allows for further plausibility in the association between CVD and cancer (50).

1.3. Problem statement

Extensive CVD and cancer data exist in South Africa. The available data and information have not been previously explored for an association between CVDs and cancer. Additionally, the association between CVDs and cancer has not been explored in South African medically insured patients. Medically insured patients presenting with CVD or cancer may all possess similar risk factors which have not been explored for the common causality of the above diseases. Data on the association between CVD and cancer are limited and sparse and herein lies the literature gap and an opportunity for research.

1.4. Justification

Medically insured patient data, as discussed in the problem statement, has not been utilised for research into NCD associations. In this study, there is an opportunity for the exploration of this data because of the vast information provided by the Government Employees Medical Scheme (GEMS). The information that the study utilised from GEMS consisted of billing details of patients that have claimed from their medical aid. The billing details provide in-depth insight into all and any chronic medical condition that the patients claimed for. This information enables the exploration of the CVD and cancer association, and the investigation of effect modifiers or cofounders that may influence the relationship between CVD and cancer.

The ET in South Africa shows a trend where the mortality is driven by an increasing amount NCDs rather than infectious disease mortality (16,19). The study of the association between the leading NCDs such as CVD and cancer may provide insight into the increasing mortality rate caused by NCDs and provide insight into NCD progression and NCD prognosis. (25) If a significant association between CVD and cancer is discovered in this explorative study, the results may lead to further investigations of the common factors that lead to these diseases and allow for further study into the processes and methods of modifying these factors. Future research on this topic may explain shared pathways and factors common to CVD and cancer that will aid in better understanding and treating these diseases (20). On a public health and economic scale, clinicians will be well disposed to and equipped to monitor and detect various cancers in patients with specific CVDs. This allows for clinicians to maintain a high index of suspicion when consulting with patients (28). Prevention plans and

strategic approaches will help to improve the quality of life for patients at risk for CVD and cancer (28). Saving on treatment cost, by preventing the onset of these disease will lessen the socio-economic burden that accompanies populations who have CVD and cancer (28). CVD outcomes in cancer patients and cancer survivors has already been established, hence this study will not investigate CVD in cancer patients (45).

1.5. Research question, aims and objectives:

1.5.1. Research questions:

1) Is there an association between cardiovascular diseases and cancer?
2) Are CVD patients more likely to develop cancer than patients without CVD?

Hypothesis:

Research hypothesis:

The hypothesis that the study investigated whether CVD patients are more likely to develop cancer than non-CVD patients.

Statistical hypothesis:

- Null hypothesis: There is no difference in the likelihood of developing cancer among patients with CVD and patients without CVD.
- Alternative hypothesis: Patients with CVD have a greater likelihood of developing cancer than patients without CVD.

1.5.2. Aim:

The aim of this study was to investigate the association between CVDs and cancer in South African adults on GEMS from 2014 to 2018 inclusive. The study will investigate cancer outcomes in patients with CVD.

1.5.3. Study objectives:

1.To estimate the prevalence of CVDs and cancer among South African adult patients, with chronic medical conditions, from GEMS between 2014 and 2018 inclusive. The prevalence outcome being measured in this objective will be indicative of the burden of CVD and cancer in the GEMS database among patients who claimed for chronic medical conditions from 2014 to 2018.

2. To establish an association between CVD and cancer among the South African adult population on GEMS. The outcome being measured in this objective is cancer in CVD patients and will be compared to cancer outcome in patients with no CVD.

3. To explore the relationship between the 5 most common CVDs and the 5 most common cancers. The outcome of this objective will indicate the type or types of cancer associated with common CVD types in South African adults with chronic medical conditions on GEMS from 2014 to 2018.

Chapter 2: Methods

2.1. Study design

This study is a retrospective cohort study among adult patients across South Africa who have one or more NCDs and are medically insured by the Government Employees Medical Scheme (GEMS).

2.2. Study setting

Routinely collected secondary claims data were provided from the GEMS database from 2014 to 2018 inclusive. The data received were on NCD medical aid claims from anonymised medical aid patients 18 years and older. When GEMS members utilise their medical insurance for health services, each claim is recorded against the patients' membership and stored in an electronic database. The database record has information on the patient demographics, disease area, prescribed medication and all healthcare services covered by the patients' insurance. Upon request of the NCD data, data were anonymised for ethical purposes and patient confidentiality, and thereafter sent to us for this study.

GEMS was registered on 1 January 2005 to meet the healthcare needs of Government Employees. GEMS provides access to healthcare benefits to all public service employees, providing lower income earners with increased access to this health benefit. Almost 2 million beneficiaries have been enrolled on GEMS, by promoting access to the workforce subsidy. The addition of these beneficiaries and dependents into the medical scheme resulted in a decrease in the public sector health burden and resources. GEMS delivers admission to equitable, comprehensive, and sustainable health subsidies to all public service employees.

2.3. Study population

This study comprises of South African adults who have utilized GEMS for their medical claims. The study population are adults above the age of 18 in males and females, from all population groups across all provinces within South Africa. The racial breakdown of the South African population is 81.17% Blacks, 8.68% Coloured, 7.6% Whites and 2,48% Indians (83). Racial breakdown in South Africa is often indicative of socio-economic status and certain race groups have a greater likelihood of developing

certain cancers and cardiovascular diseases (84, 85). The racial breakdown of the study population was requested as a potential exposure variable, but not received because racial data are not routinely collected by GEMS.

2.4. Data source

The data were sourced from GEMS as deidentified medical claims data to maintain anonymity. The data comprised of routinely collected longitudinal data from patients who have made chronic medical condition claims for their treatment, clinician visits and hospitalisation.

2.5. Variables

The data received from GEMS were ICD10 codes, demographic data on anonymised beneficiaries and medical claims data. The ICD10 codes are the International Statistical Classification of Diseases and Related Health Problems 10th Revision. The ICD10 codes are internationally utilised and accepted codes developed by the WHO to describe all health and injury conditions. These codes are used to inform all medical and health schemes on the conditions their members received treatment for, to correctly settle claims and track medical records. Demographic data comprised of the beneficiaries' geographical location, age, gender, date of birth and marital status. The claims data comprised of the anonymised unique identifier number for each beneficiary, the date of the medical claims and the ICD10 codes for which the claims were made.

The variables that were created specifically for the data analysis were:
ICD10 code, Age, Obesity, Type 2 diabetes, Asthma, COPD, gender, beneficiary, cancer, CVD, cancer subtypes (prostate, colorectal, lung, breast, cervical, colon, brain, Kaposi Sarcoma, myeloma, pancreas, ovary, skin neoplasms, non-Hodgkin's lymphoma, Oesophagus, skin melanoma, uterus, stomach) and CVD subtypes (Hypertension, Hyperlipidaemia, Ischemia, Heart failure, Angina, Atherosclerosis, Arrhythmia, Cardiomyopathy, Venous embolism and thrombosis (VET))

2.6. Data management

Stata 15 statistical software package was utilised for all data cleaning, management, and analysis.

GEMS provided unclassified ICD10 coded data, deidentified claims data and demographic data which shared the same unique beneficiary identifier code as the deidentified claims data. The unclassified ICD10 coded data contained information for all diseases that were recorded by the patients' medical aid claims.

The raw data were cleaned via the process of data checking and reduction to ensure the analysis on the statistical software was performed correctly. The data checking process included checking the data for any missing and miscoded data. The raw data were cleaned to remove strings, claims data was organised according to claim dates and ICD10 codes were arranged alphabetically. Demographic data were reduced to create interpretable date variables, de-string variables to change the format of the data, and create categorical variables grouped by their disease condition, for analysis of an association between the exposure and outcome variables. The ICD10 data were refined to include only the ICD10 variables that contained the main NCD disease code, disease category and disease description.

All the claims' data were appended to create a large claim-data data set and all demographic data were appended to create a large demographic data set. The appended claim, demographic, and ICD10 coded data were merged into a concise dataset in preparation for analysis.

Additional variables were created within the data set and specific variables pertaining to cancer, CVD, and other NCD variables were created. Cancer and CVD subgroup variables were created for further analysis into the disease subtypes.

Inclusion criteria:

- Patients above the age of 18 who claimed for chronic medical conditions from GEMS between 2014 and 2018 inclusive.

Exclusion criteria:

- Individuals below the age of 18.
- Patients diagnosed with cancer before being diagnosed with CVD.
- Patients with a medical history of childhood cancer.
- Patients who claimed purely for infectious diseases.

A baseline date variable and maximum date variable were created to indicate the first date and a last date that each beneficiary made a claim. The beneficiary variable was

sorted by the first date that the beneficiary claimed on medical aid and the last date that they claimed. The beneficiary may have claimed multiple times on the first and last date, but the data was sorted by the first claim and last claim.

An event variable for each condition was created to indicate if any disease event of interest occurred for each beneficiary. The beneficiary may have experienced one or multiple "events", and the "event" variable was created to indicate if the beneficiary had ever experienced any of the diseases of interest to the study. The event variable was tabulated to show that the observations under the beneficiary variable had at least one event for any of the diseases of interest. Duplicate records for all beneficiaries were identified and a variable was created to show the number of duplicated records for each beneficiary. Duplicated records with no event of interest were dropped, but duplicated records that indicated an event of interest, remained in the data set as the duplication may have been for claims of differing diseases, indicating there may have been more than one claim but it may have been for different conditions.

We created additional disease-specific variables to find out if each beneficiary had ever been diagnosed with any of the diseases of interest. The additional variable contained the maximum value of the initial disease type across all observations for the beneficiary. This additional variable for each disease indicated if the beneficiary had ever experienced any of the diseases of interest. The original variable for each disease will only record one event or claim for each beneficiary but the additional variable will record all events or claims for each beneficiary.

A variable that identified the number of claims made per beneficiary was created and used in the logistic regression to control the analysis for unique beneficiaries and not multiple observations of the same beneficiary. We only included one claim for each disease type per unique beneficiary identifier, in this way, logistic regression was performed on the beneficiaries and not on the multiple observations of each beneficiary.

2.7 Data analysis

2.7.1. Prevalence

The first objective of the study was to calculate the prevalence of cancer and CVD in the study population. The prevalence was calculated over the entire study period by

tabulating cancer and CVD over time and tabulating cancer and CVD by gender and age. Our dataset consisted of multiple claim records for each beneficiary as patients can claim for the same disease more than once, therefore only the very first claim recorded for cancer and CVD was used to calculate prevalence and not all claims. Using only the first instance of a claim for a certain disease to calculate prevalence, eradicates duplicate records being counted and eliminates the chance of a falsely high prevalence of disease. The prevalence T2D, COPD, obesity, asthma, cancer subgroup variables and CVD subgroup variables were also calculated.

2.7.2. Logistic regression: univariate and multivariate

The second objective of the study was to investigate the association between CVD and cancer. This objective was met using univariate and multivariate logistic regression models where the outcome was cancer, and the exposure was CVD and the CVD subtypes including gender, and age as *a priori* variable. Binomial logistic regression was performed because the outcome variable was binary, and the independent variables are a mix of categorical and continuous variables. In both the univariate and multivariate logistic regressions, cancer was the outcome variable and CVD, CVD subtype variables, T2D, COPD, obesity, age, and gender were the exposure variables.

After each logistic regression, a goodness of fit test was performed to test the fit of the model. We tested for interactions by including interaction terms in the multivariate logistic regression models. The interaction terms analysed included CVD, T2D and HT as the main variables which were combined with the other CVD subgroup variables. The significance of the interaction variables was assessed to indicate if the terms should be included in the final model. Following each logistic regression, hierarchical regression tables were created to compare the odds ratios of the independent variables on the development of cancer.

The third objective of the study was to investigate the association between the top 5 CVDs and the top 5 cancers. In the analysis phase we selected the top 6 CVDs and top 6 cancers instead of the top 5 of each, because the prevalence of the conditions in the fifth and sixth conditions in both CVDs and cancers did not differ much. Each of the top 6 cancers was the outcome variable in the analysis and a multivariate logistic regression was performed for the top 6 CVDs per each cancer type. A hierarchical

regression table was constructed for the comparison of the effect of the top 6 CVDs on the top 6 cancers, respectively.

2.7.3. Stepwise Regression, Survival analysis and Generalised Linear Models

Forward and backward stepwise regression ($p<0.10$) was performed to identify the final model and thereafter the variables selected from the stepwise regression were added to a model which included interaction terms to create a final model. Following each stepwise regression, hierarchical regression tables were created to compare all the odds ratios of all the independent variables on the development of cancer.

Lastly, the data were set for time-to-event analysis to compare the time to cancer in the total population, in beneficiaries with CVD and in beneficiaries without CVD. In all the time-to-event analyses, the time variable was set as the variable identifying claim dates, the failure event was cancer or time to the first cancer claim which indicated the end of the follow up period, the identifier variable was the unique beneficiary identifier, and the time scale was set to a yearly measurement. For the time-to-cancer analysis in the total population, the start of the follow up time (origin) was measured from the date of the first claim for any of the NCD conditions, for analysis among patients with CVD, the origin was measured from the date of the first CVD claim, and in the analysis among patients without CVD, cancer incidence was analysed in patients who had never claimed for a CVD event.

We used Kaplan Meier (KM) analysis to compare time to cancer development in patients with CVD and patients without CVD. KM analysis also indicated time to cancer and CVD in the entire study cohort and compared males and females for both cancer and CVD. A log-rank test was conducted to measure the statistical significance of the comparison between males and females.

Incidence rate analysis for cancer was performed after setting the data for survival analysis and using the survival analysis summary to generate the incidence rate. The incidence rate for cancer was investigated in all beneficiaries and was compared to the beneficiaries with and without cancer. The risk ratios were calculated as hazard ratios which were generated after the data was set for survival analysis.

Generalised Linear Models (GLM) were used instead of Cox Proportional Hazards (CPH) regression because the CPH regression showed that the assumption of proportional hazards was violated. GLM is a flexible parametric survival analysis model using Poisson distribution where the timescale is split into several time intervals to allow flexibility in the baseline hazard function instead of the hazard being constant throughout one lengthy timescale (51). The GLM assessed the risk of various risk factors – in this study they were the NCD variables, age, and gender - on cancer. The assumption of this model is the baseline hazard is constant within the smaller time intervals but has variation between intervals (51). In the GLM analysis, an offset is included to indicate the length of time each subject was at risk for cancer in each time interval, the outcome variable indicated was always cancer, time at risk was always stipulated and the timescale was split into yearly intervals.

The assumption of constant hazard in each interval is the same as assuming exponential survival distribution within the time interval. Splines are function estimates that allow for the combination of many curves from various intervals into a continuous curve to smooth out the data and smooth out random variation (51). The utilisation of splines smooths the data to obtain a parsimonious estimate of the baseline hazard function. Restricted cubic splines allow for a smooth function to be created from this piecewise model. The split points for the intervals are called knots. The data needs to be collapsed to accommodate the number of intervals (51). The mid-time of each interval was generated to calculate splines for the intervals and the number of cancer occurrences within the intervals. We used 4 degrees of freedom indicating 5 knots placed at the minimum, maximum, 25th 50th and 70th centiles of the mid-time.

2.8. Ethical considerations

The study utilised retrospective de-identified data obtained from the GEMS medical aid scheme. In the study, secondary analysis of de-identified data was performed, hence informed consent was not applicable. Permission to use the data from the GEMS database was granted by Dr Guni Goolab, the principal officer of the GEMS medical aid scheme. The data were released to the primary investigator and the supervisor of the study once the data sharing agreement was signed by both parties.

Before the commencement of research, ethics clearance was granted from the Wits Human Research Ethics Committee (M1911142).

Chapter 3: Results

In this chapter, the results corresponding to each of the study objectives are presented in diagrams, tables, and figures.

3.1 Data management Flowchart

The flowchart describes the data management from data acquisition, data cleaning to the final data analysis. The number of claim observations and number of beneficiaries are recorded in the flowchart.

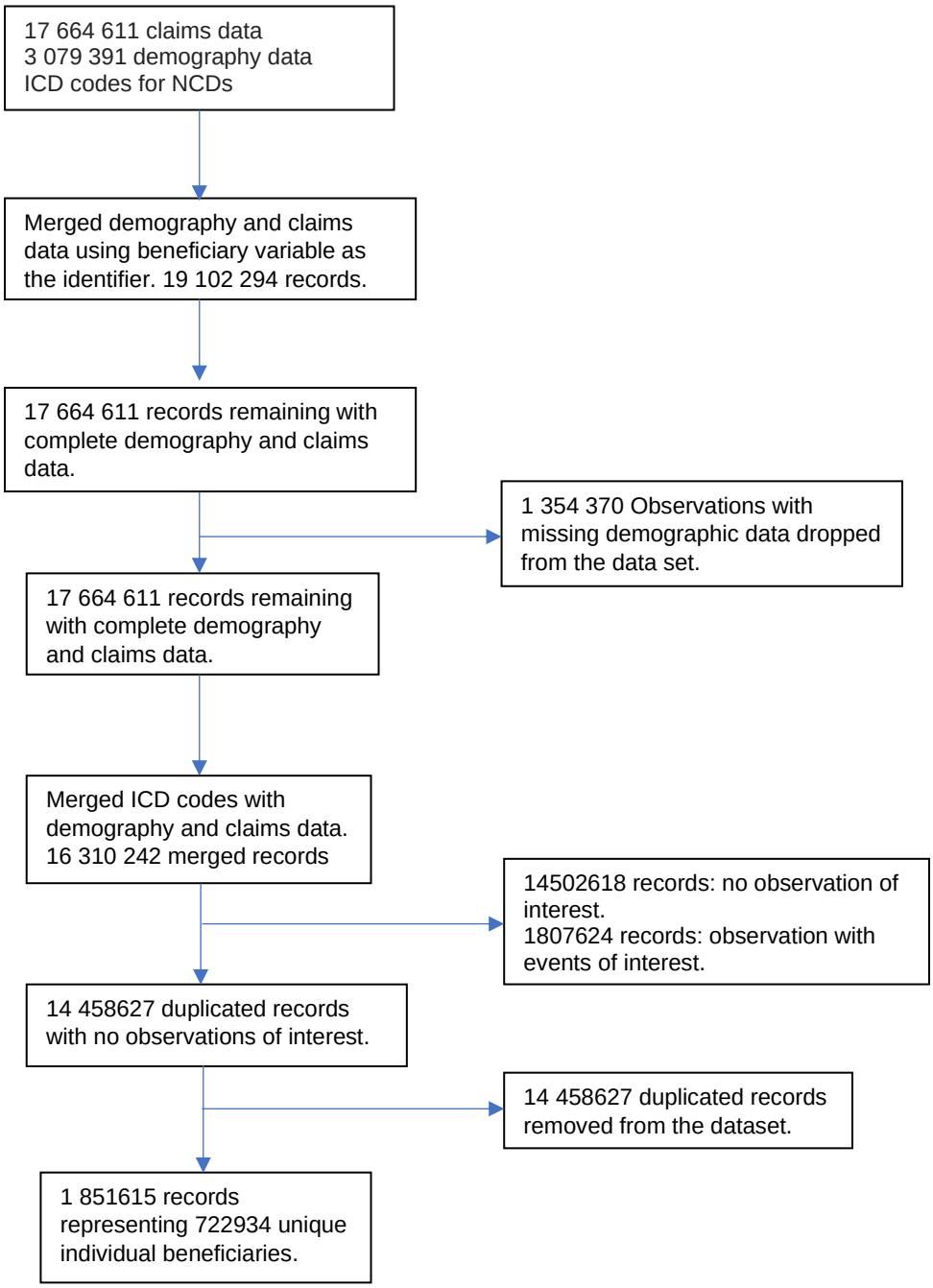


Table 1: Descriptive Analysis (N=722934)					
	n	**Freq**	**Male (%)**	**Female (%)**	**Mean**
Age (years)			52.29	52.89	52.68
Provinces					
Gauteng	168856	23,69	59864 (23,63)	108992 (23,73)	
Western Cape	104458	14,66	41479 (16,37)	62979 (13,71)	
KZN	146853	20,60	49151 (19,40)	97702 (21,27)	
Eastern Cape	92079	12,92	31253 (12,33)	60826 (13,24)	
Northern Cape	22980	3,22	8820 (3,48)	1416 (3,08)	
Limpopo	71463	10,03	25575 (10,09)	45888 (9,99)	
North West	58806	8,25	20702 (8,17)	38104 (8,30)	
Mpumalanga	47255	6,63	16547 (6,53)	30708 (6,68)	
Total			256795 (35.52)	466139 (64.48)	

The study population comprised 64,48% female beneficiaries and 35,52% male beneficiaries with Gauteng, Kwa-Zulu Natal and Western Cape Provinces claiming the most at 23,69%, 20,6% and 14,66% respectively.

3.2. Prevalence of CVD and cancer

Table 2: Prevalence (N=722934)				
	n	**Freq(%)**	**Males (%)**	**Females (%)**
Sex				
Male	256,795	35.52		
Female	466,139	64.48		
Condition				
Cancer	66268	9,17	38,517(10.81)	27,751(8.26)
Breast cancer	10515	1,45	554 (0.22)	9,961 (2.14)
Skin neoplasms	10055	1,39	4,929 (1.92)	5,126 (1.10)
Prostate cancer	9581	1,33	9,525 (3.71)	0.00
Colon cancer	3233	0,45	1,341 (0,52)	1,892 (0,41)
Cervical cancer	2445	0,34	0.00	2,433 (0,52)
Lung cancer	2409	0,33	1,218 (0,47)	1,191 (0,26)
Skin melanoma	1514	0,21	647 (0,25)	867 (0,19)
Brain cancer	1215	0,17	448 (0,17)	767 (0,16)
Non-Hodgkin's	1208	0,17	495 (0,19)	713 (0,15)
Kaposi Sarcoma	1102	0,15	468 (0,18)	634 (0,14)
Ovarian cancer	1062	0,15	0.00	1,056 (0,23)
Pancreatic cancer	1045	0,14	424 (0,17)	621 (0,13)
Stomach cancer	1038	0,14	486 (0,19)	552 (0,12)
Uterine cancer	901	0,12	0.00	898 (0,19)
Oesophageal cancer	863	0,12	408 (0,16)	455 (0,10)
CVD	573363	79,31	75,89	81,19
Hypertension	461790	63,88	157502 (61,33)	304,288 (65,28)
Hyperlipidaemia	131273	18,16	56468 (21,99)	74805 (16,05)
VET	42797	5,92	17069 (6,65)	25728 (5,52)
Angina	27010	3,74	10400(4,05)	16610 (3,56)
Heart failure	21059	2,91	6921 (2,70)	14138 (3,03)
Ischaemia	12145	1,68	6890 (2,68)	5255 (1,13)
Arrhythmia	5848	0,81	1929 (0,75)	3919 (0,84)
Cardiomyopathy	3213	0,44	1316 (0,51)	1897 (0,41)
Atherosclerosis	1999	0,28	894 (0,35)	1105 (0,24)
T2D	166753	23,07	67617 (26,33)	99136 (21,27)
Asthma	59464	8,23	19282 (7,51)	40182 (8,62)
COPD	9697	1,34	4839 (1,88)	4858 (1,04)
Obesity	7300	1,01	1141 (0,44)	6159 (1,32)

In this study, 9.17% of the population had cancer, with 10.81 % of the male patients and 8.26% of the female patients presenting with cancer, 79.31% of all patients had CVD, with 75.89% of the male patients and 81.19% of the female patients having CVD, and 23.07% of all patients had T2D.

The top 6 cancer subtypes in descending order were breast cancer, skin neoplasms, prostate cancer, colon cancer, cervical cancer, and lung cancer. The top 6 CVD subtypes in descending order are hypertension, hyperlipidaemia, VET, angina, heart failure and ischaemia.

3.3. Association between CVDs and cancer

Univariate and multivariate logistic regression analysis

Table 3: Regression analysis (N=722934)						
	Univariate analysis			Multivariate anaysis		
Cancer	Odds Ratio	95% CI	P	Odds Ratio	95% CI	P
CVD	0,55	0,54 - 0,56	< 0,001			
Age	1,04	1,04 - 1,04	< 0,001	1,05	1,05 - 1,05	< 0,001
Gender	1,35	1,32 - 1,37	< 0,001	1,41	1,38 - 1,43	< 0,001
Type 2 Diabetes	0,72	0,71 - 0,74	< 0,001	0,61	0,59 - 0,62	< 0,001
Obesity	0,59	0,53 - 0,65	< 0,001	0,7	0,63 - 0,77	< 0,001
Asthma	0,66	0,64 - 0,68	< 0,001			
COPD	2,39	2,27 - 2,51	< 0,001	1,18	1,12 - 1,25	< 0,001
VET	1,01	0,98 - 1,05	< 0,001	0,77	0,75 - 0,80	< 0,001
Angina	0,75	0,71 - 0,78	< 0,001	0,56	0,53 - 0,59	< 0,001
Heart Failure	1,49	1,43 - 1,55	< 0,001	0,71	0,68 - 0,75	< 0,001
Ischemia	1,74	1,65 - 1,83	< 0,001	0,86	0,81 - 0,90	< 0,001
Arrhythmia	1,19	1,10 - 1,30	< 0,001	0,79	0,72 - 0,86	< 0,001
Cardiomyopathy	0,95	0,84 - 1,08	0,443	0,62	0,55 - 0,70	< 0,001
Atherosclerosis	1,49	1,31 - 1,70	< 0,001	0,73	0,63 - 0,83	< 0,001
Hyperlipidemia	1,23	1,20 - 1,25	< 0,001	0,87	0,85 - 0,88	< 0,001
Hypertension	0,62	0,61 - 0,63	< 0,001	0,39	0,38 - 0,40	< 0,001

The univariate logistic regression indicates that age, gender, COPD, heart failure, ischaemia, arrhythmia, atherosclerosis, and hyperlipidaemia are all statistically significant risk factors for the development of cancer, with COPD, ischaemia HF posing the highest risk, but CVD collectively, is not a risk factor for the development of cancer. The multivariate logistic regression indicates that only age and COPD are statistically significant risk factors for cancer when controlling for the other CVD subtypes and T2D.

Top 6 cancer associations, Stepwise Regression and Time-to-event analysis

Table 4: Top 6 cancer with top 6 CVD including Age (N=722934) odds ratios						
	Prostate	**Breast**	**Skin neo**	**Cervical**	**Colon**	**Lung**
Age	1,08*	1,04*	1,09*	1,02*	1,05*	1,05*
HT	0,64*	0,44*	0,55*	0,28*	0,47*	0,39*
Hyperlipidaemia	1,17*	0,87*	1,95*	0,30*	0,92	0,81*
VET	0,94	0,69*	0,76*	0,68*	0,80*	0,73*
Angina	0,86*	0,52*	0,66*	0,42*	0,76*	1,00
Heart failure	0,59*	0,69*	0,55*	0,64*	0,71*	0,92
Ischaemia	1,12*	0,56*	1,04	0,17*	1,02	0,95
Legend	*p<0.05					

Table 4 indicates that the risk factors for prostate cancer are age, hyperlipidaemia, and ischaemia, the risk factor for breast cancer was age, the risk factors for skin neoplasms are age and hyperlipidaemia, the risk factors for cervical cancer is age, the risk factors for colon cancer are age, and ischaemia, and the risk factors for lung cancer is age. Age was included *a priori* and is a risk factor for all cancer subtypes. Hyperlipidaemia and ischaemia are the most common risk factors for 2 of the top 6 cancer subtypes.

Table 5: Stepwise regression		
Cancer	**Model 1**	**Model 2**
Age	1,05*	1,05*
CVD	0,28*	
T2D	0,56*	0,60*
Obesity	0,58*	0,68*
Asthma	0,54*	0,61*
COPD	1,19*	1,19*
Gender	1,35*	1,40*
HT		0,38*
VET		0,77*
Angina		0,55*
Heart failure		0,71*
Ischaemia		0,85*
Arrhythmia		0,77*
Cardiomyopathy		0,61*
Atherosclerosis		0,72*
Hyperlipidaemia		0,86*
Legend	*p < 0,05	

Variables generated from both the forward and backward stepwise regression analysis indicated that in model 1 where we used collective CVD, the risk factors for cancer were age, gender and COPD, and model 2 where we use CVD subtypes, again the risk factors for cancer were age, gender, COPD.

Table 6: The final model	
Cancer	**Model E1**
Age	1,05*
Gender	1,41*
T2D	0,22*
Obesity	0,61*
COPD	1,09*
HT	0,22*
Hyperlipidaemia	0,38*
VET	0,74*
Angina	0,39*
Heart failure	0,46*
Ischaemia	0,76*
Arrhythmia	0,52*
Cardiomyopathy	0,45*
Atherosclerosis	0,53*
T2D#HT	4,02*
T2D#Angina	1,86*
T2D#hyperlipidaemia	0,93*
T2D#heart_failure	1,78*
T2D#Ischaemia	1,47*
T2D#arrhythmia	1,54*
T2D#cardiomyopathy	1,78*
T2D#atherosclerosis	1,57*
HT#hyperlipidaemia	3,44*
HT#VET	1,27*
HT#angina	2,18*
HT#heart_failure	2,10*
HT#ischaemia	1,46*
HT#arrhythmia	2,78*
HT#cardiomyopathy	2,00*
HT#atherosclerosis	1,72*
Legend	*p < 0,05

Table 6 shows the models with variables selected from stepwise regression and it also includes interaction variables. T2D and HT were used as the common variables for interaction with all the other CVD subtype variables because it is common in literature that both T2D and HT are risk factors for CVD and cancer, both T2D and HT are

associated with an increasing age which is an identified risk factor for cancer. Upon analysis of the models with interactions we found that in model E1 while age, gender and COPD are the only variables that increase the risk for cancer, the interaction between T2D and HT, angina, HF, ischaemia, arrhythmia, cardiomyopathy, and atherosclerosis all increase the risk for cancer development. The interaction between HT and VET, angina, HF, ischaemia, arrhythmia, cardiomyopathy, and atherosclerosis all increase the risk for cancer development.

3.4. Time-to-event analysis

Kaplan Meier – Time to event analysis

The table below indicates that 220948 observations had both CVD and cancer

Table 7: 2x2 table showing cancer patients with CVD			
	Cancer	No cancer	Total
CVD	220,948 (13.53)	1,411,520(86.47)	1,632,468(88.16)
No CVD	42,742 (19.50)	176,405 (80.50)	219,147 (11.84)
Total	263,690 (14.24)	1,587,925(85.76)	1,851,615 (100)

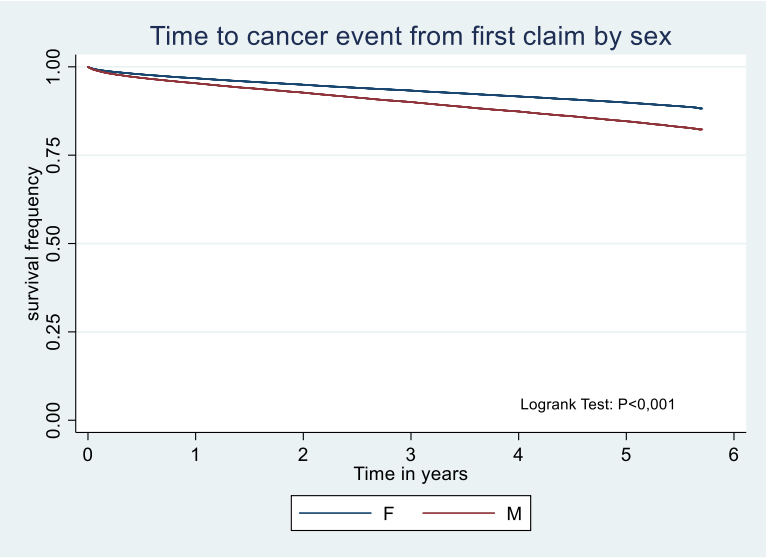


Figure 1: Time to cancer occurrence from the first claim in the total study population

In the above time to event analysis, the time interval was measured from the first claim to the last claim, with the follow up time starting from the time the first claim was made and the ending with the date that a cancer claim was made, in the period between 2014 and 2018 inclusive. In this analysis, there were 1851615 total observations

representing 722934 unique beneficiaries. After removing the errors, 1027212 observations remained, representing 566484 subjects with 44859 cancer (failure) events.

The Kaplan Meier (KM) Curve is a non-parametric statistic measurement of the survival function or time to event function from lifetime data. In this study we use the KM curve to show the time to cancer incidence. The log-rank test is used to compare the time-to-event (cancer) in 2 samples of the population, in this case the comparison was between males and females in the same study population. In the study population, males had more cancer than females and the log-rank test value is significant ($p < 0.001$).

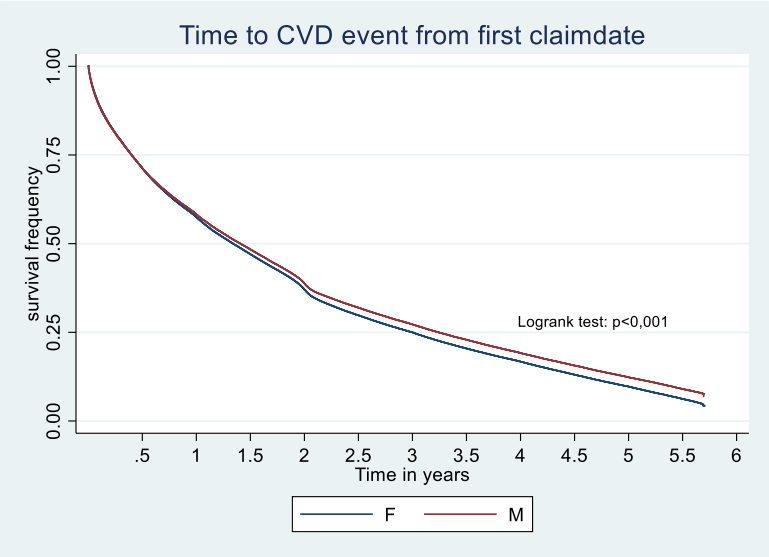


Figure 2: Time to CVD occurrence from the first claim in the total study population

In the above time to event analysis, the time interval was measured from the first claim to the last claim, with the follow up time starting from the time the first claim was made and the ending with the date that a CVD claim was made, in the period between 2014 and 2018 inclusive. The above KM curve shows that more females had CVD than males and the log-rank test was statistically significant ($p<0.001$).

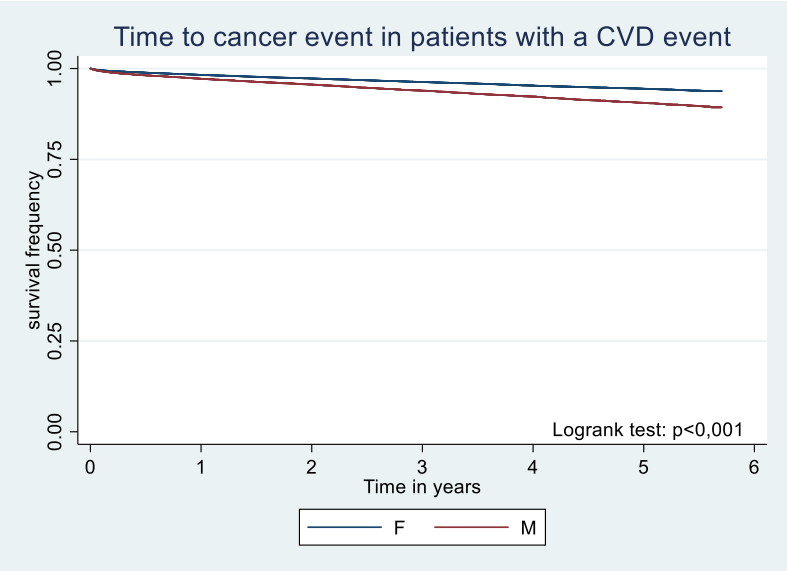


Figure 3: Time to cancer occurrence in patients with CVD

In the above time-to-event analysis only patients with CVD were analysed, the time interval was measured from the first claim to the last claim, with the follow up time starting from the time the first CVD claim was made and the ending with the date that a cancer claim was made, in the period between 2014 and 2018 inclusive. in the period between 2014 and 2018 inclusive. In this data set, there were 1851615 total observations representing 722934 unique beneficiaries. After removing the errors, 194154 observations remained, representing 161268 subjects with 5842 cancer (failure) events.

The KM curve shows that in the CVD population, more males developed cancer than females and the log-rank test is statistically significant ($p<0.001$).

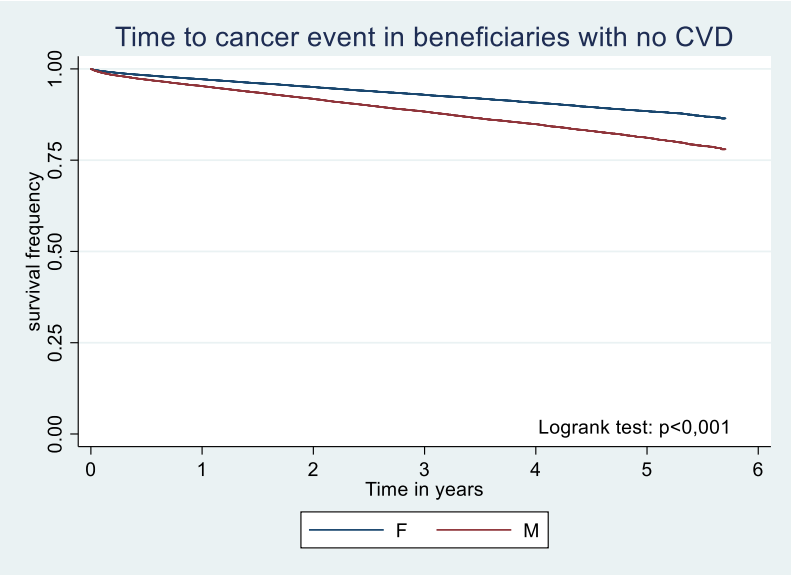


Figure 4: Time to cancer occurrence in patients without CVD

In the above time to event analysis only patients without CVD were analysed, the time interval was measured from the first claim to the last claim, with the follow up time starting from the time the first claim was made and the ending with the date that a cancer claim was made, in the period between 2014 and 2018 inclusive. In this data set, there were 1851615 total observations representing 722934 unique beneficiaries. After removing the errors, 284,309 observations remained, representing 185,930 subjects with 13425 cancer (failure) events among patients who did not claim for a CVD event.

The KM curve shows that in the population with no CVD, more males developed cancer than females and the log-rank test is statistically significant ($p<0.001$).

Table 8: Cancer incidence rate (per 100 000) from time to event analysis populations

	Total study population	Patients with CVD	Patients without CVD
Incidence rate	0.027	0.016	0.033
Failure events	44859	5842	13425
Total subjects	566484	161268	185930

The above incidence analysis shows that the incidence of developing cancer among all the observations is 0,027 per 100 1000 people with 44859 failure events from all patients in the population who claimed for any type of condition. The incidence of developing cancer among patients with CVD is 0,0216 per 100 1000 people with 5842 failure events. The incidence of developing cancer among patients without CVD is 0,033 per 100 1000 people with 13425 failure events.

3.5. Establishing cancer risk factors

Table 9: Generalised Linear Model Regression – Risk factors for Cancer outcome						
	Cancer risk factors: all beneficiaries		Cancer risk factors: beneficiaries with CVD		Cancer risk factors: beneficiaries without CVD	
	Model F3	Model F5	Model G3	Model G5	Model H3	Model H5
CVD	0,24*		1		1,32*	
T2D	0,62*	0,67*	1,38*	1,37*	0,93*	1,01
Obesity	0,81*	0,86*	1,14	1,06	1,26*	1,35*
Asthma	0,52*	0,57*	1,16*	1,04	0,91*	0,92*
COPD	1,1*	1,05	1,4*	1,1	1,01	0,99
Age	1,04*	1,04*	1,04*	1,04*	1,05*	1,05*
Gender	1,49*	1,55*	1,8*	1,70*	1,88*	1,84*
VET		0,7*		0,81*		1,1*
Angina		0,57*		0,83*		0,86*
Heart failure		0,7*		0,85*		0,91*
Ischaemia		0,81*		0,81*		1,12*
Arrhythmia		0,69*		1,05		0,93*
Cardiomyopathy		0,57*		0,64*		0,71*
Atherosclerosis		0,64*		0,81		0,68*
Hyperlipidaemia		0,78*		1,47*		1,14*
HT		0,31*		0,51		1,12*
Legend	*p<0,05					

In the above table 9 for GLM regression, the analysis indicates that the risk factors for cancer in the entire study population were age, COPD, gender (F3 and F5). The risk factors for cancer in patients who have CVD are, age, gender, COPD, asthma and T2D, and hyperlipidaemia (G3 and G5). The risk factors for cancer in patients with no CVD are age, gender, CVD, and obesity (H3) and age, gender, obesity, VET, ischaemia, hyperlipidaemia, and HT (H5).

Chapter 4: Discussion

This chapter presents a summary of the principal study findings and discusses the strengths and weaknesses of the study, comparison of study's strengths, weaknesses and findings with previous studies, the meaning of the study to clinicians and policy makers, research gaps and unanswered questions. This chapter is completed with a summary on recommendations for further research and the conclusion.

4.1. Principal findings

The study population comprised 64,48% female and 35,52% male beneficiaries, 9,17% of the study population had cancer, with 10,81 % of male patients and 8,26% of the female patients presenting with cancer, 79,31% of all patients had CVD, 75,89% of the male patients had CVD and 81,19% of the female patients had CVD. The top 6 cancer subtypes in descending order were breast cancer, skin neoplasms, prostate cancer, colon cancer, cervical cancer, and lung cancer. The top 6 CVD subtypes in descending order are hypertension, hyperlipidaemia, VET, angina, heart failure and ischaemia. The prevalence of T2D in the study population is 23,07%.

The univariate logistic regression analysis indicates that age, gender, COPD, heart failure, ischaemia, arrhythmia, atherosclerosis, and hyperlipidaemia are statistically significant risk factors for the development of cancer. CVD collectively, is not a risk factor for the development of cancer. The multivariate analysis and stepwise regression indicate that only age, gender, and COPD are statistically significant risk factors for cancer.

Analysis between the top 6 cancers and top 6 CVD subtypes indicate that age is a risk factor for all cancer subtypes, while hyperlipidaemia is a risk factor for 2 of the top 6 cancer subtypes namely, prostate cancer and skin neoplasms, and ischaemia is also a risk factor for 2 of the top 6 cancers namely, prostate and colon cancer.

In the final model, age, gender and COPD are shown to increase the risk for cancer, the interactions between T2D and HT, angina, HF, ischaemia, arrhythmia, cardiomyopathy and atherosclerosis increase the risk for cancer development, and the interactions between HT and VET, angina, HF, ischaemia, arrhythmia, cardiomyopathy and atherosclerosis increase the risk for cancer development.

The generalised linear model analysis in table 9 indicates that the risk factors for cancer in the entire study population is age, COPD, and gender. The risk factors for cancer in patients who have CVD are, age, gender, COPD, asthma, T2D and hyperlipidaemia. The risk factors for cancer in patients with no CVD are age, gender, CVD, obesity, VET, ischaemia, hyperlipidaemia, and HT.

4.2. Strengths and weaknesses

Data utilised for this study was restricted to GEMS data which is not representative of the South African adult population. Only 16% of the South African population are medically insured and GEMS contributes to less than 50% of the medically insured patients (83). GEMS being a government workers medical insurance does not include members from the South African private sector or self-employed individuals and may not be representative of the South African population (83). Potential bias may be present in this study because this only included patient claims from patients who have existing funds in their medical insurance which may lead to underrepresentation of the disease conditions. Patients whose medical aids are exhausted or who have recently found themselves unemployed would have been unable to claim via medical aid and this data would not have been recorded or considered in the analysis.

The data analysed in this study were routinely collected from pharmacy billings against the medical insurance, hospital records when patients claimed for medical treatment, and other medical claims completed at healthcare facilities that accommodate for medical insurance. This data collection may have inaccuracies and lack detail about this study's subject matter. The inaccuracies may have resulted from reporting errors from healthcare facilities, administrative errors, technological malfunctions, incomplete claim requests or incorrect information provided by the patient to their healthcare provider. The data did not contain information on confounders and interaction variables such as smoking status, weight, diet, exercise, ethnicity, alcohol, and drug use. Controlling for these confounders was therefore not possible as this information is not routinely collected in medical claims data. Acquiring the data was delayed due to approval and de-identification processes. The data contained outliers, duplicated records and records with missing information that were removed in the data management stage.

The study population comprised 64,48% female beneficiaries and 35,52% male beneficiaries, male beneficiaries were under-represented in this study cohort and this may be due to male patients not seeking medical care timeously, therefore medical records for treatment claims were less for male patients compared to females [52, 53].

The data utilised was routinely collected data by GEMS from all participating healthcare facilities, therefore, data would have captured on every visit to the pharmacy, laboratory, or health facility for auditing purposes and this implies that the data collected was reliable and provided an accurate indication of the magnitude of the claims processed. The study population was clearly defined, and the medical aid data allowed for a focused study analysis strictly on NCDs in adults above the age of 18, between from 2014 to 2018 inclusive. The GEMS data comprises of government workers and their immediate family and the population of the GEMS data set are of a higher socio-economic status than the general population, thus these medically insured patients may visit a doctor or healthcare facility whenever they deem necessary, and these visits and medical claims are all recorded in a routine database. The routinely collected data needed to be timeously compiled for auditing purposes and this ensured the data worked with was complete. The dataset was large and representative of a large population of patients which may allow for good generalisability for government workers.

4.3. Strengths and weaknesses in relation to other studies

Age being a risk factor for cancer is highlighted in this study and is supported by literature which indicates that age is indeed a risk factor for the development of cancer [54]. Aging is not an illness, but this naturally occurring process is associated with age-related illnesses such as cancer, metabolic diseases, and decreased quality of life. An increasing life expectancy allows for cancer development in the aging population as biological processes of ageing with increased inflammation and a weakened immune system allow for cancer development [54]. Aging is both a physiological and chronological process, and cellular damage along with genetic alterations that accompany the physiological and chronological passage of time, is a driver for both aging and cancer development [55].

Gender is shown to be risk factor for cancer development in both the univariate and multivariate analysis as well as the stepwise regression and GLM regression. The analysis reveals that males have 1,35 times the odds of developing cancer than females in the univariate regression, and the multivariate regression reveals that males have 1,41 times the odds of developing cancer than females. In this analysis, cancer is the collective cancer of all the cancer subtypes, and, in this population, male beneficiaries have a higher cancer prevalence than female beneficiaries, which may be the reason that males have a higher likelihood of developing cancer than females. In this population, male beneficiaries also have a higher prevalence for skin neoplasms, colon cancer and lung cancer. The gender findings in this study are aligned with literature as the South African NCR indicates that 1 in out of 6 men have a lifetime risk of developing lung, prostate, bladder, Non-Hodgkin's Lymphoma or colorectal cancer and 1 out of 7 women have a lifetime risk of developing breast, cervical, colorectal, lung or uterine cancer (38). The Italian cancer registry indicates that 50% of males and 33% of females have a risk of developing cancer in their lifetime (56). Females having 2 "X" chromosomes possess an advantage with heterogeneity since there will be variations in gene expression between the 2 X chromosomes as compared to only one X chromosome in males, this heterogeneity is beneficial with regards to immune response (57). Females have protective hormones such as progesterone and oestradiol, which have anti-inflammatory effects and positive cell-mediated immune responses, respectively. Conversely, males have the androgen hormone which suppresses the immune cell activity, and testosterone that suppresses the formation of antibodies and increases inflammation (58, 59).

COPD is only present in 1,34% of the population with more males (1,88%) presenting than females (1,04%). The low prevalence of COPD does not alter or bias the association with cancer development, the univariate analysis indicates that patients with COPD have 2,39 times the odds of developing cancer and 8 times the odds of developing lung cancer than patients with no COPD. As the length of smoking increases, the more susceptible individuals are to COPD, smoking from a young age until an older age and the non-cessation of smoking leads to a greater likelihood of developing COPD (60).

Tobacco smoking is the most common risk factor for COPD and may cause a 3 to 5-fold increase in the likelihood of developing COPD compared to non-smokers (60).

South African males show a higher prevalence of smoking than South African females, literature indicates the prevalence of smoking in males is 4 times the prevalence of smoking in females. Males smoking 4 times more than females and smoking being an identifiable risk factor for COPD may be the reason that more males have COPD than females in this dataset (61, 62). Tobacco smoking is an identifiable risk factor for at least 15 types of cancer including lung cancer, cancer of the larynx, pharynx, oesophagus and oral cavity, bladder, liver, stomach, ovary, cervix, kidney, pancreas, and colorectal cancer. Tobacco smoking may have a greater than 100-fold increase the likelihood of lung cancer compared to non-smokers and may increase the likelihood of cancer in all the oral cavities by between 3 and 13-fold (63).

Globally, COPD and lung cancer are the leading causes of illness and mortality. Oxidative stress resulting in DNA damage, chronic inflammation, genetic alterations, and tobacco smoking are common risk factors and drivers for both COPD and lung cancer. Increasing COPD severity has a positive linear relationship with lung cancer (64, 65). Air pollution and the vapours inhaled from cooking oil and other household chemicals are common risk factors between COPD and lung cancer (66, 67).
Tobacco smoking is associated with cancer outcomes as well as COPD outcomes and may be a possible confounder in the relationship between COPD and cancer. The increase in lung cancer cases among patients with COPD is a health concern because the exact mechanism for this causal relationship is not fully understood, however, genetic alterations, lung microbiome and irregularities of the immune system may be mediators in this relationship (64).

In this study, both the univariate analysis and multivariate analysis, indicate that being obese does not increase the odds of developing cancer. These findings are not supported by literature, the primary reason may be due to waist-hip ratios, BMI, and body fat percentage data not being requested or being inaccurately recorded by healthcare professionals when patients are being examined. The non-existence of a causal relationship between obesity and cancer in this study may be due to non-differential misclassification bias which is directing the measure of effect toward the null indicating that there is no relationship between the two variables. In studies that measure the association between risk factors and certain diseases, misclassification bias, a type of information bias, may alter the true effect of the association. The association may be under or overrepresented in these cases of bias.

The incidence of cancer will increase over time due to a rise in obesity and metabolic diseases as obesity is a recognised risk factor for at least 13 malignancies comprising renal and pancreatic adenocarcinomas, endometrial, oesophageal, and hepatocellular carcinoma; colorectal, postmenopausal, gastric cancers; multiple myeloma; meningioma; ovarian, breast, gallbladder, and thyroid cancers (68, 69, 70).
The growth and spread of white adipose tissue known as ectopic fat deposition resulting from obesity, causes DNA alterations via metabolic, inflammatory, and immunological changes. These DNA alterations leading to genetic mutations allow for cancer development (70). Childhood and early adulthood obesity is associated with an increased risk of pancreatic cancer, colon cancer and multiple myeloma (71, 72). The global burden of cancer attributed to obesity is 13,1% in females and 11,9% in males, for obesity-related cancers (69).

The final model in the analysis included interaction variables, T2D and HT were the common variables that were used in the interactions with the other CVD subtypes. The reason for selecting T2D and HT is because of the high prevalence of these conditions in the study population (23,07% and 63,88% respectively) and the widespread recognition of the association between T2D and HT being a risk factor for CVD and cancer development (46).

There is evidence to support that T2D increases the risk for the development of liver, pancreatic, endometrial, breast, kidney, bladder, and non-Hodgkin lymphoma (73). T2D has been shown to cause an increasing risk for pancreatic, oesophageal, and biliary-tract cancer in males; endometrial and breast cancer is females, and colorectal, liver and kidney cancer in both males and females (74, 75). In 2010, the American Diabetes Association (ADA) and American Cancer Society (ACS) published a report on risk factors linking T2D to cancer, the risk factors were categorised as biological links (hyperinsulinemia, hyperglycaemia, and fat-induced chronic inflammation), modifiable risks (overweight, obesity, physical activity, excessive alcohol consumption and smoking) and non-modifiable risks (age sex and ethnicity/race) (73, 76). Cancer mortality is higher in T2D patients than patients who do not have T2D, for a range of malignancies (74,75). Antidiabetic medication such as metformin is documented to

decrease the risk of tumour development in T2D patients and decrease the risk of cancer development (73).

In various meta- analyses comprising more than 10 individual studies, hypertension was associated with a 23% increased risk of cancer death, with liver cancer and rectal cancer responsible for majority of the cancer deaths. Studies also show that elevated blood pressure levels are positively associated with both cancer incidence and cancer mortality (77, 78). The risk of cancer increased linearly with an increase in blood pressure among patients studied in the meta-analysis. Several studies analysing the relationship between high blood pressure and cancer have shown that hypertension is a risk factor for the development of renal cancer (79, 80).

A meta-analysis of observational studies included 13 studies where the association between elevated blood pressure and cancer was analysed. The analysis showed a positive relationship between elevated blood pressure levels and the risk of renal, liver, oesophageal, breast, colorectal and endometrial cancer (81). Many studies indicate a positive relationship between hypertension and renal cancer, and this meta-analysis support these findings as it showed a 54% higher risk of developing renal cancer if patients have hypertension (81). This meta-analysis also indicated an 11% increased risk of colorectal cancer in hypertensive individuals. The causative link between hypertension and cancer remains unclear, however hypertension may promote the cessation of apoptosis which may result in an increased risk for malignancies (81, 82).

4.4. Meaning of the study: possible mechanisms and implications for clinicians or policymakers

The prevalence of CVD and cancer in this study population allowed for their association to be explored via regression analysis, to understand the relationship between these two groups of NCDs that share many risk factors.

The data analysis indicates that along with age and COPD, there are various other CVD subtypes that increase the risk for developing cancer. These findings imply that controlling for these risk factors and early treatment of these risk factors to prevent further progression, may possibly delay the onset and progression of cancer.

Historically, literature has shown that age is a risk factor for both CVD and cancer development, and this association is highlighted in this study population. This association reinforces the need for clinicians to conduct proper cancer and CVD screening in patients as they age, and to advise older patients on proper lifestyle management (diet and exercise) to maintain a healthy status and low inflammation levels to delay or prevent the onset of these NCDs.

The analysis on gender being a risk factor for different cancer types allows for clinicians to screen for certain malignancies in males and females based on their existing risk factors.

COPD is shown to be a risk factor for cancer and further exploration showed that COPD is a risk factor for lung cancer with individuals having 8 times the odds for developing lung cancer if one has COPD as opposed to not having COPD. This information implies that clinicians should conduct early screening for cancer in COPD patients. Tobacco smoking is a risk factor for lung cancer as well as COPD and may be both a confounder and a mediator in this association. Clinicians may use this information to advise patients on the risk of smoking in general and the added risk in patients with COPD.

Patients with T2D often present with HT and vice versa and interactions with these two conditions and other CVD subtypes increase the odds of developing cancer. This information will enable clinicians to suggest preventative approaches for T2D and HT which may minimise the interaction between these conditions and prevent cardiovascular damage and possibly prevent the onset and progression of cancer. Hyperlipidaemia is often an indicator of obesity or being overweight, which assists in clinicians making recommendations on the risks of obesity when obesity data is unavailable or not recorded.

The above information will possibly aid in the South African healthcare efforts to reduce the onset, progression and burden of CVD and cancer. The potential reduction in CVD and cancer burden may reduce the burden on our healthcare system by reducing medical cost to patients, caregivers, healthcare centres and facilities. This

study will reinforce the need to increase awareness of these NCDs and the association between them in both the public and private health sectors.

4.5. Unanswered questions and future research

ET indicates the increasing importance of controlling for and preventing NCDs, registries such as the South African NCR are valuable in supporting the study and analysis of the NCD trends (16). Registries on other NCDs such as CVDs, diabetes and respiratory conditions do not exist, and this may delay the investigation into these disease trends and associations (18). South Africa needs to create registries for these NCDs that have a high burden on the population and the health systems, the passing of the National Public Health Institute of South Africa (NAPHISA) bill will support the need and generation of these registries.

The analysis on this study population exposed the variables that may increase the risk for cancer development, but a few unanswered questions and inconclusive findings remained. This study revealed a need for precise obesity diagnosis and tracking of obesity by standardised measurements or markers, for studies to be conducted on the associations with obesity. GEMS comprises a fraction of medically insured South Africans who are in turn a smaller fraction of the South African population, further research is essential using all medical insurance databases to assess the associations between CVD and cancer in the private sector. Public sector data on these conditions also require investigation and analysis, and comparisons may be conducted on the association between CVD and cancer in the public sector and the private sector, as well as national population level research.

More data is needed to control for possible confounding variables and effect modifiers to stratify the analysis for potential confounders such as tobacco and alcohol use, diet and physical activity levels and other underlying conditions.

A prospective cohort study in a large sample of patients over many years will help to understand the timeframe of cancer onset in patients with CVD, T2D and COPD, a study analysed over a lengthy duration may better explain the association between CVD and cancer

4.6. Overall conclusions and recommendations

Collectively CVD is not a risk factor for the development of cancer, however, in this study population, CVD subtypes such as hyperlipidaemia and ischaemia increase the risk for cancer development. Interactions between CVD subtypes with T2D an HT, increase the risk for cancer. In this study population, age and gender are risk factors for cancer and this information is aligned with historic literature. Although COPD only comprised 1,34% of this study population, regression analysis indicated that COPD is a risk factor for cancer and may be responsible for an 8-fold increase in the likelihood of developing lung cancer. Early treatment and controlling the onset and progression of the CVD subtypes and T2D may delay the onset and progression of cancer. Awareness around the risks and dangers of tobacco smoking may decrease the incidence and severity of COPD which may result in a decrease in lung cancer. Further data collection and research is required to be conducted on CVD subtypes and their association with cancer and the common risk factors between CVD and cancer.

.

References:

1. South Africa WHO Fact Sheet. 2018;2018. Available from: https://www.who.int/nmh/countries/zaf_en.pdf
2. Juma K, Juma PA, Mohamed SF, Owuor J, Wanyoike A, Mulabi D, et al. First Africa non-communicable disease research conference 2017: sharing evidence and identifying research priorities. J Glob Health. 2019;8(2):1–13.
3. Heller O, Somerville C, Suggs LS, Lachat S, Piper J, Pastrana NA, et al. The process of prioritization of non-communicable diseases in the global health policy arena. 2019;(June):370–83.
4. Habib A, Alam MM, Hussain I, Nasir N, Almuthebi M. Erratic Behavioral Attitude Leads to Noncommunicable Diseases: A Cross-Sectional Study. 2020;2020(2014).
5. Lelieveld J, Klingmu K, Pozzer A, Po U, Fnais M, Daiber A, et al. Cardiovascular disease burden from ambient air pollution in Europe reassessed using novel hazard ratio functions. 2019;1590–6.
6. Allen L, Williams J, Townsend N, Mikkelsen B, Roberts N, Foster C, et al. Socioeconomic status and non-communicable disease behavioural risk factors in low-income and lower-middle-income countries: a systematic review. Lancet Glob Health 2017;5(3):e277–89.
7. Gouda HN, Charlson F, Sorsdahl K, Ahmadzada S, Ferrari AJ, Erskine H, et al. Articles Burden of non-communicable diseases in sub-Saharan Africa, 1990 – 2017: results from the Global Burden of Disease Study. Lancet Glob Health 2017;7(10): e1375–87. Available from: http://dx.doi.org/10.1016/S2214-109X(19)30374-2
8. Bennett JE, Stevens GA, Mathers CD, Bonita R, Rehm J, Kruk ME, et al. NCD Countdown 2030: worldwide trends in non-communicable disease mortality and progress towards Sustainable Development Goal target. Lancet 2018; 392:1072-88.
9. Benziger CP, Roth GA, Moran AE. The Global Burden of Disease Study and the Preventable Burden of NCD. Glob Heart 2016;11(4):393–7. Available from: http://dx.doi.org/10.1016/j.gheart.2016.10.024
10. WHO Fact Sheet. 2018;2018. Available from: https://www.who.int/nmh.pdf
11. Nojilana B, Bradshaw D, Pillay-van Wyk V, Msemburi W, Somdyala N, Joubert JD, Groenewald P. Persistent burden from non-communicable diseases in South Africa needs strong action. S Afr Med J 2016;106(5):436-437.
12. Hunter DJ and Reddy KS. Noncommunicable diseases. New England journal of medicine. N Engl J Med 2013; 369:1336-43.
13. World Health Organization - Noncommunicable Diseases (NCD) Country Profiles, 2018.
14. Keates AK, Mocumbi AO, Ntsekhe M, Sliwa K, Stewart S. Cardiovascular disease in Africa: Epidemiological profile and challenges. Nat Rev Cardiology, 2017;14(5):273–93.
15. Reeve B, Gostin LO. “Big” Food, Tobacco, and Alcohol: Reducing Industry Influence on Noncommunicable Disease Prevention Laws and Policies Comment on “Addressing NCDs: Challenges From Industry Market Promotion and Interferences.” Kerman Univ Med Sci [Internet]. 2019;8(7):450–4. Available from: https://doi.org/10.15171/ijhpm.2019.30
16. Nojilana B, Bradshaw D, Pillay-van Wyk V, Msemburi W, Laubscher R, Somdyala NIM, et al. Emerging trends in non-communicable disease mortality in South Africa, 1997 - 2010. South African Med J. 2016;106(5):477.
17. Gaye B, Diop M, Narayanan K, Offredo L, Reese P, Antignac M, et al. Epidemiological transition in morbidity : 10-year data from emergency consultations in Dakar , Senegal. 2019;1–7.
18. Sharman M, Bachmann M. Prevalence and health effects of communicable and non- communicable disease comorbidity in rural KwaZulu-Natal, South Africa. 2019;24(10):1198–207.
19. Kabudula CW, Houle B, Collinson MA, Kahn K, Gómez-Olivé FX, Clark SJ, et al. Progression of the epidemiological transition in a rural South African setting: findings from population surveillance in Agincourt, 1993-2013. BMC Public Health. 2017;17(1):1–15.
20. Barthelemy KD. Demographic, epidemiological, and health transitions: Are they relevant to population health patterns in Africa? Glob Health Action. 2014;7(1):17–55.
21. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, et al. Heart Disease and Stroke Statistics—2019 Update: A Report from the American Heart Association. Vol. 139, Circulation. 2019.
22. Chikafu H, Chimbari MJ. Cardiovascular Disease Healthcare Utilization in Sub-Saharan Africa: A Scoping Review. 2019;
23. Sliwa K, Ntusi N. Battling Cardiovascular Diseases in a Perfect Storm. Circulation. 2019;139:1658–1660. DOI: 10.1161/CIRCULATIONAHA
24. Yuyun MF, Sliwa K, Kengne AP, Olga A, Bukhman G. Cardiovascular Diseases in Sub-Saharan Africa Compared to High-Income Countries : An Epidemiological Perspective. 2020;15(1):1–18.
25. Statistics South Africa, Morality and causes of death in South Africa, 2016: Findings from death notification; Statistical Release P0309.3, February 2019
26. Cardiovascular disease: Heart and Stroke Foundation South Africa Cardiovascular Disease Statistics 2019.
27. Scott M. Grundy, Obesity, Metabolic Syndrome, and Cardiovascular Disease, *The Journal of Clinical Endocrinology & Metabolism*, Volume 89, Issue 6, 1 June 2004, Pages 2595–2600

28. Basu S, Wagner RG, Sewpaul R, Reddy P, Davies J. Articles Implications of scaling up cardiovascular disease treatment in South Africa: a microsimulation and cost-effectiveness analysis. Lancet Glob Heal [Internet]. 2019;7(2): e270–80. Available from: http://dx.doi.org/10.1016/S2214-109X(18)30450-9
29. Wild CP. Comment The global cancer burden: necessity is the mother of prevention. Nat Rev Cancer [Internet]. Available from: http://dx.doi.org/10.1038/s41568-019-0110-3
30. Global Burden of Disease Cancer Collaboration. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived with Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2016: A Systematic Analysis for the Global Burden of Disease Study. JAMA oncology. 2018 Nov 1;4(11):1553-1568.
31. Martel C De, Georges D, Bray F, Ferlay J, Clifford GM. Articles Global burden of cancer attributable to infections in 2018 : a worldwide incidence analysis. Lancet Glob Heal [Internet]. 2020;8(2):e180–90. Available from: http://dx.doi.org/10.1016/S2214-109X(19)30488-7
32. Sayiner M, Golabi P, Younossi ZM. Disease Burden of Hepatocellular Carcinoma : A Global Perspective. Dig Dis Sci [Internet]. 2019;(0123456789). Available from: https://doi.org/10.1007/s10620-019-05537-2
33. Somdyala NIM, Bradshaw D, Dhansay MA, Stefan DC, Supplement D. Increasing Cervical Cancer Incidence in Rural Eastern Cape Province of South Africa From 1998 to 2012 : A Population-Based Cancer Registry Study abstract. 2020;1–8.
34. Jemal A, Bray F, Forman D, O'Brien M, Ferlay J, Center M, et al. Cancer burden in Africa and opportunities for prevention. Cancer. 2012;118(18):4372–84.
35. International Agency for Research on Cancer (IARC). GLOBOCAN 2018: Latest global cancer data. CA Cancer J Clin. 2018;68(September):13–5.
36. Dent J, Manner CK, Milner D, Mutebi M, Ng A, Olopade OI, et al. Africa' s Emerging Cancer Crisis : A Call to Action. 2017;(June):1–8.
37. Parkin DM, Hämmerl L, Ferlay J, Kantelhardt EJ. Cancer in Africa 2018 The role of infections. 2019;00(May):1-15.
38. The Cancer Association of South Africa, National Cancer Registry report 2016
39. Moodley J, Stefan DC, Sewram V, Ruff P, Freeman M, Asante-Shongwe K. An overview of cancer research in South African academic and research institutions, 2013 - 2014. South African Med J. 2016;106(6):607.
40. Demetriou G. Commentary Can Endocrine Therapy Reduce the Risk of Breast Cancer in South African Women? 2020;2(1):39–41.
41. Maphayi MR, Cassim N, Bello B, George JA. Mining laboratory data to describe prostate specific antigen testing and prostate cancer in Johannesburg, South Africa. Pan African Medical Journal 2020;8688:1–10.
42. Danaei G, Hoorn SV, Lopez AD, Murray CJL, Ezzati M. Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors, Lancet 2005; 366: 1784–93
43. Minihane AM, Vinoy S, Russell WR, Baka A, Roche HM, Tuohy KM, et al. Low-grade inflammation, diet composition and health: Current research evidence and its translation. Br J Nutr. 2015;114(7):999–1012.
44. Koene RJ, Prizment AE, Blaes A, Konety SH. Shared Risk Factors in Cardiovascular Disease and Cancer. Circulation 2016;133(11):1104–14.
45. Shank C., DNP, APRN, ANP-BC J. Cardiovascular and Cancer Risk: The Role of Cardio-oncology. J Adv Pract Oncol. 2018;9(2):160–76.
46. Meijers WC, De Boer RA. Common risk factors for heart failure and cancer. Cardiovasc Res. 2019;115(5):844–53.
47. Iqbal K, Schwingshackl L, Floegel A, Schwedhelm C, Mardas MS. Gaussian graphical models identified food intake networks and risk of type 2 diabetes , CVD , and cancer in the EPIC-Potsdam study. Eur J Nutr [Internet]. 2018;0(0):0. Available from: http://dx.doi.org/10.1007/s00394-018-1714-1
48. Fedewa SA, Sauer AG, Shuval K. Global Patterns in Excess Body Weight and the Associated Cancer Burden. 2019;88–112. doi: 10.3322/caac.21499. Available online at cacancerjournal.com
49. Pontiroli AE, Ceriani V, Sarro G, Micheletto G, Giovanelli A, Zakaria AS, et al. Incidence of Diabetes Mellitus , Cardiovascular Diseases , and Cancer in Patients Undergoing Malabsorptive Surgery (Biliopancreatic Diversion and Biliointestinal Bypass) vs Medical Treatment. 2018;
50. Libby P, Kobold S. Inflammation : a common contributor to cancer , aging , and cardiovascular diseases — expanding the concept of cardio-oncology. 2019;824–9.
51. Royston P and Lambert P.C. Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model. Stata Press. Pages 47 -72.
52. Aboyade, O.M., Beauclair, R., Mbamalu, O.N. *et al.* Health-seeking behaviours of older black women living with non-communicable diseases in an urban township in South Africa. *BMC Complement Altern Med* **16,** 410 (2016). https://doi.org/10.1186/s12906-016-1378-4
53. vans, Meredith, & Olive Shisana. "Gender differences in public perceptions on National Health Insurance." South African Medical Journal [Online], 102.12 (2012): 918-924. Web. 24 Nov. 2020)
54. White MC, Holman DM, Boehm JE, Peipins LA, Grossman M, Henley SJ. Age and Cancer Risk: A Potentially Modifiable Relationship. Am J Prev Med. 2014 March ; 46(3 0 1): S7–15. doi:10.1016/j.amepre.2013.10.029.
55. Rozhok AI, Degregori J. The Evolution of Lifespan and Age-Dependent Cancer Risk. Trends in Cancer 2016;xx:1–9. Available from: http://dx.doi.org/10.1016/j.trecan.2016.09.004

56. Irelli A, Sirufo MM, Ugo CD, Ginaldi L, Martinis M De. Sex and Gender Influences on Cancer Immunotherapy Response. Biomedicines 2020;1–23.
57. Carrel, L.;Willard, H.F. X-inactivation profile reveal sex tensive variability in X-linked gene expression in females. Nature 2005, 434, 400–404.
58. Salem, M.L. Estrogen, a Double-Edged Sword: Modulation of TH1-And TH2-mediated Inflammations by Di_erential Regulation of TH1/TH2 Cytokine Production. Curr. Drug Targets Inflamm. Allergy. 2004, 3, 97–104.
59. Thompson, M.G.; Pei_er, D.S.; Larson, M.; Navarro, F.;Watkins, S.K. FOXO3, estrogen receptor alpha, and androgen receptor impact tumor growth rate and infiltration of dendritic cell subsets differentially between male and female mice. Cancer Immunol. Immunother. 2017, 66, 615–625.
60. Bai JW, Chen XX, Liu S, Yu L, Xu JF. Smoking cessation affects the natural history of COPD. Int J Chron Obstruct Pulmon Dis. 2017 Nov 16; 12:3323-3328. doi: 10.2147/COPD.S150243. PMID: 29180862; PMCID: PMC5695262
61. Vellios N, van Walbeek CDeterminants of regular smoking onset in South Africa using duration analysisBMJ Open 2016;6: e011076. doi: 10.1136/bmjopen-2016-011076
62. Reddy, P, Zuma, K, Shisana, O, Jonas, K, & Sewpaul, R. (2015). Prevalence of tobacco use among adults in South Africa: Results from the first South African National Health and Nutrition Examination Survey. SAMJ: South African Medical Journal, 105(8), 648-655. https://dx.doi.org/10.7196/SAMJNEW.7932
63. ludmil b. a. et al Mutational signatures associated with tobacco smoking in human cancer Science 04 Nov 2016: Vol. 354, Issue 6312, pp. 618-622 DOI: 10.1126/science.aag0299
64. Parris BA, Farrell HEO, Fong KM, Yang IA. Chronic obstructive pulmonary disease (COPD) and lung cancer: common pathways for pathogenesis. Journal of Thoracic Disease 2019;11(6):2155–72.
65. Mannino DM, Aguayo SM, Petty TL, et al. Low lung function and incident lung cancer in the United States: data From the First National Health and Nutrition Examination Survey follow-up. Arch Intern Med 2003;163:1475-80.
66. Kurmi OP, Lam KB, Ayres JG. Indoor air pollution and the lung in low- and medium-income countries. Eur Respir J 2012;40:239-54.
67. Lam WK. Lung cancer in Asian women-the environment and genes. Respirology 2005;10:408-17.
68. Avgerinos KI, Spyrou N, Mantzoros CS, Dalamaga M. Obesity and cancer risk : Emerging biological mechanisms and perspectives. Metabolism clinical and experimental 2019;92:121–35. Available from: https://doi.org/10.1016/j.metabol.2018.11.001
69. Dalamaga M, Christodoulatos GS, Mantzoros CS. The role of extracellular and intracellular
Nicotinamide phosphoribosyl-transferase in cancer: Diagnostic and therapeutic perspectives and challenges. Metabolism 2018;82:72–87.
70. Dalamaga M, Diakopoulos KN, Mantzoros CS. The role of adiponectin in cancer: a review of current evidence. Endocr Rev 2012;33:547–94.
71. Genkinger JM, Kitahara CM, Bernstein L, Berrington de Gonzalez A, Brotzman M, Elena JW, et al. Central adiposity, obesity during early adulthood, and pancreatic cancer mortality in a pooled analysis of cohort studies. Ann Oncol 2015;26: 2257–66.
72. Hofmann JN, Moore SC, Lim U, Park Y, Baris D, Hollenbeck AR, et al. Body mass index and physical activity at different ages and risk of multiple myeloma in the NIH-AARP diet and health study. Am J Epidemiol 2013;177:776–86.
73. Wojciechowska J, Krajewski W, Bolanowski M. Diabetes and Cancer : a Review of Current Knowledge. Exp Clin Endocrinol Diabetes 2016;263–75.
74. Colditz GA, Peterson LL. Reviews Obesity and Cancer : Evidence , Impact , and Future Directions. Clinical Chemistry 2018;162
75. Luo J, Chlebowski RT, Hendryx M, Rohan T, Wactawski-Wende J, Thomson CA, et al. Intentional weight loss and endometrial cancer risk. J Clin Oncol 2017;3: 1189 –93.
76. Gallagher EJ, LeRoith D. Diabetes, cancer, and metformin: connections of metabolism and cell proliferation. Ann N Y Acad Sci 2011; 1243: 54–68
77. Stocks T, Hemelrijck M Van, Manjer J, Bjørge T, Ulmer H, Hallmans G, et al. Blood Pressure and Risk of Cancer Incidence and Mortality in the Metabolic Syndrome and Cancer Project. Hypertension.2012;802–10.
78. Batty GD, Shipley MJ, Marmot MG, Davey Smith G. Blood pressure and site-specific cancer mortality: evidence from the original Whitehall Study. Br J Cancer. 2003;89:1243–1247.
79. Andreotti G et al. Variants in blood pressure genes and the risk of renal cell carcinoma. Carcinogenesis. 2010;31:614–620.
80. Corrao G, Scotti L, Bagnardi V, Sega R. Hypertension, antihypertensive therapy and renal-cell cancer: a meta-analysis. Curr Drug Saf. 2007;2: 125–133.
81. Seretis A, Cividini S, Markozannes G, Tseretopoulou X, Lopez DS, Ntzani EE, et al. Association between blood pressure and risk of cancer development : a systematic review and meta-analysis of observational studies. Scientific Report. 2019;(May):1–12. Available from: http://dx.doi.org/10.1038/s41598-019-45014-4
82. Hamet, P. Cancer and hypertension: a potential for crosstalk? Journal of hypertension 15, 1573–1577 (1997).
83. Statista South Africa 2018/2019 report
84. Gordon T., Booysen F. & Mbonigaba J. Socio-economic inequalities in the multiple dimensions of access to healthcare: the case of South Africa. BMC Public Health 20, 289(2020). https://doi.org/10.1186/s12889-020-8368-7
85. Özdemir BC, Dotto GP. Racial Differences in Cancer Susceptibility and Survival: More Than the Color of the Skin? Trends Cancer. 2017 Mar;3(3):181-197. doi: 10.1016/j.trecan.2017.02.002. Epub 2017 Mar 6. PMID: 28718431; PMCID: PMC5518637.

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG	HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

Office of the Deputy Vice-Chancellor (Research & Post Graduate Affairs)

TO: Ms A Ramadhar
School of Public Health
Division of Biostatistics and Epidemiology
Medical School
University

E-mail: anishkaramadhar@gmail.com

CC: Supervisor: Dr M Muchengeti <MazvitaM@nicd.ac.za>
and <HREC-Medical.ResearchOffice@wits.ac.za>

FROM: Iain Burns
Human Research Ethics Committee (Medical)
Tel: 011 717 1252

E-mail: Iain.Burns@wits.ac.za

DATE: 2020/02/19

REF: *R14/49*

PROTOCOL NO: **M1911142** *(This is your ethics application study reference number. Please quote this reference number in all correspondence relating to this study)*

PROJECT TITLE: *Investigating the association between cardiovascular diseases and cancer in South African adults from the GEMS database*

Please find attached the Clearance Certificate for the above project. I hope it goes well and that an article in a recognized publication comes out of it. This will reflect well on your professional standing and contribute to the Government funding of the University.

MSWorks2000/Iain0007/Clearscan.wps

WITS School of Public Health

Plagiarism declaration for written work: Master of Science in Epidemiology

I Anishka Ramadhar (Student number: 2244064) is a postgraduate student registered

for degree / Programme MSc Epidemiology in the Wits School of Public Health.

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

Research Project 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 stated 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 of others in my writing.

Signature: A. Ramadhar Date: 27/11/2020