2019

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

A Multilevel model of self-rated health in Gauteng: a comorbidity study

MSc GIS & Remote Sensing candidate
Ishwari Bhat (721566)
Supervisors: Dr J. Ncayiyana and Dr S. Merlo

School of Geography, Archaeology and Environmental Studies

, Johannesburg

Table of Contents

DECLARATION 5

Acknowledgements 5

Abstract 6

List of Figures 8

List of Tables 8

List of Abbreviations and Acronyms 9

Chapter I: Introduction 10

Background 11

Problem statement 12

Research questions 13

Aim 13

Objectives 13

Ethical considerations 13

Chapter II: Literature review 14

i. South Africa's health history 14

a. South African health 15

b. Health system structure 17

ii. Urban Health 18

ii. GIS in epidemiology 18

iii. Spatial statistics for GIS 20

iv. Statistics in public health 21

a. Regression models 21

b. Multilevel Model Approach 22

c. Comorbidity analyses 24

Chapter III: Methodology 27

Use of self-reports in prevalence studies 27

i. Study area 27

ii. Gauteng health ... 28

iii. Data and variables ... 29

QoL Survey 2015 dataset ... 30

Methods ... 31

iv. Study design ... 32

v. Measurement of Variables ... 32

vi. Analysis ... 33

Chapter IV: Results ... 39

i. Identification of comorbidities ... 42

ii. Spatial analysis ... 48

a. Point Pattern Analyses ... 48

iii. Regression ... 55

OLS models ... 56

Spatial error model ... 58

iv. Multilevel modeling ... 59

Linear model tests ... 59

Conventional multilevel models ... 60

Discussion ... 65

Key findings ... 65

i. Prevalence ... 65

ii. Spatial analysis ... 66

iii. Multilevel analysis ... 68

Multilevel model approach ... 74

Strengths ... 76

Limitations ... 76

Conclusion ... 78

Main findings ... 78

Implications of findings ... 78

Future research..79

References 80

DECLARATION

I [Ishwari Bhat] declare that this research report is my own work. It is being submitted for the degree of [Masters in Geographic Information Systems and Remote Sensing] at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any other degree or examination at any other university.

19th August 2019 in Johannesburg, South Africa

Acknowledgements

This research was supported by University of the Witwatersrand, School of Geography, Archaeology and Environmental Studies. We thank our colleagues from School of Public Health who provided insight and expertise that greatly assisted the research.

We thank Dr Jabulani Ncayiyana, epidemiologist for assistance with comorbidity analyses, and Dr Stefania Merlo, co-ordinator of the MSc in GIS and Remote Sensing for comments that greatly improved the manuscript.

Abstract

The report reviewed the self-rated health of Gauteng's comorbid health in 2015. The outcome variable of this report was defined as the self-rated comorbid health of Gauteng. A multilevel approach examined factors closely associated with comorbid health using both individual and community-level variables. The report addresses the symbiotic relationship between comorbidity prevalence and Gauteng's socioeconomic conditions that foster poor health.

South African healthcare has been characterised by its increasing comorbidity of communicable diseases. As the prevalence of comorbidity varies between spaces, it becomes increasingly important to examine social environments. The aim of this study estimated the prevalence of comorbidity, and determined the factors associated with self-reported comorbidity in the Gauteng province of South Africa.

A multilevel model approach was used in this cross-sectional study. Primary data was provided by the Gauteng City-Region Observatory from the Quality of Live survey (QoL) in 2015, it comprised 30 002 participants above the ages of 18 years, who were selected through numerous sampling stages. enumeration areas (EA) were drawn using probability proportional to size as the primary sampling unit. Comorbidity was illustrated as classes of two, three and four-or-more. Prevalence was estimated as a proportion of comorbid health conditions from the health section of the survey. Spatial autocorrelation was used to detect spatial patterns of comorbidity. Regression models were used to determine factors closely associated with comorbidity in Gauteng.

The estimated prevalence of self-reported two comorbidities (hypertension and diabetes) was 8.97%, three comorbidities (hypertension, diabetes and influenza/pneumonia) was 3.01% and four-or-more comorbidities (hypertension, heart disease/stroke, diabetes and asthma) was 0.96%. Ordinary Least Squares (OLS) models provided the first step for regression analysis, wherein only two comorbidities illustrated spatial dependence among the residual errors. The spatial error model of two comorbidities was interpreted as the most accurate model owing to the high Hausman test p-value. Three comorbidities and four-or-more comorbidities indicated no spatial dependence among the residuals, reiterating the OLS model as the most appropriate regression model for each class. Conventional multilevel models illustrated that self-rated comorbidity prevalence was more likely to report two comorbidities (n=829,8.97%) for *quality of property, child malnutrition, low exercise frequency, high stress* and *low-income bracket* in Gauteng. Self-rated comorbidity prevalence was more likely to report three comorbidities (n=76, 3.017%) for *migrated from another province, family living nearby, adult malnutrition, medium to high stress, low-income bracket* and *female* in Gauteng. Results of four-or-more comorbidities were defined by the limitation that data set was too small for regression analysis.

Comorbidity has a complexity in nature, in that it both influenced by Gauteng socioeconomic environments as well as influenced Gauteng health. Comorbidity became a challenge for Gauteng in addressing complexity, accessibility and cost-effectiveness. It was anticipated that these findings may advise policy interventions in mitigating health disparities in the broader context of South Africa.

Key words: GIS, epidemiology; comorbidity; multilevel model

List of Figures

Figure 3. 1. Map of the Gauteng province and its municipalities. Data source StatsSA Census 2011. 28

Figure 3. 2. Methodology flow diagram..34

Figure 4. 1. Distribution of randomly created respondent points per sub-place area in the Gauteng province. 39

Figure 4. 2 Percentages of singular health conditions..40

Figure 4. 3 Percentage of self-reported health status condition for all QoL population. 40

Figure 4. 4 Distribution of two, three and four-or-more comorbidities.. 48

Figure 4. 5 Geometric means distribution for two comorbidities. 49

Figure 4. 6 Geometric means distribution for three comorbidities..50

Figure 4. 7 Geometric means distribution for four-or-more comorbidities 51

Figure 4. 8 Thematic hot spot map measuring 99, 95 and 90 percent confidence bins for two comorbidities at sub-place level in Gauteng. 53

Figure 4. 9 Thematic hot spot map measuring 99, 95 and 90 percent confidence bins for four-or-more comorbidities in region A – Ekurhuleni..54

Figure 4. 10 Spatial regression decision process (Anselin 2005) 56

List of Tables

Table 3. 1 List of datasets used in the research..29

Table 3. 2 Selection of comorbid health conditions. Data source QoL Survey 2015. 31

Table 3.3 Measurement of variables. Data source QoL Survey 2015. 32

Table 4. 1 Descriptive statistics for demographic variables. Data source QoL Survey 2015. 41

Table 4. 2. Comorbidity combinations using exact binomial distribution 42

Table 4. 3 Description for comorbidity classes. Data source QoL Survey 2015. 44

Table 4. 4 Moran's test for comorbid health conditions. Moran's Index, z-score and p-values were evaluated for significance of the observed pattern. 52

Table 4. 5 Spatial clustering identified as hot and cold spots within Gauteng at 90, 95 and 99 percent CI. 55

Table 4. 6 OLS regression model depicting only the variables significantly associated with comorbidity 56

Table 4. 7 Spatial error model..58

Table 4.8 Linearity and normality plots for linear modelling for two and three comorbidities. 59

Table 4. 9 Two comorbidities and three conventional multilevel models. Variable estimates for two comorbidities were used to illustrate significance for each model. 61

List of Abbreviations and Acronyms

Abbreviation	Meaning
ADL	Activities of daily living
AIC	Akaike information criterion
BIC	Bayesian information criterion
CAPI	Computer Aided Personal Interviewing
CI	Confidence Interval
COPD	Chronic Obstructive Pulmonary Diseases
EA	Enumeration area
GIS	Geographic Information Systems
GCRO	Gauteng City Region Observatory
GPS	Global Positioning System
GWR	Geographic weighted Regression
IADL	Instrumental activities of daily living
MAUP	Modifiable Areal Unit Problem
MDG	Millennium Development Goal
MHF	Migrant Health Forum
NCD	Non-communicable diseases
NDP	National Development Plan
NHLS	National Health Laboratory System
OLS	Ordinary Least Squares
QoL	Quality of Life survey
SADHS	South African Demographic and Health Survey
SES	Socioeconomic Status
STI	Sexually Transmitted Infections
TB	Tuberculosis
UGCoP	Uncertain Geographic Context Problem
VIF	Variance Inflation Factor
WBOT	Ward Based Outreach Teams
WHO	World Health Organization

Chapter I: Introduction

Health is situational and is related to its surrounding environment. John Snow's 1854 Broad Street findings and illustration demonstrate the correlation between geography disease. It was clear the number of cholera deaths in London were directly related to the number of polluted water pipes in the area (Frerichs 2004). This gave rise to epidemiology, the science that explains health patterns and trends in terms of the surrounding environment. As the realm of epidemiology has developed, it has also looked to "geography to discern clues" to treating diseases (Kreiger 2003:14).

Looking into the spatial dynamics of health helps understand diseases of interest as well as serving as a means to plan for as means to plan for interventions and make important decisions. The integration of Geographic Information Systems (GIS) knowledge, statistics and health data into epidemiology prevalence research has been widely used by Public and environmental health departments and epidemiological research organizations such as Word Health Organization (WHO), United Nations Children's Fund, US Centre for Disease Control and other international organizations worldwide (Park and Kim 2014).

GIS offers epidemiology a tool to observe geographic and temporal variations in population health. Ruankaew (2005) highlights the vast capabilities of GIS from the location of intervention areas, to the identification of accessibility of service and prediction outcomes; proving that GIS plays a vital role in health studies. GIS provides a way to "move data from the project level so that it can be used by an entire organization" (ESRI 2017). Naves *et* al (2015) highlights the advantage of GIS in health studies in its ability to provide scientific accuracy for disease analysis. GIS tools include spatial data processing, topographic datasets, satellite sensing imagery and other such databases to visualize and cross-correlate public health and environmental-social factors (Tanser and le Sueur 2002). These tools allow for data analysis with the health sector to investigate the inter-relationships between health and geographical parameters. GIS has become an important application in epidemiological and etiological research because it has the ability to "integrate the digital capturing (of); (and) manage, analyse and visualize geographically referenced data". (Wang *et al* 2015). Soo *et al* (2014) defines comorbidity as the burden of illness co-existing with a disease of interest which may impact on patient outcomes. Soo *et al* (2014) also highlights the importance of high-quality data in determining comorbidity information about patients, as can be seen by the increasing amount of health surveillance, planning and clinical research being conducted.

Nkosi et al (2016) investigates risk factors with comorbidity of respiratory and cardiovascular disease among elderly in South Africa. There is rising pollution from the mine dumps. The gold mining processes that were used has resulted in air, water and soil contamination which is encroaching on residential developments. Exposed communities are of lower socioeconomic status and, within them, the most at risk are the elderly populations because of pathological ageing. The overlap of respiratory and cardiovascular diseases in this study were shared by the same risk factors such as proximity to mine dump, age group, occupational exposure history to dust/chemical fumes and main type of fuel used for heating and cooking at home. Alaba and Chola (2013) state chronic multi-morbidity in South African targets 4% of the adult population, with obesity, smoking and depression as the major determinants.

In recent years there has been a growing interest in the socioeconomic characteristics that influence health research across communities. A multilevel approach has the analytical ability to simultaneously examine individual and community-level factors. Multilevel modelling brings together the conceptual and theoretical understanding of the distinction between group and individual level and its effect on health outcomes. The unbiased estimations of this type of model helps to improve the explanatory power as compared to conventional multilevel models (Kim *et al* 2006). This research report seeks to use self-reported health data within a multilevel framework to better understand spatial dynamics.

Background

Comorbidity is defined as the simultaneous presence of two or more chronic diseases or conditions in a patient (Valderas *et al* 2007). Comorbidity health research is important for policy planning because health reforms seek to reduce costs of healthcare and improve quality of life (Katon and Unützer 2013). Policymakers have the responsibility to focus on specific subgroups that are at a high risk of excessive health costs and poor quality of life. The Robert Wood Johnson Foundation (2011) states that policy implications may include more comorbid-specific insurance coverage, care delivery redesign as well as prevention and promotion efforts to create awareness of the prevalence of comorbidities if area-specific.

Comorbidity has been shown to be associated with variables such as socioeconomic status (SES) and directly associated with variables such as age, survival, the recommended treatment, disability, and health service utilization (Leone *et al* 2012). Comorbidities may also impact treatment effectiveness, particularly when it is associated with incomplete or inadequate treatment delivery. Boulos *et al* (2006) reiterates the importance of comorbidity when studying health. Accurate measurement of comorbid illnesses is a crucial consideration for assessing the success of one's treatment delivery. Gauteng's public health care is plagued with drug overdoses and poor service delivery because of the failure of the National Health Laboratory System (NHLS). Health conditions, such as HIV Aids, tuberculosis (TB) and child mortality, overwhelm the health department (Pandie *et al* 2012). By exploring the underlying health conditions that exist with the Gauteng population, preventative action can be taken to address the weaknesses in management.

Since GIS experts can accurately analyse health outcomes by population and environmental influence, multilevel modelling of factors becomes possible (Wilkinson *et al* 2003). A multilevel model is defined as a hierarchal tool to assess neighbourhood effects (Cromley and McLaferty 2012). This is also considered a more appropriate regression model as compared to an ordinary single-level regression model as it takes into consideration intra-neighbourhood correlation between individual and neighbourhood-level variables. Arcaya *et al* (2013) state that there is potential for comorbidity analyses to be defined within a multilevel framework as it visualises spatial dynamics of health distribution.

Cost, combined with lack of access to data and inadequate technological infrastructure, hinder the application of GIS in "areas of the world inflicted by endemic diseases, poor hygiene and higher incidence of injuries" (Bertazzon 2014). Africa is one of these areas. Epidemiological studies previously applied GIS approaches to diseases that have plagued Africa such as HIV, malaria and TB (Kwandal *et al* 2009, Hay *et al* 2011, Tanser and le Sueur 2002). Tanser and le Sueur (2002) state that disease estimates in Africa are wildly speculated guesswork, therefore it becomes imperative that data collection and reliable statistical analysis be performed before GIS can intervene.

Problem statement

According to Valderas *et* al (2007:259) epidemiology of comorbidities in South Africa is an area of research that is untouched even though it has consequences for "mortality, health-related quality of life, functioning and quality of health care". A major challenge in comorbid populations has been engaging with patients in an attempt to alter their health. As comorbidity may occur for a number of reasons, such as random occurrence in conjunction with risk factors or as an overlap of two or more disorders, it is imperative that comorbidities be well defined (Klein 2004).

We need to improve our understanding of the reasons for comorbidity so that, where possible, disorders are prevented from occurring, as well as to better help individuals and their families who are affected by comorbid disorders. In response to this problem, this report proposes to find factors closely associated with comorbidity and map these locations as an indicator of Gauteng's health. This report will define Gauteng health as an outcome of its comorbidity prevalence. The outcome variables will report the comorbid health conditions observed and measured by the explanatory variables such as gender, population group, age and other relevant socioeconomic characteristics. The findings of this report may also advise policy changes in order to mitigate health disparities caused by comorbid health in the broader context of South Africa.

Research questions

i. How is comorbidity distributed in Gauteng in Gauteng?
ii. Which comorbid health conditions are spatially dependent in Gauteng?
iii. What are the factors associated with the three most prevalent comorbidities within the multilevel model?

Aim

The aim of this report is to examine the spatial associations of comorbidities and assess their effects through a multilevel model of self-reported health data in Gauteng.

Objectives

a. To estimate the three most prevalent comorbid health conditions in Gauteng;
b. To identify spatial dependency of comorbid health conditions in Gauteng;
c. To determine the multilevel model factors associated with three most prevalent comorbidities in Gauteng.

Ethical considerations

The Gauteng City Region Observatory (GCRO) has been granted ethics approval for using the Quality of Life Survey (QoL) data from the GCRO Ethics Committee for research analyses. For the purposes of this study, GCRO signed a user agreement stipulating that participants' confidentiality would be preserved by whomever conducted analysis on the data set.

Chapter II: Literature review

This chapter presents a brief review of the history of the South African health system, followed by a discussion of the role of GIS in health and GIS in comorbidity analysis. The review was an internet-based search on Scopus and Web of Knowledge using the following key terms: GIS, epidemiology, comorbidity, and multilevel model.

i. South Africa's health history

Coovadia *et al* (2009) state that within the South African borders there exists a dysfunctional health system dating back to colonial subjugation and the apartheid dispossession. This has left behind massive challenges for the current health system. Racial and gender discrimination, migrant labour systems and vast inequalities have deformed the health system and left behind a legacy of poor health. Although the public health system has gone through processes of transformation and integration, failed leadership and mismanagement has resulted in ill-informed health policies today. Coovadia *et al* (2009) outlines the health system history of South Africa as summarised below.

The *British colonialism* era sought to medically help local populations through missionary work providing orthodox medical health care for Black Africans. Although medically trained doctors and nurses dominated the health system, indigenous and traditional healers were still used by local populations. The first health legislation of 1807 established the Supreme Medical Committee, which lasted many years, to oversee all health matters. It was the 1883 Public Health Act that responded to the smallpox epidemic. Epidemics of syphilis, TB, yellow fever and cholera were only a few of the diseases plaguing the health system of South Africa. The establishment of the Union of South Africa that the *period of segregation* fragmented health services among the four provinces. Overcrowding, poor working conditions and poor sanitation within migrant labour communities put populations at a high risk of contracting syphilis, TB, malaria and venereal diseases. The 1919 Health Act established the first Union-wide Public Health Department. Community health centers were set up for primary health care in response to high mortality rates. The *apartheid years* saw poverty deepen with doctor to patient ratios steadily increasing, and health services in the Bantustans severely underfunded. The 1977 Health Act perpetuated this fragmentation of health care through provincial responsibility and became less of a local government responsibility. The breakdown of community health centres meant there was an increase in mental disorders and poverty-led health problems among the Black African and Coloured populations. The isolated South Africa was steadily declining economically and a refocus on local population was established. This refocus was exacerbated by the 1983 Tricameral Parliament, which divided health services into white, Coloured and Indian 'own affairs' departments. Towards the end of the Apartheid era, the African National Congress Health Plan built policies for primary healthcare. The *Post-Apartheid democracy* has seen a redistribution of government funding between areas, whereby plans are targeted for free care for children under six, rights-based approaches to youth sexuality, tobacco product control schemes, a focus mental health care, free-water strategies as well as firearm control policies.

a. South African health

Ntuli *et al* (2015) highlights that the increase of hypertension prevalence aligns with increasing burden of chronic diseases, whereby South Africans are experiencing an increase in hypertension among women. Peltzer (2013) states that various socio-demographic factors such as gender (female), insufficient fruit and vegetable intake, diabetes, older age, lower education level, higher body mass index, physical inactivity, lower household income, urban residence, habits and behaviours, including heart dysfunction, smoking and drinking, greater limitations on activities of daily living (ADLs) and instrumental activities of daily living (IADLs), higher frequency of doctor visits, and less social cohesion, have a direct effect on hypertension prevalence. Naiker (2003) states that malignant hypertension is an important cause of morbidity and mortality among urban Black South Africans. Hypertension patients taking specific medication can increase chances of diabetes. Environmental effects on hypertension include the intake of excessive amounts of alcohol and salt, chronic stress, obesity, and physical activity (Pickering 1997). It is important to examine the lifestyle of respondents for a better understanding on hypertension impacts.

Diabetes in Africa is more prevalent among the wealthy and powerful, and where people tend to lead a less active lifestyle (Azevedo and Alla 2008). Obesity is one of the highest contributing factors to diabetes, therefore diet and lifestyle should be closely examined. Obesity can also be linked to urban area lifestyle. Azevedo and Alla (2008) state that increased likelihood of cardiovascular disease, such as hypertension, is linked to diabetes. The targeted age for diabetes is between 40 to 70 years age. According to the American Heart Association (2015), people living with type two diabetes are at risk for cardiovascular diseases because of increased hypertension/ blood pressure; abnormal cholesterol or high triglycerides; sudden weight gain or obesity; and physical inactivity associated with insulin resistance.

Cardiovascular diseases, including hypertension, are a major cause for mortality in persons with diabetes. Sowers et al (2001) state that the risk of *heart disease and stroke* are increased with conditions such as obesity and diabetic cardiomyopathy. Factors contributing to cardiovascular diseases are “found among the elderly, especially women with type 2 diabetes” (Sowers et al 2001).

Cordero et al (2011) states that hypertension is the most widespread and poorly managed risk of cardiovascular diseases. Bhatnagar (2017) stated that changes in environment because of migration, modifications in lifestyle choices, as well as shifts in social policies and cultural practices have an impact on cardiovascular diseases. Exposure to air pollution can be linked to lung function, and in turn increased risk of heart attacks. Proximity to mining areas and landfill sites should be closely examined in relation to comorbidity prevalence.

While industrial and mining activities contribute to poor air quality, there are also many households that either cannot afford or don't access to use of electricity for their energy needs, resulting in usage of alternative fuels such as gas, paraffin, coal and wood (Maree *et al* 2017). Although Gauteng has a number of air quality monitoring stations that measure the air pollution levels, it is insufficient to only monitor large polluters from industries and mines., It is better to rather focus on domestic sources of air pollution. Jia et al (2013) states that *asthma*, of the four diseases, is one that is more prevalent among a young population, especially among women. Asthma also contributed to a seven-year loss in life expectancy. Strand et al (2018) found that there is an increased risk of mortality from cardiovascular diseases in individuals with active asthma, thereby linking heart stroke and asthma. Iribarren *et al* (2012) conducted a similar study predicting that asthma's relation to cardiovascular diseases pertains to roles of gender, concurrent allergy and asthma medications. Childhood asthma was not included in this research report as comorbidity age was above 65+ years. Asthma attacks have been linked to "exercise, respiratory infections, exposure to allergens, tobacco smoke and air pollution" (National Centre of Environmental Health., 2017). Mining areas and urban pollution should be examined in relation to comorbidity prevalence.

Influenza, commonly known as the flu, is a highly contagious viral infection that is one of the most severe illnesses during the winter season (American Lung Association 2018). It is easily spread from person to person. Mosenifar (2018) describes pneumonia as a serious infection or inflammation of the lungs, and it occurs when the air sacs fill with pus and other liquid, which in turn blocks oxygen from reaching the bloodstream. Influenza is a common cause of pneumonia especially in young children, pregnant women and the elderly with specific chronic conditions. Although influenza vaccines are necessary every season, pneumonia vaccines are only necessary once. Health24 (2018) describes influenza as being caused by a weak immune system as attributed to lack of sleep, chronic stress, poor hygiene, insufficient exercise, poor nutrition, respiratory failure, sepsis, secondary infections, and excessive smoking, drinking and use of antibiotics. Roomaney *et al* (2016) highlights that between 1997 to 2015, pneumonia and lower respiratory infections were the leading causes of death worldwide, especially in children under 5 years of age. Age and a positive HIV status are contributing factors of pneumonia along with malnutrition, indoor air pollution, parental smoking and concomitant diseases (Roomaney *et al 2016).* Comorbidities such as diabetes and renal diseases are also highlighted pneumonia risks for the elderly.

The multi-factorial risks of hypertension, diabetes, heart disease and stroke, asthma and influenza/pneumonia are closely linked, and through this the comorbidity pattern has been proven. Causes and links include SES, sedentary lifestyle, stress, diet, and genetic patterns that should be viewed from a broader perspective.

b. Health system structure

In recent years, the South African health profile has fluctuated. According to the WHO (2018) country profiling system, South African adult risk factors for raised blood pressure and obesity are high amongst men and women over 20 years of age. WHO (2018) also states that the highest causes of death that have increased include diabetes mellitus, TB, hypertensive heart disease and road injury. The causes of death that have decreased include respiratory infections, interpersonal violence and diarrhea diseases. The fluctuation in the South African health status means that it is vital to focus on smaller community impacts.

By 2030, the National Development Plan (NDP) envisions a raised life expectancy to at least 70 years; a generation of under 20s that are "largely HIV-free; an infant mortality rate of less than 20 deaths per thousand live births; an under-five mortality rate of less than 30 per thousand; and a significant shift in equity, efficiency and quality of health service provision" (South African Government 2019).

The demands on Gauteng health is greatly impacted by the rising population. According to Gauteng Department of Health Annual Report (2018), the burden of communicable diseases as well as interpersonal and motor vehicle-related trauma remains high and is putting pressure on the inherited public health system. Each year, there are over 21 million patient visits in the Department's clinics and community health centres. Health literacy, referring to degree to which individuals can make informed medical decisions, is a major determinant of health outcomes among disadvantaged populations, with a poor socioeconomic status. Griffey *et al* (2014) recognises that low health literacy can result in a number higher returns to the emergency units. The increasing demand for health services comes as a result of rapid urbanisation, the high burden of disease and the decreasing number of people covered by medical aid. Success has been seen in the steady decline of mortality ratios of non-pregnancy related infections, owing to the triumph of the Antiretroviral Treatment Programme. The Health Annual Report of 2018 highlights the positive attributes of Gauteng health, as Gauteng currently holds the highest percentage (seventy-five percent) among all the other provinces of having the ideal clinic status.

ii. Urban Health

Drexel University Urban Health Collaborative (2018) characterises urban health as process of rapid urbanisation impacting the city's wellness and health status. The definition of an urban area is often inclusive of all surrounding areas as well as the diverse segments of daily life. It is important to examine all aspects of urban life in order to understand urban health. This includes areas of safety, employment, poverty, built environment, education and housing, to name a few. Urban health disparities are a result of the inequalities found within the urban area. iLead (2018) observes that the rapid urbanisation of Gauteng, resulting from large-scale migration for employment and better services, is posing challenges for urban development. Rapid urbanisation puts pressure on natural resources, such as water. More people also means more waste production and increased traffic. Vearey *et al* (2017) highlights a local informal working group, Johannesburg Migrant Health Forum (MHF), that has enabled the sustainability of the province by playing a role in supporting the development of appropriate local responses to migration and health in Gauteng.

Vearey (2011) describes the urban context as affecting the urban health profile, whereby the "social and physical environment of cities combined with the health social service systems form the primary determinants of the urban" population (Vearey 2011: 2). The Gauteng Strategic Plan (2016) illustrates that Gauteng's urban health profile consists of HIV/AIDs, TB, Sexually Transmitted Infections (STI) and Non -Communicable diseases (NCD).

In South Africa, there has been a "transition from the Millennium Development Goal (MDG) era to that of the Sustainable Development Goals (SDGs)" (South African Health Review 2017: 13). This transition entails a focus that is predominantly on maternal and neonatal health, as well as AIDS, TB and malaria, to ensure healthy lives and promote well-being for all, at all ages (Health Systems Trust 2017).

ii. GIS in epidemiology

Epidemiology studies the spread and control of diseases as a function of time and location. Disease mapping was first introduced in the late eighteenth century for "identifying communicable diseases and describing its spread" (Rytkönen 2004). Today, disease mapping is only one of the many capabilities of GIS in epidemiology, and GIS can be used to survey, monitor, plan and evaluate risk groups for intervention (Ulugtekin *et al* 2006). Mokhele *et al* (2012) highlights that the scope for public health programmes in South Africa have been influenced by the ever-changing characteristics of the population it serves. Population growth has spurred the idea for GIS to help visualize problems in relation to existing health infrastructure. GIS has unlocked the potential to determine locations of high-prevalence areas and populations that are at high-risk. This could target public and private health disparities, as access to adequate care and quality health services is a major challenge in the rural-urban dynamic (Murad 2018).

GIS provides epidemiologists with tools for collecting, managing and reviewing socioeconomic and environment-related data (Gatrell and Löytönen 1998). GIS meets the demands of outbreak investigation and response which is central to the design of prevention and control strategies. Gatrell and Löytönen (1998) state that maps represent the epidemiological information and identify spatial trends; this is still important today in articulating policy plans for health. GIS itself ranges from daily operations to long-term planning, where GIS and epidemiology are at the intersection of the disciplines and the expertise in determining intervention and prevalence (McLafferty 2003). Data is represented in accordance to the information conveyed. Dot-density, proportional circles and choropleth maps may be used to highlight risk zones of illness (Margai 2010).

Although GIS spatial analysis effectively measures spatial prevalence in epidemiology, Rytkönen (2004) suggest that there is a concern around its accuracy. Hence, spatial statistics and GIS technology have begun to merge, and hypothesis testing and prediction modelling have become imperative to epidemiological research. Nandi and Shakoor (2010) state that geostatistical tests are used alongside GIS to improve the accuracy of the results of geographical clustering and patterns. Before the availability of computational-assistance tools, there were limitations with regard to handling vast datasets. Today, GIS has immense capacity to examine variables and preform prediction models. Nandi and Shakoor (2010) argue that GIS is limited and therefore the analyses of epidemiology is criticized. Nandi and Shakoor (2010) continue to argue that one of the concerns is scale and quality of GIS data, which is faced with bias when addressing spatial levels of disease effect. Health information aggregated at an administrative level may sometimes be inadequate in addressing wider public health concerns. Another issue with GIS in health has been that the use of high-resolution data infringes on the privacy of individuals (Tanser and le Sueur 2002). Therefore, there is a need for finer geo-referenced spatial health data. Boulos (2004) suggests that maps that have been standardized pose the issue of hiding important information of variation at smaller scales. Smoothing GIS techniques are the biggest culprits of local area averaging. Rytkönen (2004) states that the appropriation of colour usage to represent the maps can also be stated to convey a political agenda thereby underestimating the prevalence of diseases. Although GIS has its challenges when it comes to accuracy, it can be highly accurate if paired with appropriate statistical methods and knowledge.

Shahid *et al* (2009) highlights an example of GIS in epidemiology. Access to healthcare is simply not Euclidean straight-line distance but rather a complex route network. Taking into consideration the health condition and the socioeconomic variables, a multivariate spatially autoregressive model was used. The spatial weights matrix can be used to define a distance parameter, number of nearest neighbours and distance decay function. Results were able to illustrate the socioeconomic variables associated with the health condition, thereby enabling areas of social and economic concern to be highlighted. An important objective of epidemiological research is being able to identify risk factors associated with disease, therefore epidemiological analyses include regression methods and complex analytics (Ressing *et al* 2010).

The exploration of the occurrence of comorbid substance-use disorders among psychiatric patients is illustrated in Weich and Pienaar (2009). Negative outcomes indicated non-substance disorders such as behavioral problems, unemployment, social isolation, and relapse and re-hospitalization. Descriptive statistics included were mean, standard deviation and 95% confidence intervals (CI). Comorbidity was associated with 23% of patients, of which abuse, and dependence has been at controversy.

iii. Spatial statistics for GIS

The solution to GIS-bias lies in the notion of the integration with spatial statistics and modelling. The input of statistical analysis can have greater effects on the accuracy of results. Different approaches exist that merge GIS and statistic models. Orloff and Bloom (2014) compare the two most common measures of statistical modelling: 1. the *Frequentist approach* calculates CI for quantifying the significance of the hypothesis; 2. the *Bayesian approach* differs as probability distributions represent a proportion of the information and the rest is calculated from prior knowledge of calculations. This approach exemplifies the GIS assumption of interpolation, whereby neighbouring areas are more likely to be similar than remote areas. However, finding an appropriate interpolation method poses its own challenges as it may “distort the model and produce false spatial patterns” (Mitas and Mitasova 2000: 482). The key to avoid bias in GIS is to overlap the two disciplines with assistance from disease distribution mapping and statistical analyses preformed simultaneously.

As it has been previously stated, GIS and statistical analyses must be performed simultaneously to achieve utmost accuracy for this research report. For this report, descriptive statistics and spatial statistical measures are used to understanding the outcomes. Without doing descriptive and spatial measures, the research becomes vague and inaccurate. In terms of health data analysis, descriptive statistics can be used for studying and organising preventative control programmes for diseases. It is important to remember that three criteria are met: homogeneity, equality and contiguity.

iv. Statistics in public health

The increase in availability of computational power and accessible statistical software packages has been reported in recent health literature. Biostatistics is vital in critically exploring public health data. Hayat *et al* (2017) demonstrate the importance of statistics in health studies by reporting descriptive statistics in graphical or tabular form in more than 95% of public health articles.

Statistics in public health is concerned with descriptive and inferential statistics. Descriptive statistics summarizes the data numerically and examines the underlying patterns. Numerical series consist of measures that describe the central tendency of data and their dispersion. The distribution skewness symmetry and kurtosis distribution tailed explores the distributions. Graphical techniques also summarize the data relationships. Streiner (2017) states that inferential statistics allows the user to determine the probability of any differences between or among groups, and if they occur due to chance. Inferential statistics also has the ability to generalize the results.

Chung *et al* (2004) state that numerous spatial statistical studies fail to include empirical evidence and reasoning for their choice of population size or smoothing parameters. Alternatively, many studies fail to justify only one type of spatial weight matrix and its impact on the sensitivity of weight choices. There still exist gaps between GIS and health research, therefore the integration of statistical measures from other fields are prudent to addressing complex problems.

Statistics is a collaborative science and it does not make its contributions in isolation. Statistics in public health involves sectors such as medicine, chemistry, engineering and politics. Statistics in public health’s achievements have ensured: routine immunization of children, improved motor vehicle safety, a decline in heart disease mortality, safer foods, the recognition of tobacco as a health hazard, personalized medicine as well as disease modification (Wicklin 2013)

a. Regression models

The OLS model is a non-spatial model and is a regression model that is a generalized linear modelling method used to predict single responses. OLS assumes that an error term is stochastic.

The results of an OLS regression model examine the p-value, R^2 and residuals. Building a regression model is a complex process that involves determining which variables are effective predictors, then repeatedly removing and/or adding variables until one finds the best regression model possible. Even though OLS regression is a straightforward method, if the regression model does not satisfy all the assumptions required, the interpretation of results is biased and inaccurate. Misspecification of variables illustrate an over- and under prediction of residuals from the model appear to be clustered spatially. Geographic Weighted Regression (GWR) analysis provides a solution to this as it reveals spatial elements.

The assumptions of an OLS model are that: the random error term has a mean of zero; the random error terms are uncorrelated and have a constant variance; and the random error term follows a normal distribution. Spatial dependence may occur when there is a measuring error of spatial units during data collection.

Results from the OLS model examine the Variance inflation factor (VIF), Joint F-statistics, Koenker (BP) statistic, Joint Wald statistic and the Jarque-Bera statistic. Regression models with statistically significant, nonstationary data are often good candidates for GWR analysis.

b. Multilevel Model Approach

A multilevel model approach will be used throughout this research report to examine comorbidity at an individual and community level. Multilevel models are useful in analysing the effects owing to their ability to analyse two or more explanatory variables in tandem. However, there were misleading results of data exhibiting spatial dependency errors. These outcomes may arise, particularly in health-related research, because of similar economic and environmental conditions that are more likely to be exposed to locals. Park and Kim (2014) offer a more advanced spatial regression model called the *eigenvector spatial filtering method*, which is more statistically accurate because of its absence of multicollinearity. Previous multilevel models have been severely criticised for demonstrating spatial dependency bias. Spatial dependencies of health outcomes arise from similar socioeconomic and natural environmental conditions. Park and Kim (2014) suggest the integration of multilevel models and spatial regression models can decompose the spatial structure of the map into scaled-down spatial patterns. Spatial filtering techniques are more flexible and more statistically reliable because of the absence of multicollinearity issues.

Multilevel research approaches have been used to broaden the perspective and improve statistical explanation methodology, by taking into account different levels of spatial aggregation. It is important that data is aggregated at the accurate spatial level for easier relational database management for statistical and mapping modelling. Multilevel regression analysis is one of the most widely used methods in neighbourhood effects on individual outcomes. Xu (2014) states that conventional multilevel models ignore the between-neighbourhood correlations and assume independence in one neighbourhood from those in another, leading to overstatement of the statistical significance of neighbourhood effects.

The spatial econometric methodology has matured and evolved to include spatial statistics, regional science and analytical geography. Spatial interaction is limited in how it incorporates spatial econometrics into the model. Anselin (2002) states that the spatial lag and spatial error models occasionally induce radically different spatial correlation patterns, which do not match the underlying theoretical model.

Best *et al* (2000) refers to the relationship between disease rates for small geographic areas and exposure to environmental risk as *ecological regression*. For the purposes of this research report, the socio-demographic, socioeconomic and medical factors will be attributed as covariates or risk factors to comorbidity in Gauteng. The "principal goal of geographical epidemiology is to increase our understanding of disease etiology by revealing the underlying relationships between disease prevalence and spatially varying factors" (Best *et al* 2012). One must be cautious of Modifiable areal unit problem (MAUP), as the aggregation of points may introduce approximation errors and uncertainty, as well as distort the true relationship for individuals. This is called *ecological fallacy*. A Bayesian spatial modeling approach can reflect individual attributes to eliminate ecological fallacy, as well as capture spatial autocorrelation to offer coherent inference at any spatial resolution.

Multilevel models are used to understand the "importance of contextual effects in relation to individual level social and demographic factors in health outcomes" (Duncan *et al* 1998). Duncan et al (1998) criticises contextual effects as they may reduce the overall summary measures of health outcomes. Therefore, it is important to remember heterogeneity between respondents and between contexts.

"Everything in general and nothing in particular" highlights a drawback of analyses of the cause/health outcomes relationships (Duncan *et al* 1998). General statistical approaches pose serious technical difficulties as they only function at a single level. Disaggregation of higher-level factors to the individual level poses a threat to the assumption of independence and increases the chances of finding differences when none exist. Multilevel models are able to fix such problems, as it handles the micro-scale of individuals and the macro-scale of contexts. The multilevel models have the ability to distinguish different levels from a larger underlying population.

Multilevel structures can also be used to reflect temporal variations. Health researchers would be able to investigate several health statuses simultaneously. The multivariate, multilevel model approach directly measures how each factor is related to individual-level characteristics. It also has the advantage of being able to assess correlation of health outcomes between the individual and between the spaces. An individual's health is influenced both by place of work and where they live. A cross-classified structure allows individual-level characteristics to be simultaneously modelled, making it possible to identify contexts that have a confounding influence.

The geography of health-related behaviour around the world suggests that similar types of people behave differently in different types of places. "Null model" represents the estimated "between-location" and "between-region" variability without taking into account the different types of people (National Centre for Research Methods 2017). A "random-intercepts model" includes the predictor variability, so that variability between places is estimated taking into account population composition (Pillinger 2019). If there is a convergence around the 0, this represents the overall national estimate).

This research will use a multilevel model approach that will assume "individuals belonging to a particular neighbourhood are not explanatory of each other because they are presumed to share similar characteristics of that neighbourhood" (Park and Kim 2014).

c. Comorbidity analyses

There are many factors affecting patterns of comorbidity, whereby expertise in diagnosing and treating these diseases undertakes research into cause and treatment. Comorbidity may be explained in its variations across an environment. The criteria for one disorder may overlap with the criteria for another disorder, for which the differences may not appear as separate. Population surveys are credible sources of information that describe patterns of comorbidity. Understanding comorbidity is imperative to understanding trends of health and interventions for effective treatment.

Soo *et al* (2014) uses a validation study to review routine hospital administrative data to determine comorbidity information about patients. Prevalence, kappa statistic, sensitivity, specificity, positive prediction value and negative prediction value were methods used for assessment. This article highlighted comorbidity research as an important dimension in health care that is under-reported and under-investigated, owing to methodological challenges in assessment.

Laranjo *et al* (2016) uses GIS to monitor the prevalence of comorbidities (hypertension, dyslipidaemia, and obesity) influencing type two diabetes in Lisbon, Portugal. Larajno *et al* (2016) proves that the use of individual-level data, for descriptive statistics and multivariate logistic regression, is an effective method to explore comorbidities. With the use of statistical outputs and maps, variation and distribution of comorbidities can be detected.

Although this research report highlights the uses of GIS in epidemiology, the statistics behind the analyses are the core elements of the topic at hand. Comorbidity prevalence must be analysed through numerous processes to justify its accuracy. De Groot *et al* (2003) states that although there are numerous comorbidity methods available; data content, reliability and validity are always up for contest. Comorbidity of chronic conditions among the elderly is outlined by Caughey *et al* (2008) in Australia and highlights a strong relationship between age and prevalence of health conditions. Each health condition was cross-referenced by age group for the prevalence of comorbidities. This is an important method to follow as it gives more accurate results. However limitations to this study, and most studies, are that, when compared to hospital records of chronic conditions, the self-reported survey differs in results.

Four approaches are used to estimate the prevalence of common chronic conditions and co-occurrences of diseases. These are “identification of most occurring pairs and triplets of comorbid health conditions; cluster analysis of diseases; principal component analysis; and latent class analysis” (Islam et al 2014: 2). The article highlights that there remains a lack of clear understanding of an internationally accepted standard for assessing which diseases are likely to co-occur. This study proved that although different analytical methods lead to different associations, there is consistency in the associations. In addition to descriptive statistics, odds ratio and relative risk can be applied to study comorbidity, whereas multimorbidity sees observed and expected ratios (Van den Akker *et al* 2001).

Kleintjies *et al* (2006) reviews prevalence estimates by modification of prevalence risk factors. Comorbidities must be included as a weight to prevalence estimates of which reduced the prevalence of chronic disorders. The unadjusted prevalence and adjusted for comorbidity served as an accurate indicator for mental disorder prevalence rates.

Chronic pain can be linked to psychological comorbidities, as outlined by Tunks *et al* (2008). The authors highlight that there are discrepancies in regression analysis, such as small sample size and selection bias. Many examples are used to illustrate the association between chronic pain and its comorbidities, which include depression, dysthymia, anxiety disorder and migraines.

This research report will focus on Park and Kim (2014) for spatial regression methodology, and the comorbidity methods will be supported by descriptive statistics and spatial analyses.

Chapter III: Methodology

The methodology highlights the sequence in which this research was conducted and completed. As previously stated, the research report adopts a multilevel model approach that examines the individual and community-level factors associated with comorbidity in Gauteng.

First, comorbidities prevalent in Gauteng were identified through statistical analyses. For the purposes of this research report, comorbidity was congregated into three groups: two, three and four-or-more comorbidities so as to group the health conditions respondents faced on a daily basis. All analyses were adjusted accordingly.

Second, spatial dependency of comorbidities in Gauteng was determined by performing spatial autocorrelation and point pattern analyses. Global Moran's I, point frequency, density, dispersion and arrangement and OLS model were methods used to identify the comorbidity patterns.

Lastly, the effects of the factors, associated with the most prevalent comorbidities in Gauteng, were determined using a multilevel model approach. Multilevel models were produced, compared and highlighted the associated factors with effects on Gauteng health.

Use of self-reports in prevalence studies

Self-reports are characterised as the collection of information on the respondent's disease status through the use of the 2015 Quality of Life Survey's questionnaire and interview process. The use of self-reports was a more economically feasible source of data as it was efficient and readily available for use. However, there were limitations to using self-reports such as underestimation of prevalence estimates and result bias.

i. Study area

Although Gauteng is South Africa's smallest province, according to StatsSA (2017) it contains the highest population growth rates, reaching 24.1% in 2017. Gauteng encapsulates three metropolitan municipalities (City of Johannesburg, City of Ekurhuleni and City of Tshwane). Each metropolitan municipality is further sub-divided into two district municipalities, and then sub-sectioned into six local municipalities. The 2579 sub-places make up the smallest EA's of Gauteng. Figure 3.1 illustrates this research report's study area, containing the topography and land use of Gauteng in relation to other provinces.

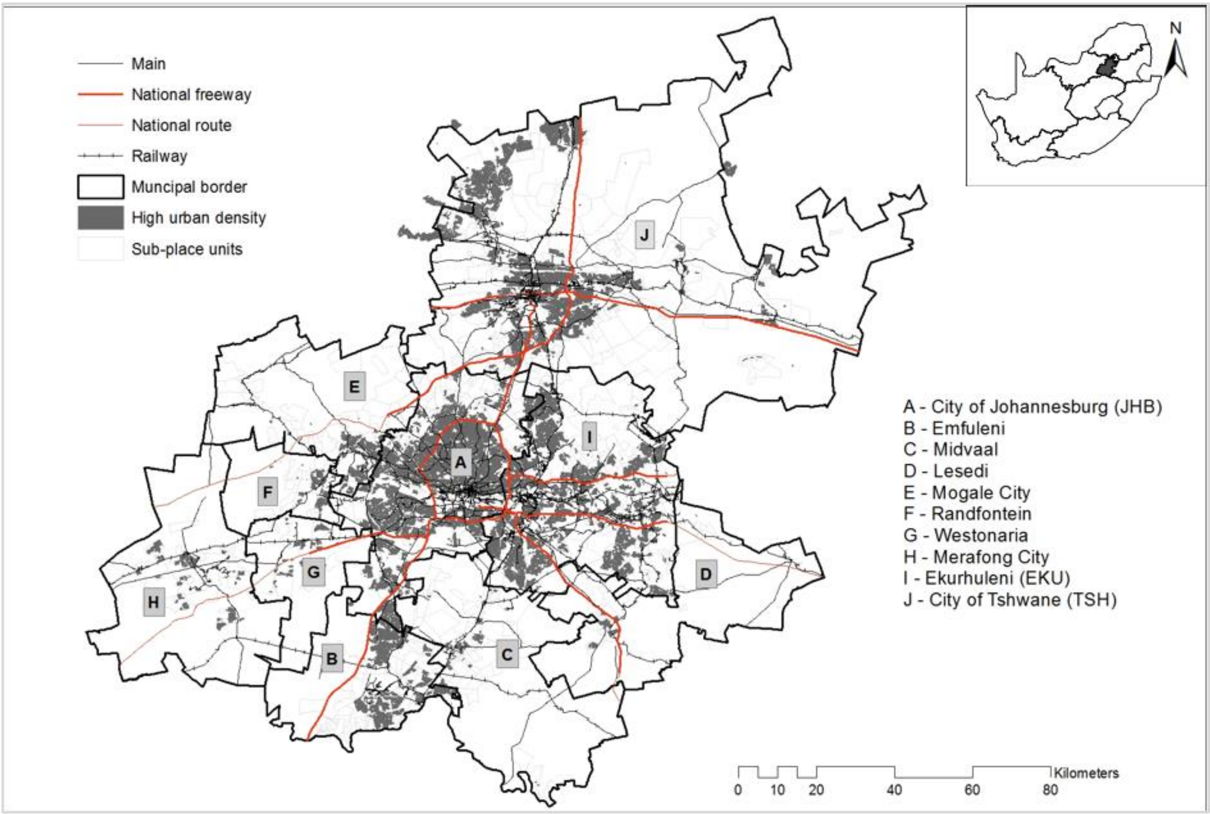


Figure 3. 1. Map of the Gauteng province and its municipalities. Data source StatsSA Census 2011.

ii. Gauteng health

According to StatsSA (2017), the leading causes of death for South Africans in 2015, that were indicated as "natural cause" on the *death notification forms,* were: certain eye infections and parasitic diseases (19.5%); multidrug-resistant TB and extensively drug-resistant TB; diseases of the circulatory system (17.8%); diseases of the respiratory system (9.6%); and external causes of morbidity and mortality (11.1%). Key findings in the Morality and Causes of Death in South Africa (2015) publication state that over 55% of deaths were attributed to non-communicable diseases, which were highest among young males. TB was once again the leading underlying natural cause of death in 2015, followed by diabetes (StatsSA 2015). Hypertensive diseases and other viral diseases followed closely behind. The publication specified that females had nine out of the ten non-communicable diseases, as compared to men who had eight out of ten. The leading cause of death for females was TB and for males was diabetes.

Health is monitored by the Gauteng Department of Health who derives its mandate from the South African Constitution, The National Health Act and other legislative frameworks. The *Development and Research* programme within the Gauteng Department of Health was created to provide sustainable development programmes facilitating the empowerment of communities through research and collection of demographics. According to the National Core Standards, “Gauteng achieved the highest score and was rated number one in the country for three consecutive years” (AllAfrica 2015: 1). The Steve Biko Academic Hospital is the star achiever of Gauteng. The Gauteng Department of Health’s Annual Report for 2015 highlighted the Ward-based Outreach Teams (WBOTs) for their success in identifying pregnant women in need of care, children who missed immunisation as well as individuals who were not receiving treatment for communicable diseases, such as TB.

South African Demographic and Health Survey (SADHS) (2016) characterises the health of the Gauteng region by its HIV and AIDS prevalence, obesity in the African race, high alcohol consumption rates, and lowered child mortality rates, when compared to other provinces. The Gauteng Department of Health has placed some importance on improving the health status of communities through improving primary health care facilities by: extending operating hours; reducing queues; improving pharmacy management; scaling-up and rolling-out Antiretroviral (ARV) treatment; and improving the Emergency Medical Service (EMS) call-response time.

iii. Data and variables

Data that were utilised throughout the research report are listed in the table below (table 3.1).

Data consists of: (1) *Administrative areas*, which determine the scale at which health is being examined. This research reported at a sub-place level; (2) *Quality of Life survey* of 2015 from the GCRO self-reported data on individual and community health. The QoL Survey 2015 was analysed in SAS, SPSS, and Excel; (3) *Census data* for the socioeconomic characteristics of Gauteng and related comorbidity prevalence. These data were extracted from the Census 2011 database in tabular format (Excel spreadsheets) and analysed in ArcGIS and RStudio.

Table 3. 1 List of datasets used in the research

Data	Brief description	Format	Source	Date acquired
Administrative areas	National, provincial, municipal, ward	Shapefile	www.ngi.gov.za	10 April 2017
Quality of Life Survey	Age, gender, address, health facility utilization, level of satisfaction with health care service, medical aid coverage, community health problem, individual health problem.	Excel table	GCRO Quality of Life survey 2015	2015

QoL Survey 2015 dataset

The GCRO conducts a QoL Survey every two years from respondents across Gauteng's province for the purpose of understanding the socioeconomic, political and social environments, and the attitudes of residents. A Computer Aided Personal Interviewing (CAPI) method was used to collect data via the answers to a survey comprising open and close-ended questions. The QoL Survey 2015 contains 12-section questionnaire, collecting continuous and discrete information about demographic and household information, migration to Gauteng, community environment, transport, access to services, satisfaction with government and services, personal life, employment, crime and safety, community participation and protest activity and health. Some of the questions were randomised to avoid possible bias, which can occur if similar questions are asked in sequence. Logic checks were programmed into the questionnaire to ensure consistency of responses. The sampling method used implicit stratification, as the population was sorted according to a health characteristic. Further analysis was carried out. Implicit stratification entailed the sub-division of the most occurring and highest frequency comorbidity class. GCRO has provided the data for this research report, which examines the general households of Gauteng's 30,000 respondents. The anonymity of the data has been protected as health data is deemed as sensitive informed; therefore, the names and any personal information provided in the QoL Survey have been omitted from this research report.

The pre-processing of these data removed potential data errors; however errors still remained and were once again pre-processed. The data was collected by the GCRO with a level of spatial accuracy to the household location but, due to issues of privacy and anonymity, the information released for this study did not have Global positioning system (GPS) coordinates. The data were provided by GCRO at the spatial unit of sub-place. Due to the requirements of this study, data had to be converted to points. Random points were created to represent the XY coordinate locations corresponding to number of respondents in each sub-pace unit. Interviews were not conducted in equal amounts at each sub-place unit because some wards consisted of fewer EA's than was required by the sample design. An addition of five visiting points was selected per EA as oversampling points. The geometric mean indicated the central tendency of the dataset and contributed to the normalisation of the dataset. Normalisation has been conducted to ensure accuracy of comparisons between sub-places.

The survey allowed the participant to answer as many questions as they could, and outcomes were analysed Further. There were various types of questions within the QoL survey, including Likert-scale, close-ended and multiple choice. Likert-scale questions are vital in measuring a respondent's opinion or attitude (ie. "How worried do you feel? 1 being little to no worry and 7 being very worried"). Likert-scale survey questions contained five to seven options. The choices arranged in descending order, including a mid-point for neutral opinions. Close-ended questions were used to answer yes or no questions (ie "*Do you receive social grant? Yes/no").* Multiple choice questions asked the respondent to choose between more than two options *(ie "To which population group do you belong? African / coloured/ Indian / Asian / white / other").* The list of questions used in this report can be found in the Appendix.

Individual-level health described the comorbidity status of the individual of that household, and community-level health examined health and relating environmental problems facing the community. Only the three highest proportional outcomes within the 95% confidence limit were analysed as comorbidities.

The main focus question, within the QoL Survey 2015, was used to define health by comorbid health conditions to meet the aims of this research report. This question used can be found in table 3.2:

> *In the last year, have you or any other member of this household had any of the following conditions*

Using the multilevel model approach, individual and community-level variables were used to examine comorbid health conditions, posed by this main focus question. Good health was described by fewer comorbid health conditions selected from the survey, and poor health was illustrated as having more than two existing comorbid health conditions. The respondent had a choice of twelve health conditions, and the combinations of two, three and four-or-more were defined as comorbidities.

Table 3. 2 Selection of comorbid health conditions. Data source QoL Survey 2015.

Focus question	Comorbid health conditions
In the last year, have you or any other member of this household had any of the following conditions	Hypertension, heart disease/stroke, diabetes, TB, Cancer, diarrhoea, emphysema/bronchitis, asthma, HIV AIDS, influenza/pneumonia, mental illness

Methods

Hox (2002) described multilevel analysis as the interaction variables characterising individuals and variables characterising groups. Multilevel model variables could be defined at any level of hierarchy and moved by aggregation and disaggregation. This research report used the multilevel model approach to examine how individual comorbidity-level variables had an impact on Gauteng's health.

Contextual variables, such as the point data from the QoL Survey 2015, were disaggregated into sub-place units to represent comorbidity in Gauteng.

iv. Study design

This report follows Levin's (2006) cross-sectional study design that estimates prevalence of an outcome for the given population. Cross-sectional research refers to the use of groups of people who share common characteristics but differ in the variable of interest, for example SES, education and ethnicity (Roundy 2018). This report illustrated a cross-sectional study design as survey data was collected at one point in time. The period in which data was collected was in accordance with the QoL Survey 2015. The cross-sectional study design highlighted the descriptive nature of the QoL Survey and described a sub-group within the population (respondents with comorbid health conditions). Non-respondents, respondents who did not complete questions within the survey, posed the challenge of bias on the probability of that outcome.

v. Measurement of Variables

The types of variables in this report inform the methods used for statistical analysis, as each unique regression method requires specific variable selection based on variable type. Categorical, binary and continuous variables are described in table 3.3. Categorical variables take on the values that are either names or labels. Categorical variables were listed as follows: *population group, age, employment status, income, place of birth, education, exercise frequency, stress level, reason for living here, improvement or deterioration in the past 12 months, biggest community problem, water quality, health status, and child and adult malnutrition*. Binary variables are the variables with only two values. *Gender*, for example, is listed as a binary variable. Continuous variables may take any given numerical value. The *household respondents'* variables were listed as a continuous variable in this report.

Table 3.3 Measurement of variables. Data source QoL Survey 2015.

Variable	Type	Variable	Type	Variable	Type
Population group	Categorical (African, coloured, Indian/Asian, White, Other)	**Gender**	Binary (1: male, 2: female)	**Age**	Categorical (has been recoded; groups ranging from 1 – 11)
Employment status	Categorical (Employed, unemployed,	**Income**	Categorical (has been recoded;	**Place of birth**	Categorical (Born in Gauteng,

	other)		groups ranging from 1 – 19)		migrated from other province, migrated from another country)
Education	Categorical (has been recoded; groups ranging from 1 – 18)	**Household respondents**	Continuous	**Social grant**	Binary (1: yes, 0: no)
Exercise frequency	Categorical (Everyday, few times per week, few times per month, hardly ever, never)	**Stress level**	Categorical (have been recoded; ranging from 0 – 10)	**Reason for living here**	Categorical (has been recoded; groups ranging from 1 – 16)
Improvement of deterioration in past 12 months	Categorical (Improvement, deterioration, no change)	**Biggest community problem**	Categorical (has been recoded; groups ranging from 1 – 29)	**Water quality**	Categorical (always, usually, sometimes, hardly ever, never)
Health status	Categorical (excellent, good, poor, very poor)	**Child malnutrition**	Categorical (never, seldom, sometimes, often, always, no children)	**Adult malnutrition**	Categorical (never, seldom, sometimes, often, always)

Data limitations

Sample size determination places limitations on the type of analysis performed. This research report depicted three data sets of varying sample size. Four-or-more comorbidities contained a small data set which could not support high-resolution spatial distribution maps and hot spot analysis. Owing to the small sample size, four-or-more comorbidities was unable to perform conventional multilevel model analysis. As there were more explanatory variables than respondents, four-or-more comorbidities was excluded from multilevel model analysis.

vi. Analysis

Pre-processing of the data revealed errors in how the data was captured. These errors included misguided information and misspelt words. Misguided information was re-worked and edited to make sense of data, and misspelt words were corrected as this would have had implications on analysis.

a. Data pre-processing

Pre-processing of the data was a necessity as the dataset was highly detailed and contained numerous errors from data collection. Errors included misguided information, illegible notation and grammatical mistakes. Pre-processing was vital to comorbidity analyses as the QoL Survey did not contain GPS locations of respondents but rather an aggregation of data at sub-place unit. Random points were created and represented the exact number of respondents per sub-place unit. This was one of the major limitations of research as the exact coordinates could not be matched to the respondent, and instead respondents were matched to their location within the sub-place unit. The methodology followed three steps, as outlined by figure 3.2.

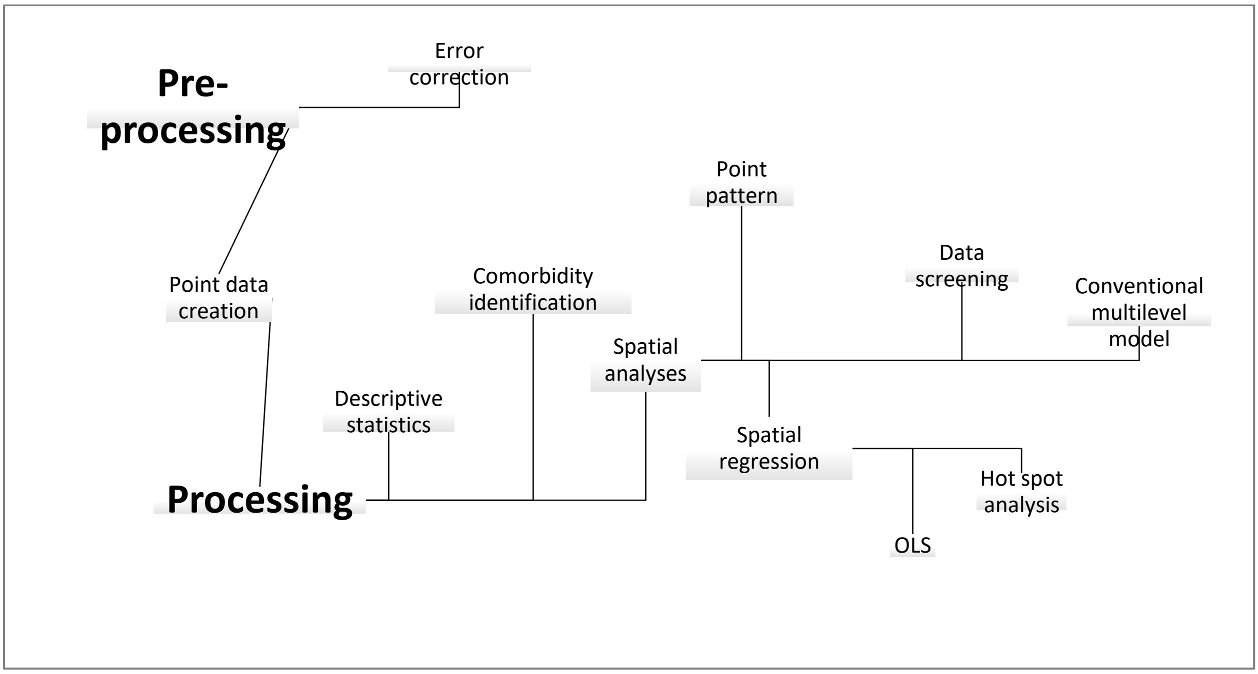


Figure 3. 2. Methodology flow diagram

b. *Processing*

i. To estimate the three most prevalent comorbid health conditions in Gauteng;

Comorbidity prevalence estimated prevalence as a proportion, typically a presented as a percentage throughout all reported tables in the report. Health conditions were disaggregated from sub-place unit and transformed into point data to represent combinations of two, three and four-or-more comorbidities. Data processing was applied, using Excel functions, in order to create random points corresponding to respondents with comorbidity. Random points corresponded to the exact number of questionnaires, which represented the respondents of the QoL Survey. These XY geographic locations were calculated as the sum of health problems in groups of two, three and four-or-more to form comorbidities. Comorbidity has been ranked to illustrate what each disease accounts. The 95% CI is a range of values that that has a 95% certainty that the sample contains the true mean of the population. 95% CI as a mean statistic has been used for comorbidity data as it derives the central tendency of population data. Greenland *et al* (2016) states that the 95% CI has the ability to provide the range of plausible value estimate of the population parameter. As the nature of this research highlights the probability distribution of comorbidity, the CI is able to estimate population proportion from categorical variables. Similar to the Sahai and Khurshid (1993) review, this report also highlights the usage of the Poisson distribution of the 95 % CI as an appropriate model for probability distribution.

Comorbidity combinations represented the respondent's health into groups of two, three and four-or-more comorbid health conditions. The top 5 comorbidities were selected within each combination but only the highest frequency comorbidity was chosen within each group for further analysis. These combination groups represented the most occurring comorbidities in Gauteng and were considered as the comorbidity variable for analysis in this research report. Because health condition disorders precede the previous condition, as they share a common risk factor between both comorbid health conditions, two comorbid health conditions are likely to be present in other comorbidity classes. The first objective aligned with the research question based on how comorbidity is distributed.

ii. To identify spatial dependency of comorbid health conditions in Gauteng;

The comorbidities were then located and analysed. Spatial dependency was identified using Global Moran's I spatial autocorrelation method. Each comorbidity class is located across Gauteng. However, simply locating the comorbidity is not accurate enough as it does not give sufficient information about point dependencies. According to Goodchild (1992) spatial dependence is the tendency of nearby locations to influence one another and possess similar characteristics. Therefore, an appropriate statistical measure is to use a spatial autocorrelation method, which tests for similarity of attributes. Global Moran's I is used to test this. It is important to detect spatial autocorrelation, as it reinforces the notion of spatial dependency and patterns among data. The Moran's I Index value, p-value and z-score evaluated the significance of the index. The pattern expressed random, clustered or dispersed associated with attribute values. The spatial dependency of the comorbidities indicated that the nearby sub-places were more similar than those in distant neighbourhoods. ArcGIS Pro (2018) states that since Global Moran's I is an inferential statistic, the results of analysis will always be interpreted within context of the null hypothesis. The null hypothesis stated that the attribute is as a result of spatial randomness. Global Moran's I statistic was used to illustrate point dispersion and arrangement. The use of the spatial autocorrelation method of Moran's I indicated the statistical measure of spatial dependency of point data. OLS models were tested for spatial analyses. The results stipulated that there was spatial dependency bias. The Moran's I and both the z-score and p-value were used to evaluate the significance of that index. As Moran's I is an inferential statistic, the statistical significance must be determined first using a hypothesis test of the p-value and z-scores. A statistically significant p-value indicated that the observed pattern was random by chance. The interpretation of the z-score illustrated high/low clustering of the observed pattern.

Spatial analysis was conducted using point pattern analyses. The point pattern analyses explained the comorbidity distribution, density of point clustering and point intensity and frequency. Point pattern analyses is an important methodology as it describes the comorbidity pattern across Gauteng. A geometric mean was used to normalise the dataset in order to create more accurate maps. Point pattern analyses, such as frequency, density, dispersion and arrangement, were also used to illustrate the relationship between respondents and the effects on health status in Gauteng. Point patterns are an evolution of pattern and point distributions. Point pattern analyses can refer to spatial or temporal locations. To analyse the occurrence of point patterns; point frequency, intensity or abundance must be considered. These describe the density of the pattern, the centre of the pattern and how dispersed these points are. Point density calculates the magnitude per square meter area features that fall within Gauteng sub-places.

Hotspot analysis was used to detect high and low values of spatial clustering within two comorbidities. Although hotspot results illustrated spatial clustering, they did not indicate why clustering occurred in those locations. This type of spatial analysis examined the aggregation of points of occurrence based on a distance-function. When testing for the presence of clustering, Moran's I must be tested to show statistically significant spatial autocorrelation. Hotspot analysis requires the presence of clustered data. Hot spots are the statistically significant clusters of features with high values, whereas cold spots are statistically significant clusters with low values. This analysis reiterates Tobler's first law of Geography stating that "all things are related, but near things are more related than distance things". This method has been used for incident-based mapping for crime, traffic congestion and population data.

The second objective, aligned with the second research question, was to determine which comorbid health conditions are spatially dependent. Spatial dependency of each comorbidity class highlighted that, if comorbidity was a random pattern and observations were independent from one another, then the observed pattern could be interpreted with a spatial context. If spatial dependence illustrated a dispersed pattern, and observations were not independent of one another, then the observed pattern could not be interpreted with a spatial context.

iii. To determine the multilevel model factors associated with 3 most prevalent comorbidities in Gauteng.

Objective three produced results in line with the proposed research questions based on conventional multilevel model effects on Gauteng' s health. Multilevel models were determined by examination of Anselin (2005) spatial regression decision process. ArcMap, RStudio and Excel generated the statistical outputs for this research report. Maps were produced in ArcMap and plots and graph were produced in the R environment. It is because regression methods have the ability detect variables associated with the phenomenon, an OLS regression was used to as a starting point for regression techniques. It provided a global model for comorbidity analyses and explanatory variables for Gauteng. As OLS is a non-spatial model and does not account for space, so it was simply used to explain the relationship between comorbidity and its explanatory variables. The equation for an OLS model is $y = X\beta + \varepsilon$, where the regression coefficients (β) are the computed regression components that represent the strength and type of relationship the explanatory variables have to the dependent variable. Although OLS model is a linear model and can only predict linear relationships within the data, and comorbidities contain spatial XY data, OLS must be conducted as a pre-requisite to performing spatial regression. Without disproving a linear relationship within the data, the spatial component cannot be proven to exist. Adjusted r squared, Akaike information criterion (AIC), Bayesian information criterion (BIC), standard residual error and the pr(<t) values were examined to determine measure of quality of linear regression fit and probability of a relationship between predictor and response variables. A suitable OLS model indicated the appropriate spatial regression model.

The spatial regression models were compared, and they addressed the spatial dependency among the comorbidities. Spatial dependence was determined from the results of conducting LaGrange multiplier tests for spatial lag and spatial error dependencies. The results from the OLS regression model highlighted the comorbidity relationship with explanatory variables statistics. As per Anselin's spatial regression decision process, LaGrange multiplier diagnostics were examined when there was an absence of a linear relationship within the data. Only if spatial dependencies were absent, OLS models would be retained. Matthews (2006) states that spatial error models evaluated the influence of unmeasured explanatory variables with reference to the clustering of error terms. In contrast, spatial lag models incorporated the unmeasured explanatory variables as well as the additional effect of neighbouring attribute values. Florax and Nijkamp (2003) state that ignoring spatial interaction can create serious model misspecification, therefore it is an important aspect of regression modelling.

Park and Kim (2014) prefer a multilevel model approach as it enables researchers to examine the hierarchal structure of variables. However, this research report incorporates the results from spatial regression as a follow-up to the step-by-step process for all regression models. Best *et al* (2000) studies the disease incidence to individual level variables that geographically varied. The community level in which the ecological regression methods are used relate geographical variations of disease occurrence to the geographical distribution of suspected risk factors. Multilevel modelling was produced through the two conventional models (intercept-only model and generalized least squares fit by a maximum likelihood model) across all comorbidity classes in the R environment. The intercept for explanatory variables was assumed to be random in this study. Babyak (2004) states that size of dataset hinders the complexity of modelling as overfitting in regression analysis is evident in smaller datasets (Babyak 2004). Before conventional multilevel models were produced, data screening was applied to test if data met assumptions for linear modelling. Examination of the linearity and normality plots and graphs determined whether or not linear multilevel modelling was a suitable analysis, and if conventional multilevel modelling could follow. Linearity examined the normal q-q- plot for data linearity; and normality tested to see if the histogram of standardized residuals were centred around zero.

Conventional multilevel models produced only the significant explanatory variables estimate results associated with comorbidity. Estimates defined which categorical variables were more likely to report comorbidity in Gauteng, with reference to estimate, p-value and reference category. Comorbidity from small data sets were excluded from multilevel analysis. Conventional multilevel models were categorised as model 1 (intercept-only model) and model 2 (generalised linear mixed-effects model), respectively. Models' 1 could not be compared as intercept-only models contained no explanatory variables, therefore models' 1 were examined on their own. The third objective, aligned with the third research question, determined the factors associated with the three most prevalent comorbidities within the multilevel model. Only significant factors were compared.

Chapter IV: Results

It is important to note that, as a limitation of GCRO data, 70 interviews (out of all the possible total respondents) did not answer all questions. It should be noted that maps were labelled individually, and similar labelling did not correlate with corresponding maps in this report. As part of pre-processing, figure 4.1 illustrates data as a map depicting the point locations of respondents in the sub-place area-unit of Gauteng from the QoL Survey 2015.

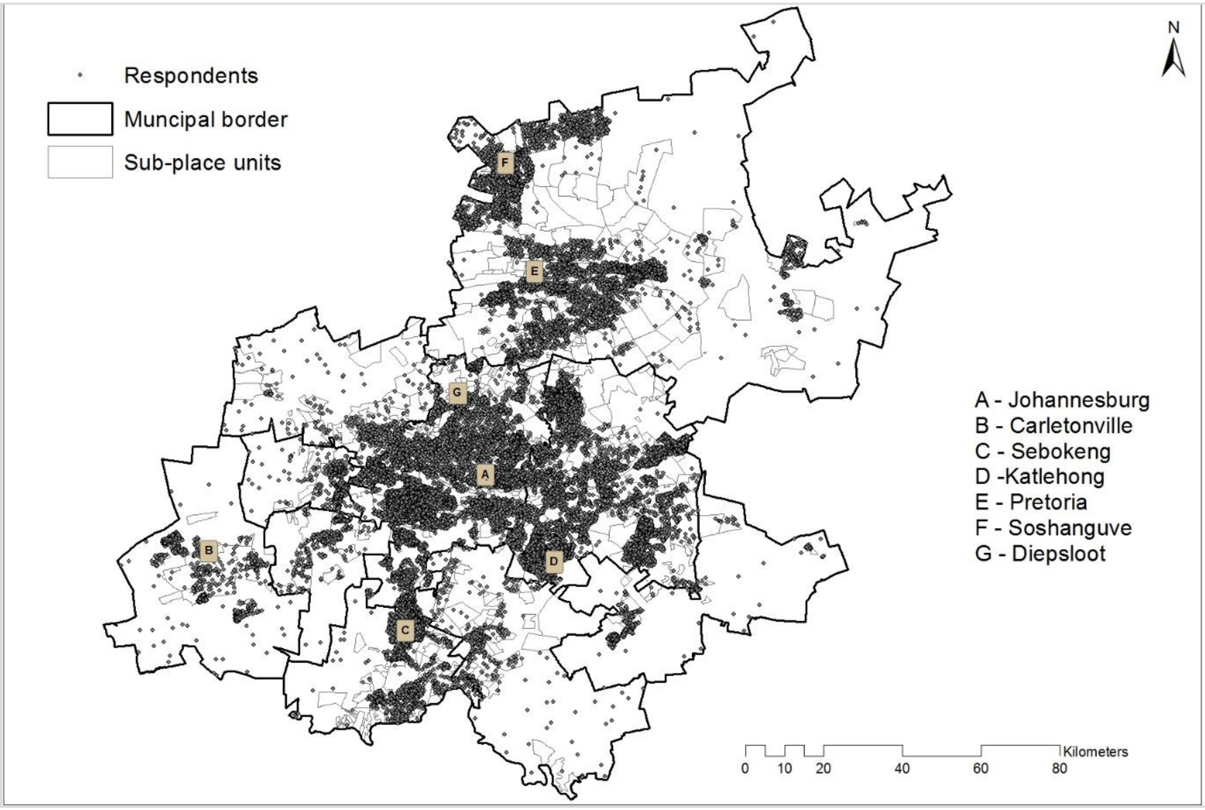


Figure 4. 1. Distribution of randomly created respondent points per sub-place area in the Gauteng province. Johannesburg, Pretoria illustrates a high distribution of comorbidity in a city; Carletonville, Sebokeng, Katlehong, Soshoanguve and Diepsloot illustrate high distribution of comorbidity in a township.

As previously stated, it was vital to explore the data before further analyses could be done, as the inappropriate method may result in inaccurate outcomes. Descriptive statistics were produced in an Excel environment. It is important to note that the descriptive statistics represented singular health conditions, not comorbidity values. Although this report focused on comorbidity prevalence, these results outlined the underlying socioeconomic factors that contributed to the general health of Gauteng.

Figure 4.2 highlights the percentages of health conditions in the QoL Survey 2015. Hypertension, influenza/pneumonia and diabetes presented as the three most selected health conditions of the respondents This may have influenced the comorbid health conditions as the most frequently selected health conditions were found to be among comorbidities.

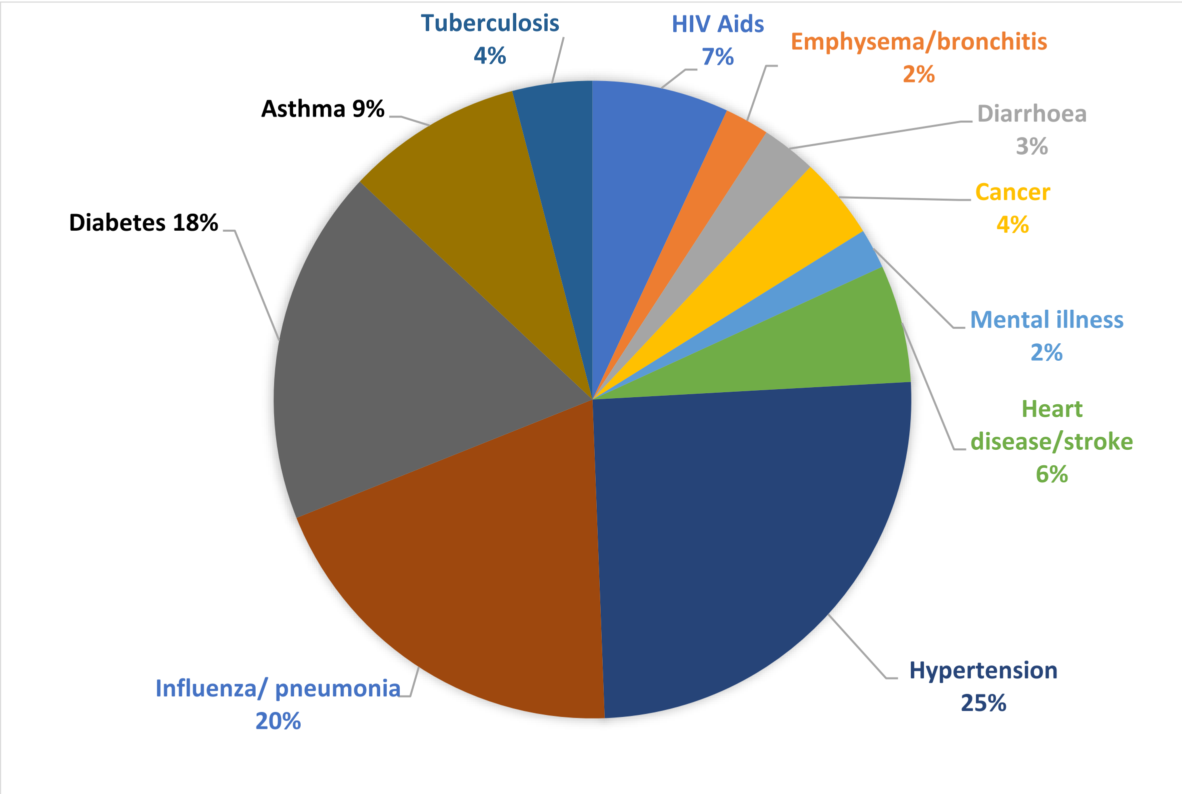


Figure 4. 2 Percentages of singular health conditions

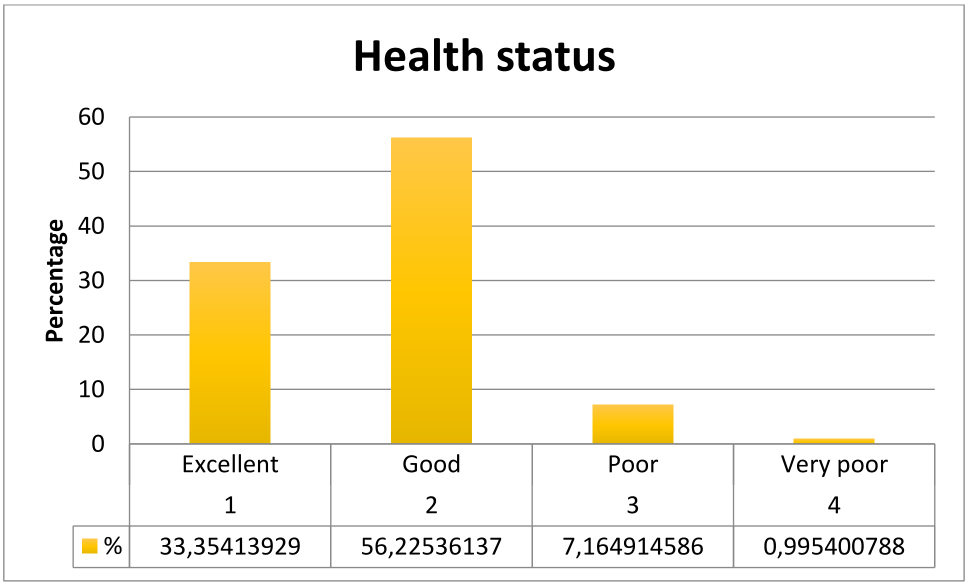


Figure 4. 3 Percentage of self-reported health status condition for all QoL population.

The health status of the respondents was conducted over a four-week period before the survey and highlighted health conditions that affected the individuals at that time. The health status was categorised as *excellent, good, poor or very poor*. The bar graph illustrates that in the past four weeks, the health status was good for more than half of the population in the QoL Survey 2015 (figure 4.3). Health status outcome could have significant impacts on prevalence of comorbidity as it explores the underlying conditions of health in a community. Although health status was revealed as a measure of *good,* this result will be used in juxtaposition of the final results of comorbid health in Gauteng. Descriptive statistics indicate that majority of respondents are between 25-34 years of age, female, Black African, employed full-time in the formal sector, possess high school level of education and were born in Gauteng (table 4.1).

Table 4. 1 Descriptive statistics for demographic variables. Data source QoL Survey 2015.

Variable	Recoded	N	%	Variable	Recoded	N	%
Age				**Gender**			
18-19	1	1196	4.0	Male	1	14052	46.68
20-24	2	3449	11.6	Female	2	16051	53.32
25-29	3	4165	14.0	**Population group**			
30-34	4	4064	13.7	African	1	24227	80.53
35-39	5	3833	12.9	Coloured	2	1156	3.84
40-44	6	3007	10.1	Indian/Asian	3	634	2.11
45-49	7	2600	8.7	White	4	3944	13.11
50-54	8	2053	6.9	Other	5	125	0.42
55-59	9	1754	5.9				
60-64	10	1305	4.4				
65+	11	2346	7.9				
Employment status				**Education**			
Employed full time, formal sector	1	7909	25.98	No education	0	490	1.61
Employed part time, formal sector	2	1813	5.59	Gr 1 – 7 (primary school)	1-8	3259	11.26
Employed full time, informal sector	3	1813	5.59	Gr 8 – 12 (high school)	9-13	18746	60.58
Employed part time, informal sector	4	1563	5.13	A certificate from a college, Technikon or university	14	2492	8.14
Self-employed, own business, NOT working from home	5	636	2.09	A diploma from a college, Technikon or university	15	2091	6.87
Self-employed, own business, working from home	6	745	2.45	Technikon or university degree	16	1334	4.38
Non-respondents	.	15373	50.50	Post graduate degree - e.g. Hons, MA, PhD	17	763	2.50
Migration status				Unspecified	18	586	1.92
Born in Gauteng	1	19234	63.17				
Migrated into Gauteng from a province in South Africa	2	8539	28.05				
Migrated into Gauteng from another country	3	2341	7.69				
Non-respondents	.	146	0.48				

i. Identification of comorbidities

The top five comorbidities are selected within each combination class using exact binomial calculation for the 95% CI (table 4.2). Comorbidity class represented the comorbidity groups which were ranked in line with count of comorbid condition. This process fulfilled objective one of this research report and was essential to preforming spatial dependency of objective two, as well as in reporting the associated factors in objective three. The highest frequency of comorbidity class was *two comorbidities* of which made up 8.97% of comorbidities.

Lalkhen and Mash (2015) stated that NCD's such as hypertension had the highest rates of comorbidity; whereas multimorbidity was most common in patients with chronic obstructive pulmonary diseases (COPDs) and osteoarthritis. The high frequency of hypertension-combinations illustrates Lalkhen and Mash's statement (figure 4.2). Total comorbidity cases accumulate to a small total of 12.947% for two, three and four-or-more comorbidities (table 4.2).

Table 4. 2. Comorbidity combinations using exact binomial distribution

Comorbidity class	Rank	N	% of comorbidity within total population sample	95 % CI	
				Lower	Upper
TWO		2684	8.97	0.00321	0.00329
Hypertension & Diabetes	**1**	**829**	**2.77**	**0.017**	**0.018**
Hypertension & Influenza/Pneumonia	2	247	0.83	0.011	0.012
Diabetes & Influenza/ Pneumonia	3	142	0.47	0.00816	0.00915
Hypertension & Asthma	4	133	0.44	0.0079	0.0089
Asthma & Influenza/ Pneumonia	5	107	0.36	0.00708	0.00811
THREE		903	3.017	0.00191	0.002
Hypertension, Diabetes & Influenza/ Pneumonia	**1**	**86**	**0.29**	**0.018**	**0.0210**
Hypertension, Heart disease/ Stroke & Diabetes	2	78	0.26	0.018	0.02
Hypertension, Diabetes & Asthma	3	67	0.22	0.016	0.019
Hypertension, Diabetes, TB	4	23	0.08	0.00925	0.012
Hypertension, Diabetes & Cancer	5	10	0.03	0.00856	0.012
FOUR-OR-MORE		286	0.96	0.00107	0.00117

Hypertension, Heart disease/Stroke, Diabetes & Asthma	**1**	**16**	**0.05**	**0.024**	0.033
Hypertension, Diabetes, Asthma & Influenza/Pneumonia	2	15	0.05	0.023	0.033
Hypertension, Heart Disease/Stroke, Diabetes & Influenza/Pneumonia	3	12	0.04	0.020	0.033
Hypertension, Heart Disease/Stroke, Diabetes & Influenza/ Pneumonia/Hypertension, Heart Disease/Stroke	4	8	0.02	0.016	0.026
Hypertension, Heart Disease/Stroke, Diabetes & Influenza/ Pneumonia	5	5	0.01	0.012	0.023

Table 4.3 illustrated the descriptive of each comorbidity class in this research report. Individual-level variables highlighted *population group, gender, age*, *health status, employment, income, place of birth, education, social grant, perceived level of stress exercise frequency and healthcare provider*. Community-level variables demographics highlighted *community health problems*, *community social problems*, *number of people in household, improvement or deterioration the past 12 months, and reason for living in your suburb*. Comorbidity class represents the comorbidity groups of which have been ranked in line with count of comorbid condition. The relationship among comorbid conditions and explanatory variables will be examined in conjunction with the multilevel model analysis.

Table 4. 3 Description for comorbidity classes. Data source QoL Survey 2015.

	Two comorbidities	**%**	**Three comorbidities**	**%**	**Four-or-more comorbidities**	**%**	**Total**	**%**
N	829	100	86	100	16	100	931	100
Population group: ***Black African***	829	100	86	100	16	100	931	100
Male	304	36.67	38	44.19	9	56.25	351	37.70
Female	525	63.33	48	55.81	7	43.75	580	62.30
Age: **55 – 59 years**	89	10.74	7	8.14	1	6.25	97	10.42
60 – 64 years	98	11.82	6	6.98	0	0	104	11.17
65 + years	224	27.02	23	26.75	7	43.75	254	27.28
Health status: ***Excellent***	142	17.13	13	15.12	3	18.75	158	16.97
Good	500	60.31	55	63.95	6	37.50	561	60.26
Poor	177	21.35	18	20.93	7	43.75	202	21.70

Very poor	14	1.69	0	0	0	0	14	1.50
Biggest social problem facing community: ***Drugs***	157	18.94	11	12.79	1	6.25	169	18.15
Crime	269	32.45	33	38.37	2	12.50	304	32.65
Unemployment	144	17.37	15	17.44	3	18.75	162	17.40
Poverty	22	2.65	1	1.16	2	12.50	25	2.69
Foreigners	13	1.57	2	2.33	2	12.50	17	1.83
High cost of living	28	3.38	4	4.65	2	12.50	34	3.65
Employment status: ***Employed***	253	30.40	27	27.91	7	43.75	287	30.83
Unemployed	151	18.09	22	25.58	3	18.75	176	18.90
Other	428	51.51	40	46.51	6	37.50	474	50.91
Income bracket: ***R1 601- R3 200***	135	16.28	24	27.91	1	6.25	160	17.19
R3 201 – R6 400	104	12.55	12	13.95	1	6.25	117	12.57
Respondent refused	177	21.35	10	11.63	3	18.75	190	20.41
Place of birth: ***Born in Gauteng***	592	71.41	57	66.28	10	62.50	659	70.78
Migrated from another province	215	25.93	25	29.07	6	37.50	246	26.42
Migrated from another country	22	2.65	4	4.65	0	0	26	2.79
Education: ***Grade 10***	71	8.56	11	12.79	5	31.25	87	9.34
Grade 12	205	24.73	24	27.91	3	18.75	232	24.92

Certificate from college, Technikon or institute	60	7.24	2	2.33	3	18.75	65	6.98
No. of people in household: **1-3**	376	45.35	33	38.37	6	37.50	415	44.58
4-6	333	40.17	42	48.84	8	50.00	383	41.14
7+	120	14.48	11	12.79	2	12.50	133	14.29
Receive social grant: ***No***	281	33.89	27	31.40	3	18.75	310	33.30
Yes	548	66.10	58	67.44	13	81.25	619	66.49
Exercise frequency: ***Everyday***	180	21.71	18	16.28	2	12.50	200	21.48
Few times a week	191	23.04	16	18.60	5	31.25	212	22.77
Few times a day	88	10.62	13	15.12	3	18.75	104	11.17
Hardly ever	124	14.96	12	13.95	3	18.75	139	14.93
Never	246	29.67	27	31.40	3	18.75	276	29.65
Improvement of deterioration in the past year: ***Improvement***	232	27.99	28	32.56	6	37.50	266	28.57
Deterioration	129	15.56	12	13.95	0	0	141	15.15
No change	468	56.45	46	53.49	10	62.50	524	56.28
Reason for living in your suburb: ***Always lived here***	229	27.62	28	32.59	7	43.75	264	28.36
Affordability of property	275	33.17	22	25.58	7	43.75	304	32.65
Security	12	1.45	3	3.49	3	18.75	18	1.93

Perceived level of stress: ***zero means they did not experience feeling "at all"; ten means they experienced feeling "all of the time':*** ***Category -0***	0	0	0	0	2	12.50	2	0.21
Category -1	214	25.81	32	37.21	3	18.75	249	26.75
Category -2	107	12.91	10	11.63	2	12.50	119	12.78
Category -3	76	9.17	6	6.98	2	12.50	84	9.02
Category -4	110	13.27	4	4.65	1	6.25	115	12.35
Category -6	76	9.17	12	13.95	1	6.25	89	9.56
Category- 10	31	3.74	2	2.33	0	0	33	3.54
Healthcare provider: ***Public healthcare facility***	575	69.36	53	61.63	8	50.00	636	68.31
Use public and private facilities	51	6.15	4	4.65	1	6.25	56	6.01
Private healthcare facilities	169	20.39	17	19.77	7	43.75	193	20.73
Traditional healers	2	0.24	0	0	0	0	2	0.21
Spiritual healer	2	0.24	0	0	0	0	2	0.21
Not applicable, dont usually need healthcare	30	3.62	2	2.33	0	0	32	3.44

ii. Spatial analysis

This section examined spatial analyses such as point patterns, distributions and point density, as well as the spatial regression models like OLS and multilevel modelling. The spatial distribution of comorbidities across Gauteng analysed the relationship between comorbidity classes and its predictors. Figure 4.4 highlights the distribution of the highest number of cases of two, three and four-or-more comorbidities. Clustering cannot be viewed in the distribution map alone and will therefore be further analysed in the sections to follow.

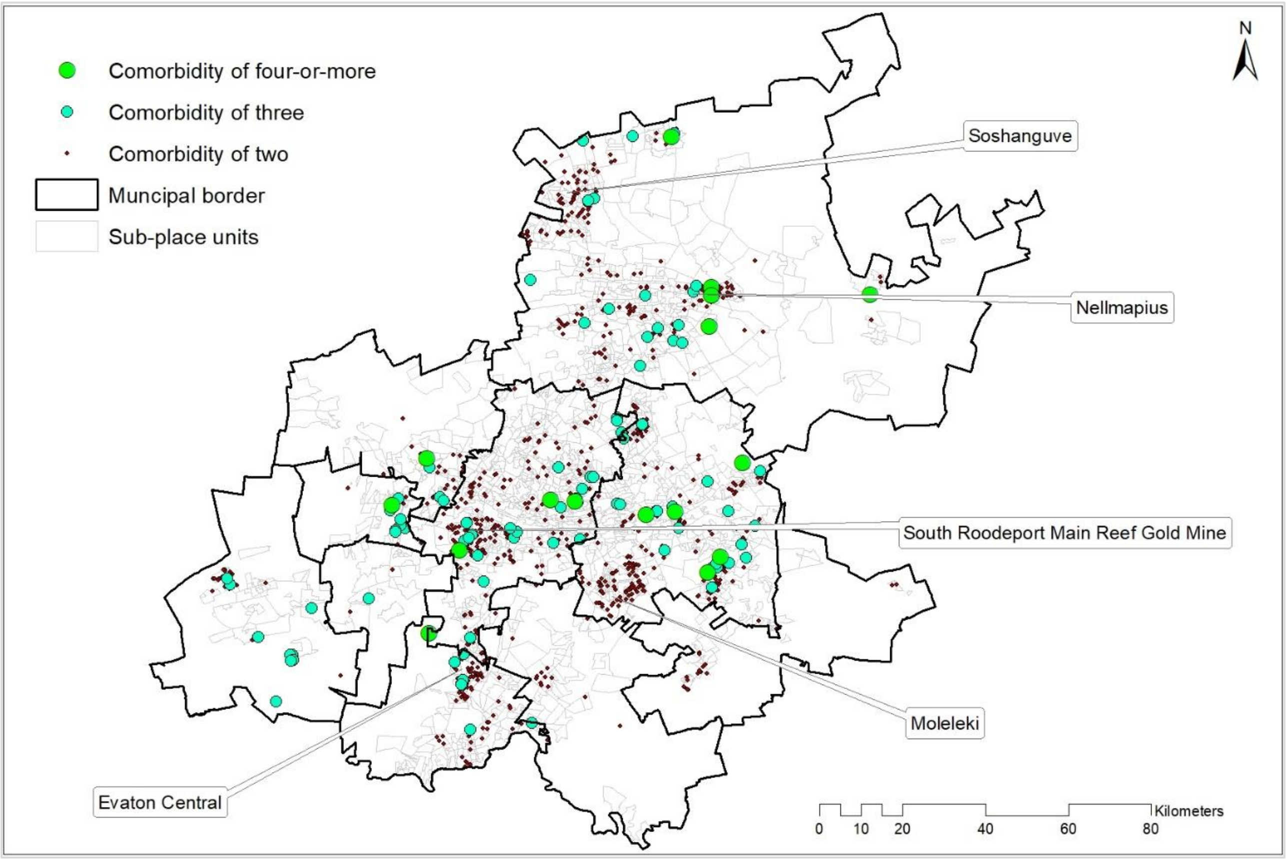


Figure 4. 4 Distribution of two, three and four-or-more comorbidities. Moleki and Soshanguve represented two comorbidity prevalence; Evaton Central and South Roodeport Main Reef Gold Mine represented three comorbidities; and Nellmapius represented four-or-more comorbidities.

a. Point Pattern Analyses

Frequency and Density

As previously stated, the frequency of comorbidities, indicates the number of respondents with comorbidities (table 4.2). A density surface was computed and visualised as a continuous surface where the points were concentrated in Gauteng. The point density of comorbidities was illustrated as per the respondent's self-reported health.

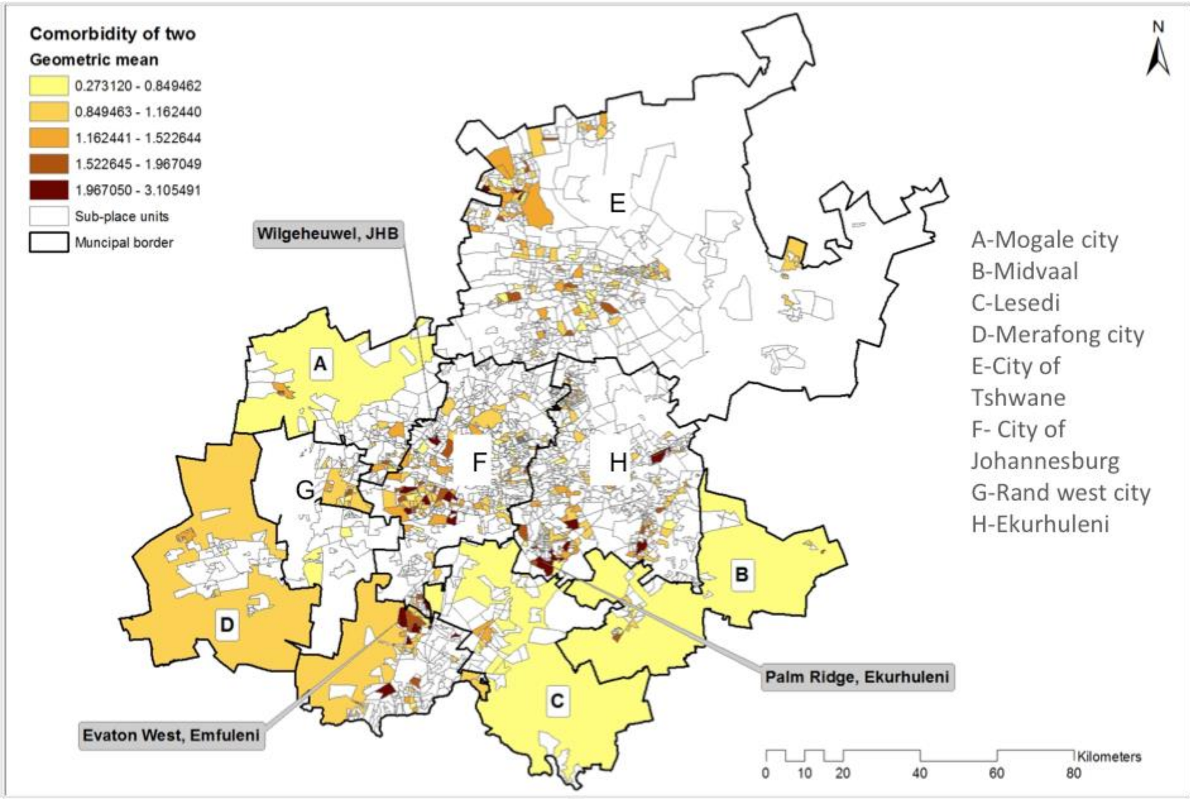


Figure 4. 5 Geometric means distribution for two comorbidities. Geometric means highlighted the central tendency of two comorbidities where Palm Ridge, Evaton West and Wilgeheuwel illustrated highest density of comorbidity prevalence.

The geometric mean calculates the normalized count of comorbidities per sub-place (figure 4.5) for two comorbidities. The higher the geometric mean the higher the density per square kilometer of comorbidity prevalence. Table 4.3 was used in the discussion of results for two, three and four-or-more comorbidities. Two comorbidities illustrated the point density per square kilometer in cases of hypertension and diabetes in Gauteng. *Natural breaks classification* was used to minimize variation within classes and maximize variation between classes. Wilgeheuwel, Evaton West and Palm Ridge were some of the areas highlighted as a high-density locations for two comorbidities. It may be important to note that Mogale City, Midvaal and Lesedi foster low density areas for two comorbidities.

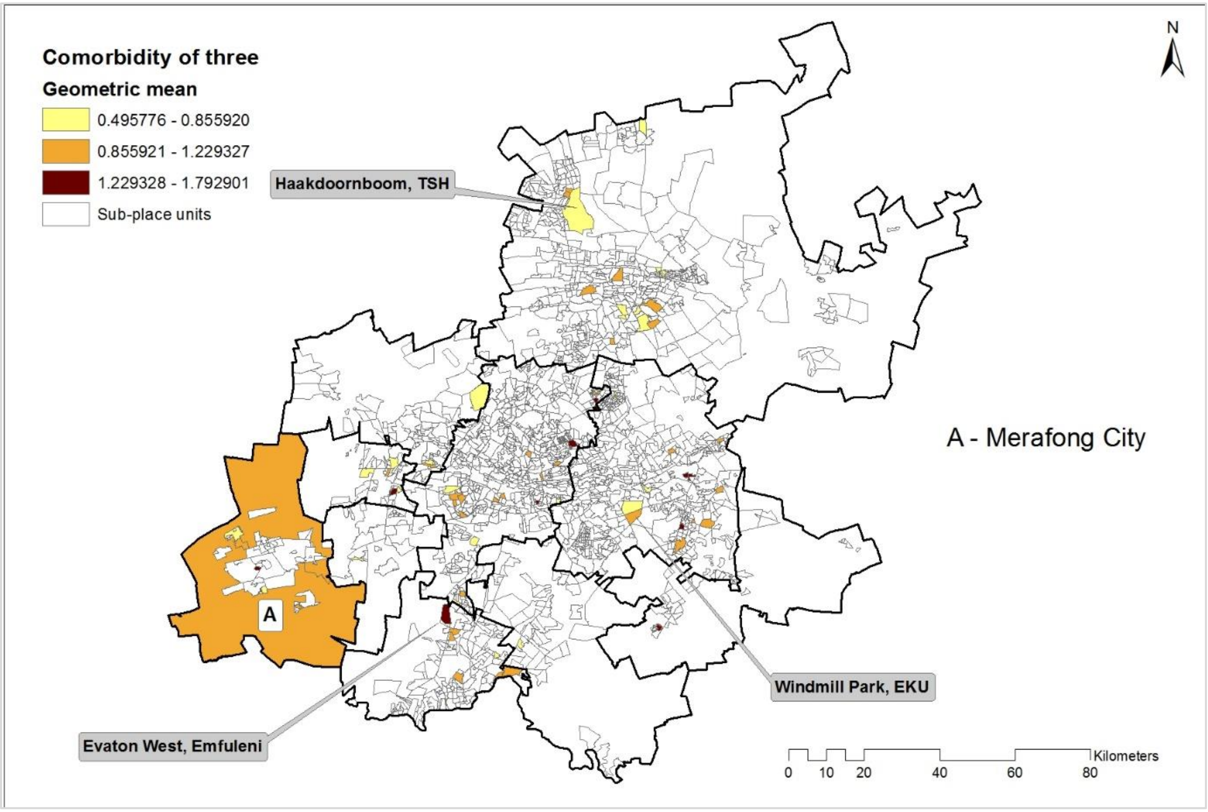


Figure 4. 6 Geometric means distribution for three comorbidities. Evaton West illustrated high point density; Merafong City, Windmill Park and Haakdoornboom illustrated areas of low density for three comorbidities.

Evaton West presents a high frequency of three comorbidities (figure 4.6). Windmill Park and Haakdoornboom presented as areas of low frequency, and Merafong City continued to be an isolated area with little to no frequency for three comorbidities. It is interesting to note the juxtaposition of high and low areas of comorbidity were in close proximity to one another, which emphasised the disparity of health in Gauteng. The independence of health status of each sub-place is again evident when noting where high and low frequencies of comorbidity were dispersed across the province.

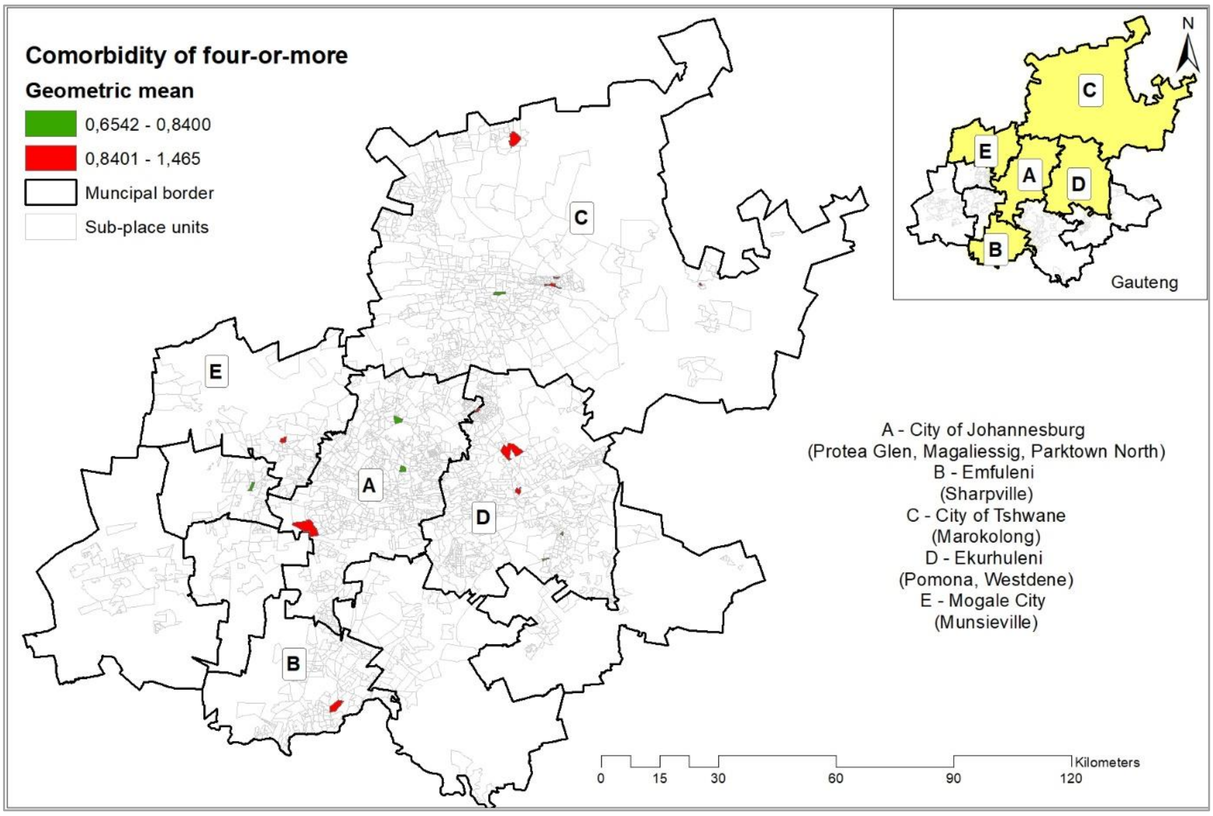


Figure 4. 7 Geometric means distribution for four-or-more comorbidities .

Examination of data at a sub-place level affected the resolution of the map owing to the small geography of sub-place demarcation. Marokolong, Magaliessig and Parktown North are characterised as high frequency areas of four-or-more comorbidities, as compared to Protea Glen, Sharpeville, Pomona and Westdene which reflect as areas of low frequency of four-or-more comorbidities (figure 4.7).

Point dispersion and arrangement

H_0: Complete Spatial Randomness
H_a: spatial distribution is more spatially clustered than would be expected if the underlying spatial processes were random.

The Moran's I result for *two comorbidities* indicated that, given the positive z-score (**1.502896)**, data was spatially clustered in some way. The p-value (**0,132866**) was positive and not significant, therefore the pattern supported the null hypothesis that the spatial distribution was a result of spatial randomness.

The Moran's I result for *three comorbidities* illustrated that given the negative z-score (**-0.328319),** data was clustered in a competitive way, where high values may have been repelling one another, and negative values may have been repelling one another. Although the p-value (**0,019977**) was significant and z-score was negative, the null hypothesis was rejected as a dispersed pattern (which often reflected when a type of competitive process occurs).

The Moran's I result for *four-or-more comorbidities* illustrated that given the negative z-score (-**0,124244**), data was clustered in a completive way. Because the p-value (**0,901122**) was not significant and positive, the Moran's I result supported for four-or-more comorbidities the null hypothesis and could have been a version of a random pattern.

Two comorbidities and four-or-more comorbidities illustrated complete spatial randomness (table 4.4). Three comorbidities highlighted a more spatially dispersed pattern than would have been expected if underlying processes were random, therefore three comorbidities was excluded from spatial analyses in the sections to follow. A Moran's I result of complete spatial randomness meant that two and four-or-more comorbidities were the only comorbidity classes which were further examined in the sections to follow.

Table 4. 4 Moran's test for comorbid health conditions. Moran's Index, z-score and p-values were evaluated for significance of the observed pattern.

	Two comorbidities	Three comorbidities	Four-or-more comorbidities
Moran's Index	0.026151	-0.328319	-0.093099
Z-score	1.502896	-2.326772	-0.12424
P-value	0.132866	0.019977	0.90112
Pattern	random	dispersed	random

Hotspot analysis

Significant *hot and cold spots* of comorbidities at sub-place level in Gauteng are illustrated together with their corresponding CI percentages (table 4.6). A feature with a high value may not be always be a statistically significant hot spot, rather it needs to be a feature with a high value, surrounded by similar features with high values, to be considered a significant hot spot. The CI illustrated that, at each interval, the population mean was contained. The hot spot analysis table is supported by hot spot analysis maps (figure 4.8, 4.9). Three comorbidities have been excluded from hot and cold spot maps as per the aforementioned Moran's I result reflecting a dispersed pattern.

Two comorbidities

Respectively, City of Johannesburg, Emfuleni, Midvaal and Ekurhuleni observed the highest and lowest values of clustering. This result is attributed to the number of observed points of two comorbidities across Gauteng. Locations that were described as “not significant” indicated sub-places where only one respondent was located.

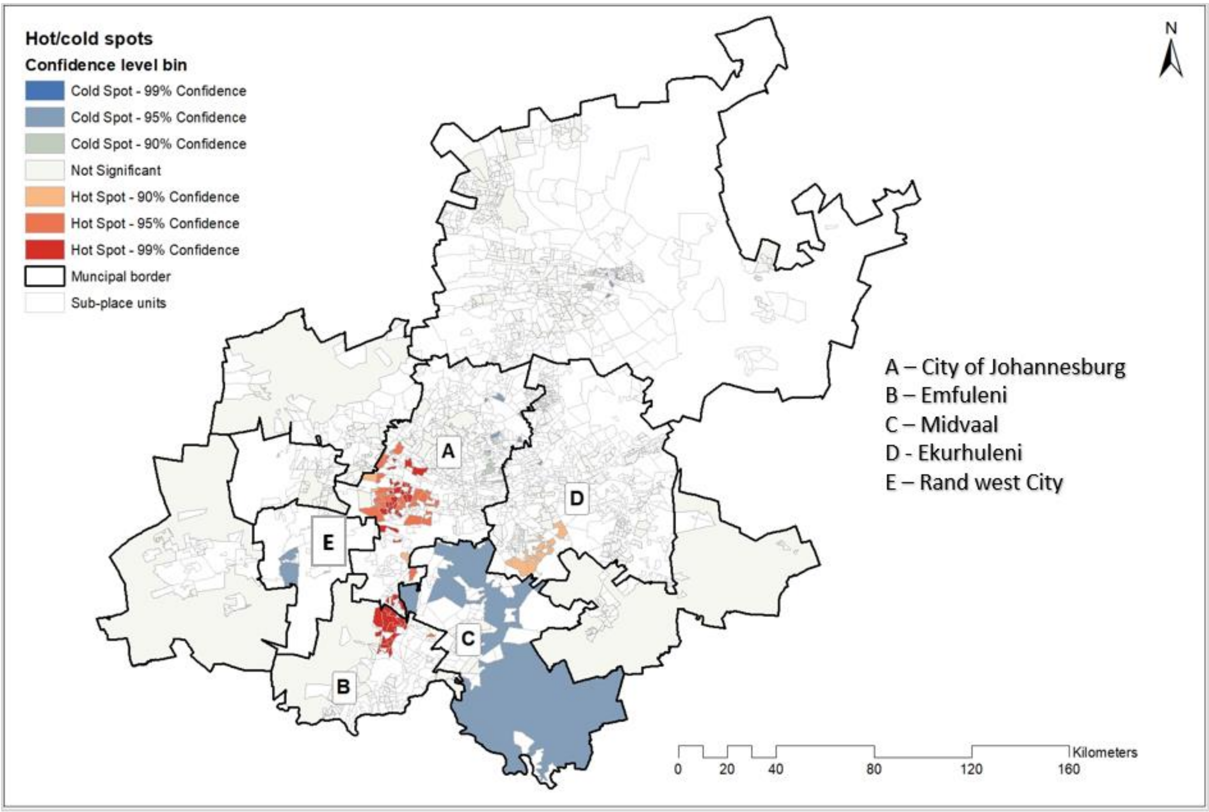


Figure 4. 8 Thematic hot spot map measuring 99, 95 and 90 percent confidence bins for two comorbidities at sub-place level in Gauteng. City of Johannesburg, Emfuleni and Ekurhuleni illustrated hot spots from 90 to 99 percent confidence. Midvaal and Rand West City illustrated cold spots at 95 percent confidence.

Four-or-more comorbidities

Because four-or-more comorbidities contained a small dataset of 16 respondents, pockets of health are scattered across Gauteng sub-places (figure 4.8). Hot spot analysis revealed that only Ekurhuleni municipality comprised of four-or-more comorbidities. *Highland* and *Geluksdal ext 3* were the only locations with statistically significant hot spots at 90% CI. *Pomona* revealed a cold spot at 90% CI.

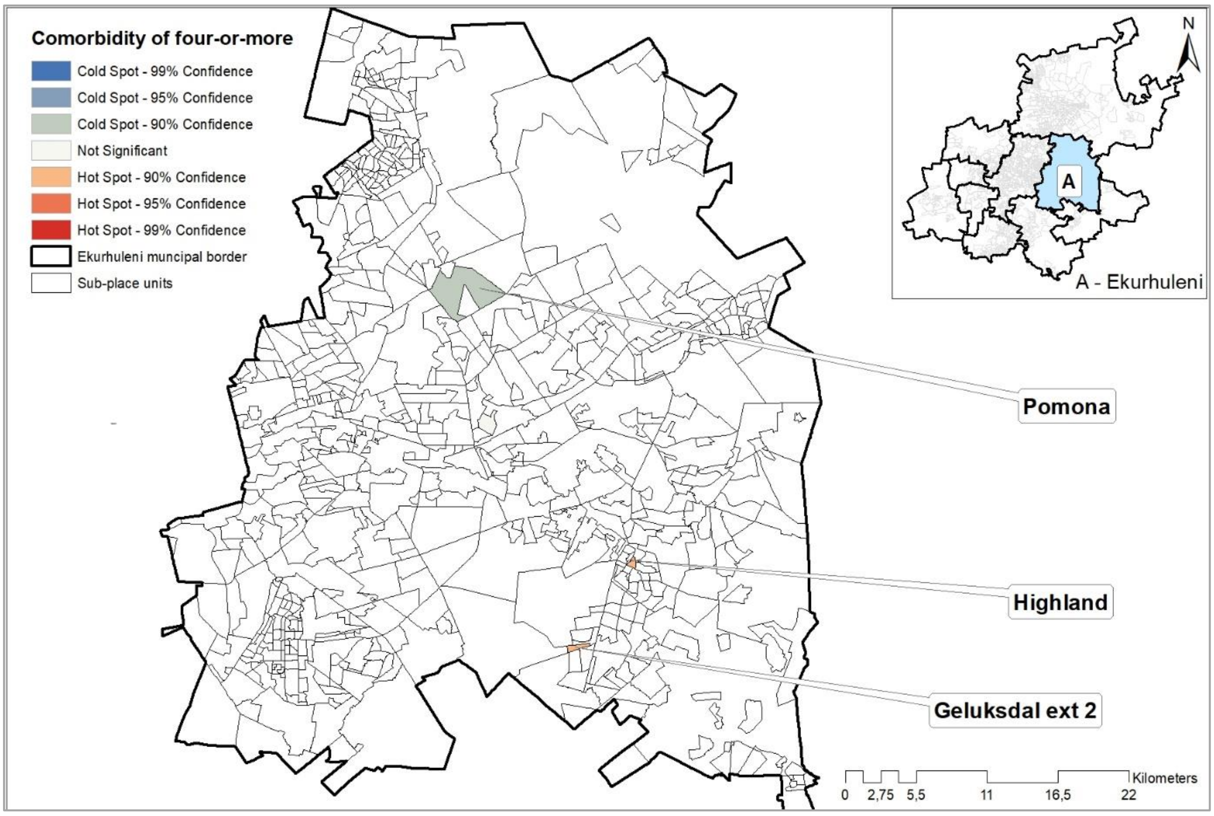


Figure 4. 9 Thematic hot spot map measuring 99, 95 and 90 percent confidence bins for four-or-more comorbidities in region A – Ekurhuleni.

Hot spot analysis illustrated very high and very low clusters of comorbidity prevalence. The CI bins identified statistically significant hot and cold spots, whereby features in the ± 3 bins reflected statistical significance with a 99 percent confidence; features in the ± 2 bins reflected a 95 percent CI; and features in ± 1 bin reflected a 90 percent confidence level. The clustering of features in bin 0 was not statistically significant.

Only Ekurhuleni highlighted hot and cold spots for four-or-more comorbidities. It was interesting to note that Ekurhuleni municipality is highlighted as a common location with comorbidity with two comorbidities, but they did do not share comorbidities in the same sub-places. *Not applicable* locations were regarded as such when the locations reflected an extreme low proportion of respondents. Hot and cold spots are identified by sub-place units for two comorbidities and four-or-more (table 4.5).

Table 4. 5 Spatial clustering identified as hot and cold spots within Gauteng at 90, 95 and 99 percent CI.

Municipality symbol	Municipality	Sub-place unit	CI (%)
Two comorbidities			
			Hot spots
A	City of Johannesburg	Soweto	99 % Confidence
		Eldorado Park	95% Confidence
		Roodepoort	95% Confidence
		Florida	99% Confidence
			Cold spots
		Midrand	99% Confidence
			Hot spots
B	Emfuleni	Evaton	99% Confidence
		Sebokeng	99% Confidence
		Lakeside	99% Confidence
			Cold spots
		Not applicable	None
			Hot spots
C	Midvaal	Not applicable	None
			Cold Spots
		Midvaal NU	99% Confidence
			Hot spots
D	Ekurhuleni	Katlehong	90% Confidence
		Dawn Park	90% Confidence
		Zonkizizwe	90% Confidence
		Nguni	90% Confidence
			Cold spots
		Not applicable	None
Four-or-more comorbidities			
			Hot spots
A	Ekurhuleni	Highland	90% Confidence
		Geluksdal ext 2	90% Confidence
			Cold spots
		Comet	90% Confidence

iii. Regression

Because spatial dependencies existed within two and four-or-more comorbidities, a non-spatial OLS model was produced to predict values by using explanatory variables to identify the strength of the relationship of each comorbidity. The p-value from Moran's I results was used as an indicator of validity for spatial regression. This research report adopted the Anselin (2005) spatial regression decision process for regression analyses (figure 4.10).

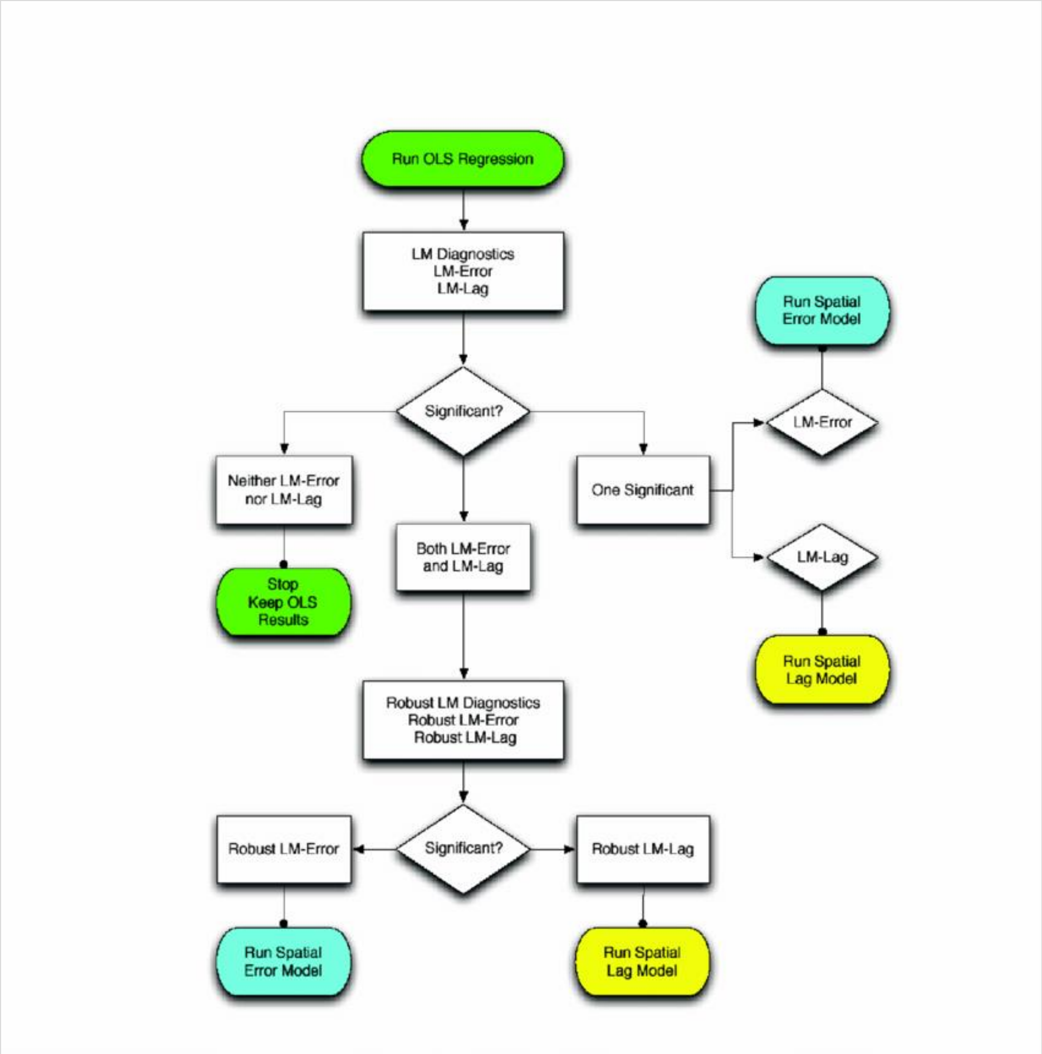


Figure 4. 10 Spatial regression decision process (Anselin 2005)

OLS models

Following Anselin (2005) spatial regression decision process, results from the OLS regression model highlighted the comorbidity relationship only with significant associated explanatory variables (table 4.6). Significance was reported as $p < 0.05$*.

Table 4. 6 OLS regression model depicting only the variables significantly associated with comorbidity

Two comorbidities	
Variables	**Estimate**
Constant	3.618803
Population group	-
Gender2 (female)	0.273677*
Exercise frequency4 (hardly ever)	-0.498069*
Health status4 (very poor)	1.205473*
Water quality2 (usually receives clean water)	-0.557433*

Adult malnutrition3 (sometimes goes hungry when there is no food at home)	0.695271*
Stress9 (feels stressed most of the time)	0.768431*
Test on spatial dependence	
Lagrange Multiplier (Lag)	7.095 (p-value)
Lagrange Multiplier (error)	**0.0003798**
Summary output: Spatially dependence	
Adjusted r-squared	36.01%
Standard residual error	2.127 on 751 degrees of freedom
AIC	1325.339
Pr(<t)	0.01647
Three comorbidities	
Variable	**Estimate**
Constant	1.131182
Population group	-
Stress10 (feels stressed all of the time)	-1.137308*
Place of birth2 (migrated from another province into Gauteng)	-0.350375*
Employment2 (unemployed)	-0.428323*
Healthcare provider2 (uses public and private healthcare facilities)	0.563693*
Test on spatial dependence	
Lagrange Multiplier (Lag)	0.8345
Lagrange Multiplier (error)	<2.2016
Summary output: No spatial dependence in the residuals	
Adjusted r-squared	7.9015%
Standard residual error	0.3622 on 18 degrees of freedom
AIC	-147.8369
Pr(<t)	0.4173

A classic model was produced as a starting point for all OLS regression. In line with Anselin's theory, two comorbidities highlighted the Lagrange Multiplier error as the only significant LM diagnostic, therefore a spatial error model is produced (table 4.7).

The small p-value from the OLS regression model for two comorbidities rejected the null hypothesis and concluded that there was a relationship between comorbidity prevalence and significant explanatory variables. LM diagnostics stated that, as the p-value was "small" and significant (0.0003798), a spatial error model would improve the fit of the model.

The p-value of three comorbidities illustrated that there was no significant relationship between comorbidity prevalence and explanatory variables. As there was no spatial dependence, the OLS model was deemed to be suitable model for three comorbidities.

Four-or-more comorbidities depicted inconclusive results, as data did not overlap or include explanatory variables, and was therefore excluded from OLS regression models.

Spatial error model

A spatial error model has been produced for two comorbidities, with only the statistically significant factors (table 4.7). Spatial error effects suggested that there was dependence among the errors as OLS estimates become inefficient. There was a link between significant explanatory variables from the spatial error model and the hot spot locations, suggesting that these variables may be present at the hot spot locations.

Table 4. 7 Spatial error model

Variable	Estimate
Constant	0.003291
Household respondents	-0.0427134*
Population group	2.6717577*
Water quality2 (usually receives clean water)	-0.9162573*
Water quality 3 (sometimes receives clean water)	-0.6717928*
Refuse disposal1 (removed from house at least once a week)	-0.5558839*
Refuse disposal3 (placed on communal refuse dump)	-0.9140137*
Energy source2 (Gas or LPG)	1.1487516*
Energy source5 (Candles)	0.8515493*
Reason for living here5 (Quality of property)	-0.3885189*
Reason for living here11 (I can be myself)	-0.3477423*
Child malnutrition1 (never goes hungry when there is no food at home)	0.2269478*
Child malnutrition2 (seldom goes hungry when there is no food at home)	-0.3773808*
Employment1 (employed)	-0.2019985*
Healthcare provider1 (public healthcare facility)	0.2807013*
Healthcare provider3 (Private healthcare facility)	0.3510735*
Healthcare provider4 (Traditional healer)	1.3071834*
Education1 (primary only)	0.56322296*
Exercise frequency1 (Everyday)	0.18746902*

Stress2 (feels very little stress)		-0.23530509*
Stress4 (feels some stress)		-0.51814893*
Stress8 (feels a lot of stress)		-0.38108495*
Gender1 (male)		-0.17765577*
Summary output		
LAMDA: **0.046835** ***p-value (0.10833)***	**Log likelihood**: -1751.427	**AIC:** 3656.9

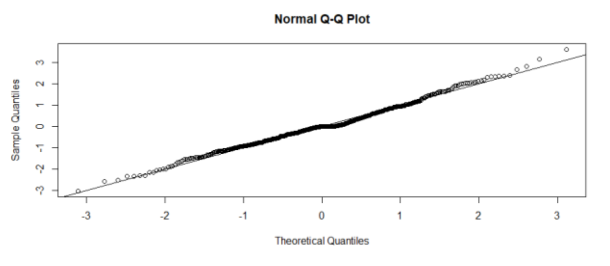
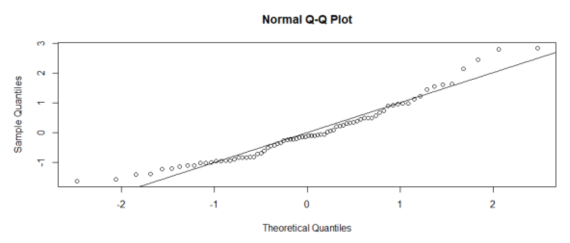
Table 4.8 illustrates the significant explanatory variables for a spatial error model for two comorbidities. By running a spatial error model, standard errors were estimated accurately. However, based on the Hausman test p-value (0.0877), two comorbidities could not, on its own, lead to the conclusion that the spatial error model is the most accurate model. Neither OLS nor spatial error model estimated the coefficients therefore, other models were explored in the sections to follow.

iv. *Multilevel modeling*

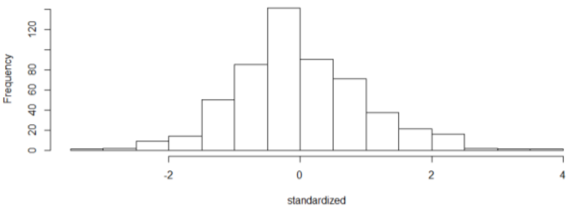
Linear model tests

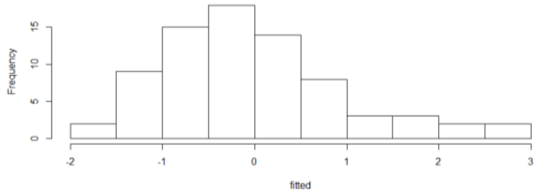
Table 4.8 depicts the data screening for linearity, normality and homogeneity of comorbidities. Linearity examined the normal q-q- plot for data linearity, and normality tested if the histogram of standardised residuals were centred around zero. Two and three comorbidities are tested for linearity and normality as they indicate significant OLS results (table 4.8).

Table 4.8 Linearity and normality plots for linear modelling for two and three comorbidities.

Two comorbidities	**Three comorbidities**
Linearity	
	

Normality





a. Linearity

Two comorbidities illustrated that points followed the trend line closely, and this made for a good model fit. Although three comorbidities illustrated points that followed the trend line, there was indication of a slight deviation from the trend line at the tail ends. This was still a good model fit. Both two comorbidities and three comorbidities indicated that linearity among points was evident.

b. Normality

Two comorbidities illustrated data that was centred around the mean and was normally distributed. Three comorbidities depicted that, although data is slightly skewed to the left, it was still considered a normal distribution. Both two and three comorbidities demonstrated linearity among points and a normal distribution, therefore could recommence with linear conventional multilevel modelling in the section to follow.

Conventional multilevel models

Conventional multilevel models were conducted in the R environment. Two and three comorbidities highlight only the significant explanatory variables from multilevel model analysis (table 4.9). Four-or-more comorbidities continued to be limited by its dataset size and was not illustrated within a conventional multilevel model. Model coefficient estimates were used to compare Model 1 and Model 2 conventional multilevel models.

Table 4. 9 Two comorbidities and three conventional multilevel models. Variable estimates for two comorbidities were used to illustrate significance for each model.

Variables	Conventional multilevel models	
	Model 1: Intercept-only model	Model 2: Generalised least squares fit by a maximum likelihood model
	Estimate	Estimate
Intercept	2.6538	3.618803
Reason for living here5 (quality of property)	-	-1.326993*
Child malnutrition3 (sometimes goes hungry when there is no food at home)	-	-1.266922*
Exercise frequency3 (few times per month)	-	-0.617889*
Exercise frequency5 (never)	-	-0.528381*
Stress9 (feels stressed most to all of the time)	-	0.768431*
Income	-	-0.032490*
Summary output		
AIC	3637.258	3611.939
BIC	3636.699	3459.557
Loglikelihood	-1816.629	-1760.97
Residual standard error	2.164981	0.7107936
Variables	**Conventional multilevel models**	
	Model 1: Intercept-only model	Model 2: Generalised least squares fit by a maximum likelihood model
	Estimate	Estimate
Intercept	1.131579	1.4837974
POB2 (migrated from another province)	-	-0.3477613*
Reason for living here3 (family lives nearby)	-	-1.6589423*
Adult malnutrition2 (seldom goes hungry when there is no food at home)	-	0.5884869*
Adult malnutrition4 (often goes hungry when there is no food at home)	-	2.7739817*
Stress6 (feels stress sometimes)	-	-0.6009056*
Stress7 (feels stress sometimes to most of the time)	-	-0.8600252*
Stress8 (feels stress most of the time)	-	1.20131388*
Stress10 (feels stress all of the time	-	-1.2349071*
Income	-	-0.0471636*
Gender2 (female)	-	-0.4576384
Summary output		
AIC	70.56922	64.69989
BIC	75.23068	207.5517
Loglikelihood	33.28461	19.65006
Residual standard error	0.374923	0.06560461

a. Two comorbidities

Model 1: Intercept-only model

As the intercept-only model contained no explanatory variables, it predicted the mean as the best estimate of comorbid health conditions for each respondent. Intercept estimate, AIC, BIC, loglikelihood and residual standard error parameter estimates compared models. Two comorbidities compared to Model 1 and Model 2 to deem a best fit model. The best fit model highlighted the model that best depicted the relationship among explanatory variables and two comorbidity. A good model fit was described as model estimates were p-value < 0.05, a small AIC, small BIC value, small residual standard error as well as a small loglikelihood value.

The intercept estimate (2.6538) revealed the average number of respondents per sub place with hypertension and diabetes. Null model did not account for any explanatory variables. Summary outputs illustrated the AIC (3637.258) as a combination of model fit with a measure of model complexity. The smaller the AIC, the better the model fit. Helwig (2017) described the Bayesian information criterion (BIC) (3636.699) as aimed to approximate the posterior probability of the model. The smaller the BIC, the better the model fit. Loglikelihood (-1816.629) measured how the model fit the data. The smaller the Loglikelihood, the better the model fit. The closer to zero of Residual standard error (2.164981), the better the model fit.

Model 2: Generalised least squares fit by a maximum likelihood model

Intercept estimate (3.618803) revealed the average number of respondents per sub place with hypertension and diabetes. The significant explanatory categorical variables were *reason for living here, child malnutrition, exercise frequency, stress and income*. *Quality of property* was more likely to report two comorbidities for *reasons for living in Gauteng* (-1.326993; p-value (0.0114)), with the reference category as *affordability of property*. *Sometimes goes hungry when there is no food at home* was more likely to report for two comorbidities for *child malnutrition* in Gauteng (-1.266922; p-value (0.0083), with the reference category as *never goes to bed hungry when there is no food at home*. *Exercises few times per month* was more likely to report two comorbidities for *exercise frequency* in Gauteng (-0.617889; p-value (0.0325)), with the reference category as *always exercises*. *Never exercises* was more likely to report two comorbidities for *exercise frequency* in Gauteng (-0.528381; p-value (0.0wa65)), with a reference category as *always exercises*. *Feels stressed most to all of the time* was more likely to report two comorbidities for *stress* in Gauteng (0.768431; p-value (0.0462)), with the reference category as *feels no stress at all*. As *income* bracket increased (-0.032490; p-value (0.0086)), the number of respondents reporting two comorbidities decreased. Therefore, the lower the *income* bracket, the more respondents per sub place reported two comorbidities.

Model 2 accounts for the fact that *quality of property, sometimes goes hungry where is no food at home, exercises few times per month, never exercises, feels stressed most to all of the time* and *low-income bracket* are more likely to report two comorbidities in Gauteng. Respondents who reported two comorbidities were influenced by the quality property when deciding on a place to live. One can presume that the grander the quality of property the less likely respondents would present two comorbidities (hypertension and diabetes). Respondents who reported two comorbidities were influenced by levels of hunger when stating if they faced malnutrition in their household. One can presume that the hungrier respondents felt the more likely respondents would present two comorbidities. Respondents who reported two comorbidities were influenced by exercise duration when questioned about their physical activeness. One can presume that the less exercise respondents did the more likely to present two comorbidities. Respondents who reported two comorbidities were influenced by levels of stress when stating their mental state. One can presume that the more stress respondents felt the more likely to present two comorbidities. Respondents who reported two comorbidities were influenced by level of income when describing their household income. One can presume that the lower the income bracket household the more likely to present two comorbidities. It is important to note that explanatory variables have shown high association with two comorbidities when analysed as a collective. If analysed individually or in conjunction with other explanatory variables, association to two comorbidities may vary.

Summary outputs illustrated the AIC (3681.939), BIC (3636.699), Loglikelihood (-1816.629) and Residual standard error (2.164981). A best fit model deemed Model 2 as a good model fit owing to smaller AIC and BIC values, smaller residual standard error and a smaller loglikelihood value as compared to Model 1.

b. Three comorbidities

Model 1: Intercept-only model

The intercept estimate (1.31579) revealed the average number of respondents per sub place with hypertension, diabetes and influenza/ pneumonia. Null model did not account for any explanatory variables. Summary outputs illustrated the AIC (70.56992) as a combination of model fit, with a measure of model complexity. BIC (75.23688) as aimed to approximate the posterior probability of the model. Loglikelihood (-33.28461) measured how the model fit the data. The closer the Residual standard error (0.374923) is to zero, the better the model fit.

Model 2: Generalised least squares fit by a maximum likelihood model

The intercept estimate (1.4837974) revealed the average number of respondents per sub place with hypertension, diabetes and influenza/ pneumonia. The significant explanatory categorical variables were *place of birth, reason for living here, adult malnutrition, stress, income and gender*. *Migrated from another province* was more likely to report three comorbidities for *place of birth* in Gauteng (-0.3477613; p-value (0.0490)), with the reference category as *born in Gauteng*. *Family lives nearby* was more likely to report three comorbidities for *reason for living here* in Gauteng (-1.6589423; p -value (0.0099)), with the reference category as *affordability of property*. *Seldom goes hungry when there is no food at home* was more likely to report three comorbidities for *adult malnutrition* in Gauteng

(0.5884869; p-value (0.0440)), with the reference category as *never goes to bed hungry when there is no food at home*. *Often goes to bed hungry when there is no food at home* was more likely to report three comorbidities for *adult malnutrition* in Gauteng (2.7739817; p-value (0.0116)), with the reference category as *never goes to bed hungry when there is no food at home*. *Feels stress sometimes* was more likely to report three comorbidities for *stress* in Gauteng (-0.6009056; p-value (0.0144)), with reference category as *feels no stress at all*. *Feels stress sometimes to most of the time* was more likely to report three comorbidities for *stress* in Gauteng (-0.8600252; p-value (0.0432)), with reference category as *feels no stress at all*. *Feels stress most of the time* was more likely to report three comorbidities for *stress* in Gauteng (1.20131388; p-value (0.0393)), with reference category as *feels no stress at all*. *Feels stress all of the time* was more likely to report three comorbidities for *stress* in Gauteng (-1.2349071; p-value (0.0291)), with reference category as *feels no stress at all*. As the *income* bracket increased (-0.0471636; p-value (0.0038)), the number of respondents per sub reporting three comorbidities decreased. Therefore, the lower the *income* bracket, the more respondents per sub place who reported two comorbidities. *Females* were more likely to report three comorbidities as compared to *males* in Gauteng (-0.4576384; p-value (0.0114)), with reference category as *males*.

Model 2 accounts for the fact that *migrated from another province, family living nearby, seldom goes hungry when there is no food at home, often goes hungry when there is no food at home, feels stressed sometimes, feels stress some to most of the time, feels stress most of the time, feels stress all of the time, low-income bracket* and *female* are more likely to report three comorbidities in Gauteng. Respondents who reported three comorbidities were influenced by migration when stating their place of birth. One can presume that migration from another province are more likely to present three comorbidities (hypertension, diabetes, influenza/pneumonia). Respondents who reported three comorbidities were influenced by knowing your neighbours in the area you live in family when deciding on a place to live. One can presume that respondents felt if family lived nearby they were are less likely to present three comorbidities. Respondents who reported three comorbidities were influenced by levels of hunger when stating if they faced malnutrition in their household. One can presume that the hungrier respondents felt the more likely to present three comorbidities. Respondents who reported three comorbidities were influenced by levels of stress when stating their mental state. One can presume that the more stressed respondents felt the more likely to present three comorbidities. Respondents who reported three comorbidities were influenced by levels of income when describing their household income. One can presume that the lower the income bracket household the more likely to present three comorbidities. It is important to note that explanatory variables have shown high association with two comorbidities when analysed as a collective. If analysed individually or in conjunction with other explanatory variables, association to three comorbidities may vary.

Summary outputs illustrated the AIC (74.69989), BIC (207.5517), Loglikelihood (19.65006) and Residual standard error (0.0650461). Even though BIC value was smaller in Model 1 compared to Model 2, the best fit model deemed Model 2 as a good model fit owing to a smaller AIC value, smaller residual standard error and a smaller loglikelihood value.

Discussion

Key findings

i. Prevalence

Objective one estimated the three most prevalent comorbid health conditions (hypertension, diabetes and influenza/pneumonia) in Gauteng as comorbidity classes within two, three and four-or-more health conditions. As previously stated, frequency did not accurately depict prevalence of comorbidity as it referred to the number of cases instead of the proportion of cases, and this report illustrated proportion of comorbid health conditions. Therefore the use of population proportion has been used to describe comorbidity by model estimate and p-value. Comorbid prevalence was illustrated within its class, where two comorbidities described hypertension and diabetes (n=829,8.97%), three comorbidities described hypertension, diabetes and influenza/ pneumonia (n=76, 3.017%), and four-or-more comorbidities described hypertension, heart disease/ stroke, diabetes and asthma (n=16, 0.96%).

In line with the research questions, the descriptive statistics highlighted general Gauteng health from the QoL Survey as *good,* and the top three singularly selected health conditions as hypertension, heart disease/ stroke and diabetes. Although comorbidity is not classified as a good health outcome, the *good* health outcome represents total population health and does not support comorbid health. Comorbidity only characterises a small proportion of Gauteng's health. Descriptive statistics of health conditions, although premature to the study, highlight major health conditions that inform the comorbid health conditions of this study. As comorbidity cases accumulated to a mere 12.947% across all comorbidity classes, it is an indication that Gauteng is not plagued by comorbid health conditions. It can be suggested that comorbidity prevalence does not affect Gauteng on a large scale. Although the majority of Gauteng is not affected by comorbid health conditions, measures must still be taken to ensure that the minority population, that do suffer with comorbidity of conditions, are well taken care of.

ii. Spatial analysis

Objective two identified spatial dependency of comorbid health conditions in Gauteng through methods of frequency and density patterns, spatial autocorrelation, hot spot analysis and OLS regression models. Point frequency patterns illustrated density surfaces of comorbidity points across Gauteng. High point density across all comorbidity classes indicated comorbidities were located in built-up areas. Low density areas were located on the outskirts of Gauteng, where populations tend to be smaller.

Spatial autocorrelation highlighted two comorbidities and four-or-more comorbidities as random patterns. This reinforces the notion that there is spatial dependence across these comorbidity classes and spatial analyses should be extended. As three comorbidities indicated a dispersed pattern, this proves that the points did not associate with one another; and further suggests that comorbid health conditions are challenging to relate to on a spatial level.

Hot spot analysis was conducted to visualise clusters of high and low values of comorbidities. City of Johannesburg, Emfuleni, Midvaal and Ekurhuleni observed the highest number of hot and cold spots for two comorbidities. This correlates to the high point density result which reiterates that comorbidities are located in built-up urban areas. As four-or-more comorbidities contained a smaller dataset, pockets of health are distributed across Gauteng. Only Ekurhuleni municipality was described as a hot spot location, and other municipalities were recorded as *not-significant* locations.

Following Anselin (2005) spatial regression decision process, regression methods were used to identify the appropriate spatial regression model. An OLS model, also known as an unfiltered model, was tested. OLS model was a non-spatial model that tested for significant spatial dependence among residuals. Results stated that only two comorbidities highlighted spatial dependence; three comorbidities highlighted there was no spatial dependence among residuals; and four-or-more comorbidities highlighted spatial dependence.

In line with the research questions posed, spatial dependency of two comorbidities with a Moran's I (0.026151) represented a high clustering value. This means that hypertension/ diabetes is not random or dissimilar and, instead, it follows a spatial pattern where the highest distribution of two comorbidities are most likely near other high two comorbidity locations. This phenomenon is further tested by the regression model which explains how this phenomena is associated to the comorbid condition.

Spatial dependency of three comorbidities with a Moran's I (-0.328319) represented a dissimilar pattern. The z-score (-2.36772) and positive p-value (0.09977) indicates that hypertension, diabetes and influenza/pneumonia do not follow a spatial pattern and are therefore removed from further analyses.

Spatial dependency of four-or-more comorbidities with a Moran's I (-0.93099), z-score (-0.12424) and positive p-value (0.90112) represented high clustering values. This means that hypertension, heart disease/stroke and asthma are not random or dissimilar. Instead, it follows a spatial pattern where the highest distribution of four-or-more comorbidities are most likely near other high four-or-more comorbidity locations. The phenomenon was further tested in the regression models to reiterate the factors associated to the comorbidity.

a. Hot spot analysis

GIS has adapted to rapid population growth and has evolved to help visualise problems in high-prevalence areas and populations that are at high-risk. Hot spot analysis has unlocked the potential to determine high-prevalence and high-risk populations. Although density mapping can illustrate where clusters occur, it does not indicate statistical significance. Hot spot analysis uses vectors to identify locations of significant high and low values of clustering in data. Beyond assessing the density of points in a given area, hotspot techniques also measure the extent of point event interaction to understand spatial patterns (Mudiyanselage and Senanayake 2016). In this report, hot spots were been identified in urban municipalities like City of Johannesburg, Emfuleni and Ekurhuleni. Urban health disparities are a result of the inequalities found within these urban area.

The rapid urbanisation of the City of Johannesburg has led to socioeconomic disparities in housing, employment and income. Sobngwi *et al* (2002) states that urbanisation has become closely associated with changes in dietary habits, stress, physical activity and employment. This may have an impact on type of comorbidities that are present in urban areas. Comorbidities included hypertension, diabetes, heart disease/ stroke, influenza/ pneumonia and asthma; all of which were closely related to effects on stress, income, employment, frequency of exercise, and malnutrition in adults and children, as per the QoL Survey variables.

Criminology has been using hot spot mapping since the early 1990s in order to explore unprecedented increase in gun homicides among youth (Wortley and Mazerolle 2008). Poverty rates and crime rates are frequently used characteristics in determining visual hot spots of crime with geographic precision. Geocoding, by transforming an address into precise geographic coordinates, is the method used for geographic precision., Using mathematical matching algorithms, geo-spatial coordinates are used to visualise points in space. Point pattern maps, point aggregation and geo-visualisation are just a few GIS capabilities used in the real world.

Hot spot analysis has evolved in traffic and accident analysis. Erdogan *et al* (2008) uses hot spot analysis tools to analyse accidents in order to identifying location with a high rate of accidents as well as safety deficient areas on highways in Turkey. Tabular data was georeferenced and plotted onto highways. GIS was used as a management system for accident analysis, with the purpose of creating of hot spots, so that traffic officials could implement precautionary measures for traffic safety.

The potential for hot spot analysis continues to evolve in health care. It is used to determine: locations in need of nutrition programmes; the effects of locations with air pollutants on asthma; and how South African epidemiologists are using HIV hotspot mapping to develop location-specific, population-sensitive and evidence-informed HIV prevention responses (Stopka *et al* 2014, Tuluri *et al* 2007, Rousseau 2016).

iii. Multilevel analysis

Objective three determined the multilevel model effects of factors associated with comorbidities through comparison of conventional multilevel models in determining the best fit model for each comorbidity class. Results of two comorbidities confirmed that model 1 was the best fit model. According to regression results, self-rated comorbidity prevalence was significantly associated with: *quality of property* (-1.326993; p-value (0.0114)) , *child malnutrition* (-1.266922; p-value (0.0083)), *exercise few times per month* (-0.617889; p-value (0.0325)), *never exercises* (-0.528381; p-value (0.0165)), *feels stress most to all of the time* (0.768431; p-value (0.0462)) and *low-income bracket* (-0.032490; p-value(0.0086)); and were more likely to report two comorbidities in Gauteng.

Three comorbidities reported that model 1 illustrated a high explanatory power. According to regression results, self-rated comorbidity prevalence was significantly higher for respondents that had: *migrated from another province* (-0.3477613; p-value (0.0490)), *family living nearby* (-1.6589423; p -value (0.0099)), *adult malnutrition* (0.5884869; p-value (0.0440)), *Often goes to bed hungry when there is no food at home* (2.7739817; p-value (0.0116)), *feels stress sometimes*(-0.6009056; p-value (0.0144)), *Feels stress sometimes to most of the time* (-0.8600252; p-value (0.0432)), *Feels stress most of the time* (1.20131388; p-value (0.0393)), *Feels stress all of the time* (-1.2349071; p-value (0.0291)), lower income bracket (-0.0471636; p-value (0.0038)) and female (-0.4576384; p-value (0.0114)); were more likely to report three comorbidities in Gauteng.

b. Influences on Gauteng socioeconomic conditions

In line with the research questions posed, model factors are further examined. Objective three determines the factors associated with the most prevalent comorbidities within the multilevel model. These factors are the socioeconomic characteristics of respondents within their environment. Consequently, it becomes important to understand the how the society has an influence on. Soul City Institute (2016) highlights the link between health and location through the example that substance abuse has direct and indirect bearing on the health status, which, in turn, increases risks of unemployment and crime, and eventually lead to premature deaths. Health and location are often interlinked to accessibility, utilisation and affordability. As per results, comorbidity prevalence is found to be linked to socioeconomic characteristics such as race, quality of property, malnutrition, low exercise frequency, high stress, low income bracket, gender, age, migration, social issues and education.

Population group

Table 4.3 examines various factors associated with comorbid health in Gauteng. *Black African* population group makes up hundred percent of all population groups, therefore defines comorbidity in Gauteng by race. Considering this result, policy should target population groups for comorbidity healthcare. A focus on race reveals that there may be underlying social challenges of ill-health between socioeconomic groups. It becomes imperative to recognise the socioeconomic characteristics of comorbidity in order to target comorbid epidemics. GCRO (2017) determines that *Black African* is the only population group where the average quality of life is below provincial average.

Diet may have had an impact on *Black African* health, resulting from a shift towards a more westernised diet, which has led to the breakdown of macronutrients (Bourne *et al* 2002). Hypertension and diabetes had also been attributed to *Black African* health status resulting from obesity among women and poor access to physical activity (Schwandt *et al* 2016). As two comorbidities has reported population group as a significant variable,

Quality of property

It is important to recognise the differences between rural and urban communities, and how this effects well-being. Well-being is associated with longer life expectancy; better health outcomes and the effects of residence may improve or diminish their well-being (Bates 2018). The physical environment may explain the higher number of factors associated with well-being and health. Bates (2018) suggests there are links between well-being and feeling positive about your community, that lead to health benefits. Florida (2017) reiterates that "quality of place" is linked to high levels of better health and happiness, as less advantaged communities may experience poorer quality of place because of pollution, which in turn leads to higher levels of stress and anxiety. Urban planners have a responsibility to ensure that environments promote good health behaviours and protect against unhealthy behaviours, as this could mean saving on health care costs in the future (Brazier 2016).

Malnutrition

NHS (2019) describes food poverty as the inability to acquire or consume adequate foods in a socially acceptable way. This has dangerous effects on physical and psychological well-being in the long run. NHS (2019) highlights food poverty and health inequalities as associated with socioeconomic status, ethnicity and geographic location. Food poverty often involves sacrifice and diet adjustments, in the form of skipping meals, to save on costs or to feed others first. Sustain (2019) states that diet-related health diseases include diabetes, cancer, cardiovascular diseases, obesity and malnutrition. FRAC (2017) states that children who live in households that are" food insecure" are more likely to be sick more often and recover from illnesses slower. There is a "higher probability of chronic diseases such as hypertension, coronary heart disease, hepatitis, stroke, cancer, asthma, arthritis, chronic obstructive pulmonary disease and kidney disease" associated with food insecurity (FRAC 2017:1).

Low exercise frequency

Exercise can have lasting health benefits. A study by Xi *et a*l (2019) found that as little as ten minutes per week of leisurely exercise is associated with decreased risk of death. Physical activity is not limited to age but rather stamina, strength and suppleness. Mayo Clinic (2018) states that exercise has health benefits for chronic diseases such as heart disease, diabetes, asthma and joint or back pain. Warburton *et al* (2006) highlights physical inactivity in the development of chronic diseases and premature death as a modifiable risk.

High stress

FRAC (2017) states that growing up in poverty is associated with toxic stress, which in turn has negative effects on child development and health. By undergoing stress, the stress hormone levels become increasingly high and may lead inhibit normal brain and physical development, as well as the metabolic processes in children. Ganster and Schaubroeck (1991) state there is a relationship between self-reported stressors in the work environment (lack of social support, lack of participation, hectic job demands and low decision latitude) and diastolic blood pressure. Stressful experiences become a way for gender, racial-ethnic, marital status and social class inequalities to manifest into physical and mental health problems. Thoits (2010) highlights that "coping" and support interventions should be more widely employed, and there should be a presence of focus programmes and policies at macro levels for intervention. Thoits (2010) suggests that there is a need for more comprehensive measures of stressors to explain high rates of disease, injury, disability and mortality.

Low income bracket

The burden of ill-health in South Africa has shown that "relative to the wealthy, the poor suffer from more disease and violence" (Ataguba *et al* 2011: 10). Even though income inequality does not have a direct effect on health, it does have a follow-through impact. *Income and education* can be pooled in the same group as non-direct impacts on health, although both affect the way in which health information is received and understood.

Social grants support a large of part of the South African population and are responsible for pension pay-outs, disability, child support and other household services (Huffington Post 2017). Social grant households survive on whatever they receive as support from government. However, social grants do not cover a lifelong poverty. Income poverty unemployment cannot be alleviated by social aid, but rather by integration of policies on education, health care and housing.

As per the Socio-Economic Review and Outlook (2018), Gauteng employment has been declining since 2012 and unemployment has rapidly increased reaching its highest rate in 2017. The rapid population growth has spurred competition among the work force which has left many struggling for work. Household income has increased, and the increase in lower-income household rates can be attributed to post-1994 job creations (SocioEconomic Review and Outlook 2018). The increase in income per household increases has a direct effect on medical treatment. However, medical treatments are expensive, and many households may not have insurance or sufficient social grant funds to reverse ill-health.

Gender

Gender is higher in females in both the two and three comorbidity classes; while comorbidity class of four-or-more is higher in males. Although the comorbid health conditions are interlinked in terms of health conditions following in-suit of one another, the impact of health on gender should be further analysed by examining the effects of disease on each gender type. Morrell *et al* (2012) highlights the idea of South Africa being dominated by *hegemonic masculinity* to make sense of a gendered landscape.

Female populations have always been more at risk of carrying the disproportionate burden of illnesses, such as HIV infection, TB and depression. Fischer *et al* (2007) highlights the correlation between alcohol abuse, HIV infection and gender, whereby alcohol abuse is a high-risk factor for HIV infection among men and women. Among female university students, Janse van Rensburg and Surujlal (2013) highlight that females report higher incidence of stress across Gauteng campuses. Behaviours and lifestyle patterns often reflect health outcomes.

Age

As the average age of comorbidity is categorised between the ages of 55 to of 65 years and older, it is important to examine why comorbidities occur among the elderly. Gauteng's elderly population over 65 years of age consists of more women than men, and the elderly, white population make up the large proportion (Gauteng Provincial Government Review 2014).

Sudore *et al* (2006) suggests that as the *elderly population* grows, the burden of chronic diseases increases. Descriptive statistics highlighted that the average age was over 65 years, thus proving that positive health outcomes will significantly decline among the elderly population. WHO (2018) explains that, at a biological level, the gradual decrease in physical and mental capacity puts individuals at risk for disease and ultimately death.

Older age is also characterised by the rise of more chronic health problems, including hearing loss, eyesight deprivation, osteoarthritis, obstructive pulmonary diseases and dementia (WHO 2018). According to StatsSA (2011), the number of elderly persons has increased. Among the Black African and coloured population groups, the old-age dependency ratio has increased by three percentage points, indicating a higher burden of family support and care of the elderly (Hock and Weil 2012). This may have an impact on well-being of households.

Migration

Increasing rates of urbanisation and lead to urban population growth and is now associated with migrant patterns across major cities. These high levels of *migration* have led to social tensions and poverty in areas of Gauteng, where crime rates have risen and competition among residents' results in unemployment and poverty (Vearey *et* al 2014).

Migration has an impact on health as it may induce distress on wellbeing. The impacts on Gauteng health have been attributed to presence of high blood pressure and diabetes, reaffirming results from comorbidity prevalence (Vearey *et al* 2014). Hypertension prevalence is attributed to distress in any environment and could contribute to respondents facing issues of unemployment, who often become displaced from homes. As per descriptive statistics, unemployment is relatively high among respondents with comorbid health conditions which include hypertension.

Urbanisation vulnerabilities have an impact on place of residence, gender, level of formal education and access to services (Vearey *et al* 2014). The implications of migration may result in an increase in the 'urban poor' population, who experience interlinked deprivations within that society. Migration is acknowledged as a determinant to health as it poses challenges to health outcomes (Anarfi 2005, WHO 2008).

Social issues

Crimes against women, such as reports of hijacking of motor vehicles, street robbery, sexual offense and assault, poses a large threat to women's health (StatsSA 2018). Fear of crime affecting activities, such as walking to fetch food and water or walking to and from home, is more significant for women in rural areas. In urban areas, fear of crime impacts household homicide rates. Similarly, Seedat *et al* (2009) examine how drug abuse is as dangerous to health, where the health culture and associated social factors are based on *violence, alcohol and drug abuse, and a weak culture of enforcement to uphold safety*. Unemployment and social inequalities leave women vulnerable and is a contributing factor for an AIDS pandemic (Hunter 2007).

The social dynamic of Gauteng sees the occurrence of violence against women and children, trafficking, poverty, unemployment and income inequality (Kistner 2003). Social issues in health are posed not posed as clinical but are considered structural issues at a government level. Gauteng health is greatly impacted by the rising population demands on health. The Gauteng Department of Health Annual Report (2018) identifies that the burden of communicable diseases and interpersonal and motor vehicle-related trauma remains high and is putting pressure on the public health system. Health literacy is a major obstacle to delivering positive health outcomes among disadvantaged groups, because of poor SES.

Education

Deaton (2003) highlights that socioeconomic status has a fundamental impact on health, and there are correlations between *education* and health. He explains that *education* provides tools for understanding how health information is perceived. Health literacy is vital, as it provides the best opportunity to tackle long and short-term health behaviours. Life Orientation is a school subject that is incorporated into school curriculums. The education of health should be integrated into other domains, such as the workplace.

As per the comorbidity results, education levels from grade 10 to 12 are adequate in understanding health, but not sufficient to enable understanding of how health should be managed and treated. Mahlakoana (2018) states that the low levels of education have an impact on mental stunting of growth among young children. Malnutrition could be a leading cause of inability to understand and perceive information and this has a direct effect on how health is understood. This study proves that the lower the education level, the less health information is understood and managed.

As both the Gauteng health and education departments are integral components of Gauteng's welfare, health promotion should be a priority in the agenda. The Department of Basic Education (2019) programmes on health education aims to create health awareness through progress such as the *Integrated School Health programmes*.

Multilevel model approach

Multilevel models, because of their ability to analyse two or more explanatory variables in tandem, are useful in analysing the effects. Multilevel regression analysis is one of the most widely used methods in determining neighbourhood effects on individual outcomes. The emergence of multilevel models has led to the ability to understand geographic clustering in health at the specified area level. *Place,* as a source of location, exposes welfare policy, social environment quality and neighbourhood resources (Arcanum *et al* 2013). Multilevel models, using small-area unit data, have also been used in combatting UGCoP. Kwan (2012) highlights the use of multilevel models to capture influences across different levels. Multilevel models provide the ability to understand complex hierarchal or nested data types.

Conventional multilevel models can be criticised for demonstrating spatial dependency bias, because of the poor explanatory power and little spatial dependency among residuals. Xu (2014) states that standard multilevel models ignore the correlations between neighbourhood and assume independence in one neighbourhood from those in another. This leads to overstatement of the statistical significance of neighbourhood effects. Health multilevel modelling gives insights to sociodemographic and built-environment variables associated with health outcomes.

Increasingly, multilevel models are used in public health policy, as it enables analysis of the health of people in their context (Leyland and Groenewegen 2003). The use of multilevel analysis is multidisciplinary and spans across fields of health service research, epidemiology and public health, with the assistance of multilevel software and statistical packages. Leyland and Groenewegen (2003) state that before multilevel analysis, OLS regression was the only tool of use. However, posed problems, known as ecological and atomistic fallacy resulted in problems with estimates of higher levels of characteristics. Multilevel analysis makes it possible to test individual characteristics on health between communities, and their contextual characteristic effects, and health.

There are a variety of ways in which multilevel modelling has been conducted on diverse topics. Li *et al* (2005) uses multilevel modelling to explore the relation between built environment factors and walking activity at a neighbourhood level and a resident level in Portland, Oregon. A two-level data structure allows neighbourhoods to correspond to the primary sampling unit, and residents within each primary sampling unit to the secondary unit. Therefore, simultaneous analysis of neighbourhood and residents is possible. Mobley *et al* (2008) examines how variables at different geographic levels impact women's mammography decision-making, using contextual influences on a local and a county level.

The interest in multilevel modelling has grown out of the potential that macro determinates of health are across disciplines, as well as the notion that variables are relevant in understanding the distribution of health outcomes. With the growing interest in health context, there has been accelerated development in statistical methods and techniques (Diez-Roux 2000). Multilevel analysis illustrates that all disease determinants are best measured and conceptualised at the individual level. Group-level variables are only used as proxies when individual-level data is unavailable. Diez-Roux (200) states that a challenge in public health research is its ability to develop practical explanations and move beyond the statistics of the models. Multilevel models do run the risk of being reduced to examining variations that have no real meaning or are difficult to interpret. There is a potential application in health economics which includes the exploration of medical practice , examination of inequalities in resource distribution, as well as an indicator of variable performance.

Strengths

The strengths of this research report lie in its methodology. The identification of comorbidities, although a demanding process, was done manually as this enhances the accuracy of the results. The creation of comorbidity classes allowed for a better understanding of how health conditions are correlated with one another. The way in which the comorbidity was classed enabled associations amongst the health conditions to be identified. The use of numerous socio-demographic variables gave more information about association with comorbidity.

Spatial analysis produced hot spot areas to examine comorbidity prevalence. Public health strategies should be used to intervene in these high-risk locations. Following Anselin's spatial regression decision process, regression models could be accounted for. OLS model, spatial lag and spatial error regression models indicated presence or absence of spatial dependency, which is important in understanding if the data could be explained with or without spatial dynamics.

The multilevel model approach portrayed simultaneous analysis of the individual and community level of comorbidity prevalence. By introducing the a more robust regression model, through the spatial filtering approach, variables expressed the fundamental spatial structures of the model.

Limitations

This research report has limitations that should be thoroughly considered for future research.

Firstly, given that the majority of respondents in this research report were above the age of 65+ years, the employment variable may be problematic as many these respondents are likely to retire after 60 years of age. This could affect the 'age' factor of regression results. One should consider restricting the analysis to this age group.

Secondly, geographic locations were produced within sub-place unit and did not correspond with exact locations within the aggregated unit. Maclaurin et al (2015) states that the implication of aggregating data challenges the migration-environment linkages in spatial analysis. If household level data is displayed as a sub-place level outcome, the structure of aggregation is compromised, and this produces the MAUP, as the aggregation effects may impact spatial autocorrelation analysis and disrupt correlations between defined neighbours.

Thirdly, data included too many categorical variables that could not produce a thorough spatial regression analysis. For future research, data set should include more continuous variables.

Fourth, in terms of comorbidity counts, the data sets appeared to be too small for thorough regression analysis and smaller data sets had to be excluded from analyses. A larger sample should be drawn from population in order to produce thorough results.

Fifth, the Uncertain Geographic Context Problem (UGCoP) defined by Kwan (2012) highlights that, as a result of human mobility, complex geographical and temporal contexts were compromised, especially considering the timing and duration of when respondents experienced comorbidities were unclear. To obtain more accurate results, actual contexts that impact health outcomes should include geography and time of outcomes. Data restrictions did not allow for the report to reflect outcomes below sub-place level, therefore "density calculations and neighbourhood level could not further interrogate the hot spot analysis results" (Weimann et al 2016:155). As a result, hot spots should be interpreted with caution.

Finally, there are limitations around the usage of cross-sectional designed studies, as causality cannot be inferred. Chen et al (2015) highlights that cross-sectional studies do not illustrate temporal relationships between cause and effect. Thelle and Laake (2015) reiterates this statement, and supports Chen et al (2015), by suggesting that cross-sectional studies are poorly suited to examine disease and illness because there is no dimension of time, and therefore they are unable provide new information during that specific time period. As the QoL Survey 2015 describes one year of data, if a respondent has experienced a change in health, then this report will not reflect that change in one year. As cross-sectional studies examine prevalence rather than incidence, data will always reflect determinates of survival. However, Deen et al (2014) states that cross-sectional designs provide other important health information, such as the effectiveness of prevention and control strategies. This report highlights socioeconomic links to comorbidity in Gauteng that may inform health policies to better understand and contain the spread of comorbid health conditions. Barrat et al (2018) states that, as the methodology is used to assess the burden of disease on a population, as well as the health needs of the given population, cross-sectional design is useful in planning the allocation of health resources. This report seeks to assess the links to comorbidity that may better equip healthcare facilities to manage and care for patients experiencing poor health as a result of comorbid health conditions. Sedgwick (2014) suggests that results from a cross-sectional study may inform more complex investigation within a cohort study.

Conclusion

This research report explored the effects of factors on self-reported status of respondents from the QoL Survey 2015.

Main findings

The first objective of this report estimated comorbid health conditions in groups of two, three and four-or-more. The second objective identified spatial dependence of comorbidity that reflected two comorbidities and four-or-more comorbidities were spatially dependent among the residual errors and that the highest distribution of two and four-or-more comorbidities were most likely near other high comorbidity locations. According to regression results, the third objective determined that self-rated comorbidity prevalence was more likely to report two comorbidities (n=829,8.97%) for *quality of property, child malnutrition, low exercise frequency, high stress* and *low-income bracket* in Gauteng. Self-rated comorbidity prevalence was more likely to report three comorbidities (n=76, 3.017%) for *migrated from another province, family living nearby, adult malnutrition, medium to high stress, low-income bracket* and *female* in Gauteng. Results of comorbidity for four-or-more were defined by its limitation, that the data set was too small for regression analysis.

Implications of findings

In line with the research questions posed, the effects of comorbidity prevalence had wider implications on Gauteng's health. Health effects were defined as the changes in health from resulting exposure (Sherwin 1983). Health effects take into consideration the fields of occupational health and safety, nutrition, pollution studies and hygiene. Important implications arise from interactions between comorbidity and population health. Comorbidity prevalence in Gauteng's health has impacts on complex clinical management, health costs and ability to worsen health outcomes. As Gauteng is plagued with a shortage of medical staff, there is a higher risk of the burden of comorbidity to worsen healthcare outcomes (Child 2019). Comorbidity healthcare needs to be increasingly managed with patients who have two or more chronic diseases, as costs rapidly increase as the number of health conditions increases. The impact of comorbidity has an impact on mortality, quality of life and quality of health care.

By defining comorbidity as the *simultaneous presence of two or more chronic diseases or conditions in a patient,* informed thinking amongst researchers and practitioners may update more appropriate health measures on specific comorbid outcomes. Brand *et al* (2013) highlights that primary healthcare is the foundation of the South Africa's healthcare system and, if ill-managed, it could impact all levels of health management. As Gauteng comorbidity was defined by major chronic conditions such as *hypertension, diabetes, heart disease* and *asthma*, it becomes imperative that these high-risk respondents be recruited in more intensive and aggressive health programmes with access to more hospital visits and better drug exposure.

Various approaches have been taken to convey comorbidity as a single measure on a scale. The Charlson Index, Cumulative Illness Rating Scale (CIRS), Kaplan Index and others have focused on classifying patients into groups based on disease conditions, age and gender (Valderas *et al* 2007). However, because of patient complexity, comorbidity is not only related to health-related characteristics but also to cultural, environmental and socioeconomic characteristics, and capturing this measure of complexity remains a challenge. Comorbidity becomes a challenge for Gauteng to address its complexity, as well as its challenges for accessibility and cost-effectiveness. Brand *et al* (2013) poses the idea to study primary healthcare systems of the Japanese, German and Brazilians.

Future research

Further work is needed to develop a more comprehensive mix of GIS and socio-demographic measures relating to comorbidity. Given the complexity of comorbid health conditions, although the QoL Survey covered many general household topics, a more inclusive health-based profile should be established. Environmental factors (smoking, diet, and live near a landfill site) should be included in future research. These factors are excluded from this study, as they are not evident in the QoL Survey. Future research should use data from a more credible health source so to legitimise the study. Comorbidity is a complex construct and should be fully analysed in order to retain appropriate outcomes for policy planning.

In summary, this research report contributes to multilevel research by evaluating the influence of socio-demographic, particularly the spatial dynamics, on comorbidity in Gauteng. Findings from this research suggest more work needs to be developed around multilevel modelling of comorbidities, as there is a gap in knowledge around the factors that best explain comorbidity prevalence. It would be useful to apply the multilevel model against other datasets to test for different outcomes. Finally, it is hoped that the present findings may advise policy planners in mitigating health inequalities in the broader context of South African epidemiology.

References

Alaba, O. and Chola, L., 2013: The social determinants of multimorbidity in South Africa in *International Journal for Equity in Health*, 12(63), 1-10.

American Heart Association., 2015: Why are people with diabetes at increased risk for CVD, Cardiovascular Disease & Diabetes. Available from: www.heart.org (Accessed 12 July 2018).

American Lung Association., 2018: What is the connection between influenza and pneumonia*, Lung Health & Diseases*. Available from: www.lung.prg/lung-health-and-diseases (Accessed 01 December 2018).

Anarfi, J., 2005: Reversing the spread of HIV/AIDS: What role has migration? *International Migration and the Millenium Development Goals. Selected papers of the UNFPA Expert Group Meeting, Marrakech, Morocco, 11-12 May 2005,* UNFPA. New York.

Anselin, L., 1988: Studies in Operational Regional Science *in Spatial Econometrics: methods and models*, 4(1), XVI-284.

Anselin, L., 2002: Under the hood. Issues in the specification and interpretation of spatial regression models in *Agricultural Economics*, 27(3), 247-267.

Anselin, L., 2005: Exploring spatial data with GeoDa: A workbook, Department of Geography, University of Illinois, March 6.

Arcaya, M, Brewster, M, Zigler, C.M. and Subramanian, S.V., 2013: Area variations in health: A spatially multilevel modelling approach in *Health Place*, 18(4), 824-831.

Ataguba, J.E, Akazil, J. and McIntyre, D., 2011: Socioeconomic-related health inequalities in South Africa: evidence from General Household Surveys in *International Journal for Equity in Health*, 10, 1-10.

Azevedo, M. and Alla, S., 2008: Diabetes in Sub-Saharan Africa: Kenya, Mali, Mozambique, Nigeria, South Africa and Zambia in *International Journal of Diabetes in Developing Countries*, 28 (4), 101-108.

Babyak, M., 2004: What you see may not be what you get: A Brief, nontechnical introduction to overfitting in regression-type models in Psychosomatic Medicine, 66, 411-421.

Barrat, H, Kirwan, M. and Shantikumar, S., 2018: Design, applications, strengths and weaknesses of cross-sectional, analytical studies (including cohort, case-control and nested case-control studies), and intervention studies (including randomised controlled trials), HealthKnowledge. Available from: (https://www.healthknowledge.org.uk) (Accessed 19 June 2019).

Bates, D., 2018: Where you live affects your happiness and health, but how exactly? Available from: www.houstonpublicmedia.org (Accessed 27 May 2019)

Best, N.G, Ickstadt, K. and Wolpert, R.L., 2000: Spatial Poisson Regression for Health and Exposure data measured at disparate resolutions in *Journal of the American Statistical Association*, 95 (452), 1076 – 1088.

Bertazzon, S., 2014: GIS and Public Health, *ISPRS International Journal of Geo-information*, 3, 868-870.

Bhatnagar, A., 2017: Environmental determinants of cardiovascular disease in *Circulation Research*, 121(2), 162-180.

Bourne, L.T, Lamber, E.V. and Steyn, K., 2002: Where does the black population of South African stand on the nutrition transition in *Public Health Nutrition*, 5 (1A), 157-162.

Boulos, M.N.K., 2004: Towards evidence-based, GIS-driven national spatial health information infrastructure and surveillance services in the United Kingdom in *International Journal of Health Geographics*, 3 (1), 1-50.

Brand, M, Woodiwiss, A.J, Michel, F, Booysen, H.L, Majane, O.H.I, Maseko, M.J, Veller, M.G. and Norton, G.R., 2013: Chronic diseases are not being managed effectively in either high-risk or low-risk populations in South Africa in *South African Medical Journal,* 103(12), 938-941.

Brazier, Y., 2017: Where you live affects your health and risk of obesity, Medical News Today.Available from: www.medicalnewstoday.com (Accessed 27 May 2019).

Brundson, C, Fotheringham, A.S. and Charlton, M.E., 1996: Geographically Weighted Regression: A method for exploring spatial non-stationarity in *Geographical Analysis*, 28 (4), 281-298.

Caughey, G.E, Vitry, A.I, Gilbert, A.L. and Roughead, E.E., 2008: Prevalence of comorbidity of chronic conditions in Australia in *BMC Public Health*, 8, 1-13.

Chen, J, Jiang, F, Liu, S.S, Li, s, Kulu, K.S, Eckert, G, Ind, I. and Xia, Z., 2015: Limitation of a cross-sectional study in *American Journal of Orthodontics and Dentofacial Orthopaedics*, 148(2), 205.

Child, K., 2019: Gauteng health hamstrung by critical shortage of doctors and nurses, Times Lives. Available from: www.timeslives.co.za (Accessed 28 May 2019).

Chung, K, Yang, D-H. and Bell, R., 2004: Health and GIS: Towards spatial statistical analyses in *Journal of Medical Systems,* 28(4), 349-360.

CI, A.C.I. and Rule, R.O., 1987. Confidence intervals. *Lancet*, *1*, pp.494-497.

Coovadia, H, Jewkes, R, Barron, P, Sanders, D. and McIntrye, D., 2009: The health and health system of South Africa: historical roots of current public health challenges in *Health in South Africa 1*, 374, 817-834.

Cordero, A, Bertomeu-Martínez, V, Mazón, P, Fácila, L, Bertomeu-González, V, Cosín, J, Galve, E, Núñez, J, Lekuona, I. and González-Juanatey, J.R., 2011: Factors Associated With Uncontrolled Hypertension in Patients With and Without Cardiovascular Disease in Revista Española de Cardiologiá, 64 (7), 587-593.

Cromley, E.K. and McLaferty, S.L., 2012: Mapping Health Information in *GIS and Public Health*, 2nd Edition, Chapter 4, The Guilford Press, New York.

Deaton, A., 2003: Health, Income and Inequality, The National Bureau of Economic Research. Available from: www.ncber.org (Accessed 12 March 2019)

Deen, J, Von Siedlein, L. and Clemens, J.D., 2014: Issues and challenges in public health research in developing countries in *Manson's Tropical Infectious Diseases (Twenty-third Edition)*, 40-48.

Department of Health., 2019: Health Promotion, Department of Basic Education. Available from: www.education.gov.za (Accessed 26 March 2019).

De Groot, V, Beckerman, H, Lankhorst G.J. and Bouter L.M., 2003: How to measure comorbidity. a critical review of available methods *in* Journal of Clinical Epidemiology,56, 221-229.

Diez-Roux, A.V., 2000: Multilevel analysis in public health research in *Annual Review of Public Health*, 21, 171 – 192.

Duncan, C, Jones, K. and Moon, G., 1998: Context, composition and heterogeneity: using multilevel models in health research in *Social Science Medicine*, 46(1), 97 - 117.

EQ-5D., 2017: Valuation, EQ-5D Instruments. Available from: www.euroqol.org (Accessed 28 January 2017).

Erdogan, S, Yilmaz, I, Baybura, T. and Gullu, M., 2008: Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar in *Accident Analysis & Prevention*, 40(1), 174 - 181.

ESRI., 2017: *GIS for health care*, ArcUser Online. Available from: http://www.esri.com (Accessed 01 May 2017).

Fischer, J.C, Bang, H. and Kapiga, S., 2007: The Association Between HIV Infection and Alcohol use: A Systematic Review and Meta-Analysis of African Studies in *Sexually Transmitted Diseases*, 34(11), 856-863.

Florida, R., 2017: Where you live impacts your health, Quora. Available from: www.forbes.com (Accessed 27 May 2019).

Florax, R.J.G.M. and Nijkamp, P., 2003: Misspecification in linear spatial regression models in Tinbergen Institute Discussion Paper, 1-28.

FRAC., 2019: The impact of poverty, food insecurity and poor nutrition on health and well-being, Hunger and Health, Food Research and Action Centre.

Frericher, R.R., 2003: John Snow and the Broad Street Pump, UCLA: Department of Epidemiology, 31(3), 23-31.

Ganster, D.C. and Schaurbroeck, J., 1991: work stress and employee health in *Journal of Management*, 17, 235-271.

Gatrell, A. and Löytönen, M., 2001: Cartographic perspective in *GIS and Health*, 39, 53-55.

Gauteng Department of Health., 2015: Gauteng Department of Health Annual report 2014/2015. Available from: www.gauteng.gov.za/goverment/departments/health (Accessed 24 November 2018).

Gauteng Provincial Government Review., 2014: Gauteng Provincial Government Review of 20 years of Democracy, 20 Year Review. Available from: www.gauteng.gov.za (Accessed 26 March 2019).

Gauteng Department of Health Annual Report., 2018: Department of General Information, Chapter 1: General Information. Available from: www.gauteng.gov.za/government/health/Annual%20Report/GDH%20Annual%20Report%202017-2018.pdf (Accessed 20 November 2018).

Gauteng Strategic Plan., 2016: The HIV and Tuberculosis epidemic in Gauteng. Available from: www.gauteng.gov.za/government/departments/office-of-the-premier/ (Accessed 19 November 2018).

GCRO., 2017: Quality of life remains lowest for African respondents and is not catching up, 5th Quality of Life Survey 2017 -2018. Available from: www.grco.ac.za (Accessed 26th March 2019).

Goodchild, M.F, Haining, R.P. and Wise, S.,1992: Intergrating GIS and spatial analysis in problems and possibilities *in International Journal of Geographic Information Systems*, 6, 407-423.

Greenland, S, Senn, S.J, Rothman, K.J, Carlin, J.B, Poole, C, Goodman, S.N. and Altman, D.G., 2016: Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations in *European Journal of Epidemiology*, 31, 337-350.

Hay, S.I, Omumbo, J.A, Craig, M.H. and Snow, R.W., 2011: Earth Observation, Geographic Information Systems and Plasmodium falciparum Malaria in Sub-Saharan Africa in *Advances in Parasitology*, 47, 173-215.

Hayat, M.J, Powell, A, Johnson, T. and Cadwell, B.L., 2017: Statistical methods used in the public health literature and implications for training of public health professionals in *PLoS One*, 12(6), 1-10.

Health24., 2018: When the flu turns deadly, *Colds and flu.* Available from: www.health24.com (Accessed 01 December 2018).

Health Systems Trust., 2017: South African Health Review. Available from: www.hst.org. (Accessed 15 March 2019).

Helwig, N.E., 2017: Model selection and diagnostics. Available form: http://users.stat.umn.edu/~helwig/notes/msd-Notes.pdf (Accessed 27 May 2019).

Hock, H. and Weil, D.N., 2012: On the dynamics of the age structure, dependency, and consumption in *Journal of Population Economics*, 25(3), 1-26.

Hox, J., 2002: Introduction to multilevel analysis in *Multilevel Analysis: techniques and applications*, Chapter 1, Lawerance Elrbaum Associates and Publishers, Mahwah, New Jersey.

Huffington Post., 2017: Why social grants matter in South Africa: they support 33% of the nation. Available from: www.huffingtonpost.co.uh (Accessed 11 March 2019).

iLead., 2018: Rapid urbanization poses urban development challenges in Gauteng. Available from: www.practicegroup.co.za/rapid-urbanization-poses-urban-development-challenges-in-gauteng/ (Accessed 19 November 2018).

Iribarren, C, Tolstykh, I.V, Miller, M.K, Sobel, E. and Eisner, M.D., 2012: Adult Asthma and Risk of Coronary Heart Disease, Cerebrovascular Disease, and Heart Failure: A Prospective Study of 2 Matched Cohorts in *American Journal of Epidemiology, 176 (11), 1014-1024.*

Islam, M.M, Valderas, J.M, Yen, L, Dawda, P, Jowsey, T. and Mcrae, I.S., 2014: Multimorbidity and Comorbidity of Chronic Diseases among the Senior Australians: Prevalence and Patterns in *PloS ONE*, 9(1), 1-11.

Janse van Rensburg, C. and Surujlal, J., 2013: Gender differences related to the health and lifestyle patterns of university students in *Health SA Gesondheid*, 18 (1), 1-8.

Jia, H, Zack, M.M. and Thompson, W.W., 2013: The effects of diabetes, hypertension, asthma, heart disease, and stroke on Quality-Adjusted Life Expectancy in *Value Health*, 16 (1), 140-147.

Kang, E-J. and Ko, S.K.K., 2009: A catalogue of EQ-5D utility weights for chronic diseases among noninstitutionalized community residents in Korea in *Value Health*, 12 93), S114-S117.

Katon, W.J. and Unützer, J., 2013: Health reform and the Affordable Care Act: The importance of mental health treatment to achieving the tripe aim in *Journal of Psychosomatic Research*, 74(6), 533-537.

Kim, J-S, Frees. and Frees, E.W., 2006: Multilevel Modelling with correlated Effects, research supported by National academy of Education/ Spencer Foundation and National Science Foundation, University of Wisconsin.

Kistner, U., 2003: Gender-based violence and HIV/ AIDS in South Africa: a literature review, CADRE, Johannesburg.

Klein, D.N., 2004: Different reasons for comorbidity require different solutions in *World Psychiatry*, 3(1), 28.

Kleintjies, S, Flisher, A.J, Railoun, A, Lund, C, Molteno, C. and Robertson, B.A., 2006: Occurrence of comorbid substance use disorders among acute psychiatric inpatients at Stikland Hospital in the Western Cape, South Africa in *South African Psychiatry Review*, 9, 157-160.

Kwan, M.P., 2012: How GIS can help address the uncertain geographic context problem in social science research in *Annals of GIS*, 18 (4), 245 – 255.

Kwandal, R, Gard, P.K, Garg, R.D., 2009: Health GIS and HIV/AIDS studies: Perspective and retrospective in *Journal of Biomedical Informatics*, 42(4), 748-755.

Lalkhen, H. and Mash, R., 2015: Multimorbidity in non-communicable diseases in South African primary healthcare in *The South African Medical Journal*, 105 (2), 134-138.

Laranjo, L., Rodrigues, D., Pereira, A.M., Ribeiro, R.T. and Bavida, J.M., 2016: Use of electronic health records and Geographic Information Systems in public health surveillance of Type 2 Diabetes : a feasibility study, *JMIR Public Health Survey*, 2(1), 1-12.

Leone, T, Coast, E, Narayanan, S. and de Graft Aikins, A., 2012: Diabetes and depression comorbidity and socioeconomic status in low- and middle-income countries (LMICs): a mapping of the evidence in *Global Health*, 8, 29-39.

Levin, K.A., 2006: Study Deisgn III: cross-sectional studies in *Evidence-based dentistry*, 7(1), 24.

Leyland, A.H. and Groenewegen, P.P., 2003: Multilevel modelling and public health policy in *Scandinavian Journal of Public Health*, 31, 267 – 274.

Li, F, Fisher, J, Brownson, R.C. and Bosworth, M., 2005: Multilevel modelling of built environment characteristics related to neighborhood walking activity in older adults in *Journal of Epidemiology in Community Health*, 59, 558 – 564.

Maclaurin, G, leyk, S. and Hunter, L.M., 2015: Understanding the combined impacts of aggregation and spatial non-stationarity: The case of migration-environment associations in rural South Africa in *Transactions in GIS*, 19(6), 877-895.

Mahlakoana, T., 2018: Poor education in SA keeps it from economy of the future in *Business Day.* Available from: www.businesslive.co.za (Accessed 24 March 2019).

Maragi, F., 2010: Environmental health hazards and social justice in *Geographical perspectives on race and class disparities,* London, Washington DC.

Matthews, S.A., 2006: GeoDa and Spatial Regression Modelling, GISPopSci. Available from: http://ibis.geog.ubc.ca/~brian/workshop/GeoDa_Spatial_Regression.pdf (Accessed 11 August 2019).

Mayo Clinic., 2018: Exercise and chronic disease: get the facts, The Mayo Clinic. Available from: www.mayoclinic.org (Accessed 27 May 2019).

McLafferty, S.L., 2003: GIS and Health care in *Annual Review of Public Health*, 24(1), 25-42.

Mitas, L. and Mitashova, H., 2000: Spatial Interpolation. Available from: www.geos.ed.ac.uk (Accessed 20 August 2018).

Mobley, L.R., Kuo, T.M., and Andrews, L., 2008. How sensitive are multilevel regression findings to defined area of context: a case study of mammography use in California in *Medical Care Research and Review*, 65 (3), 315–337.

Morrel, R, Jewkes, R. and Lindegger, G., 2012: Hegemonic Masculinities in South Africa: Culture, power and gender politics *in Men and Masculinities*, 15 (1), 11-30.

Mosenifar, Z., 2018: Viral Pneumonia in *Medscape.* Available from: www.emedicine.medscape.com (Accessed 01 December 2018).

Mudiyanselage, P. and Senanyake, C., 2016: GIS based spatial analysis and hotspot detection of kidney disease in *International Journal of Scientific and Research Publications*, 6(11), 330 – 336.

Murad, A., 2018: Using GIS for determining variations in health access in Jeddah City, Saudi Arabia in *International Journal of Geo-Information,* 7, 1-12.

Naiker, S., 2003: End-stage renal disease in sub-Saharan and South Africa in *Kidney International*, 63 (83) 119-122.

Nandi, A. and Shakoor, A., 2010: A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses in *Engineering Geology,* 110 (1-2), 11-20.

National Centre for Environmental Health., 2017: Health effects asthma and the environment-CDC Tracking Network, Centers for Disease Control and Prevention. Available from: www.ephtracking.cdc.gov (Accessed 02 September 2018).

National Centre for Research Methods., 2017: 5.5.2 Model 2: multilevel model: null model, Multilevel models for macro and micro data. Available from: www.restore.ac.uk (Accessed 11 August 2019).

Naves, L.A, Porto, L.B, Corrêa Rosa, J.W.C, Casulari, L.A. and Corrêa Rosa, J.W., 2015: Geographic information system (GIS) as a new tool to evaluate epidemiology based on spatial analysis and clinical outcomes in acromegaly in *Pituitary*, 18, 8-15.

NHS., 2019: Food poverty and health inequalities, Food Poverty. Available from www.healthscotland.scot (Accessed 27 May 2019).

NIH., 2017: Health consequences of drug misuse, Drugs of Abuse. Available from: www.drugabuse.gov (Accessed 18 July 2018).

Nkosi, V, Wichmann, J. and Voyi, K., 2016: Comorbidity of respiratory and cardiovascular disease among the elderly residing close to mine dumps in South Africa: A cross-sectional study in The South African Medical Journal (SAMJ), 106(3), 290-297.

Ntuli, S.T, Maimela, E, Alberts, M, Choma, S. and Dikotope, S., 2015: Prevalence and associated risk factors of hypertension amongst adults in a rural community of Limpopo Province, South Africa in *African Journal of primary health care & family medicine,*7, 842-847

Orloff, J. and Bloom, J., 2014: Comparison of frequentist and Bayesian inference, Introduction to Probability and Statistics. Available from: www.owc.mit.edu (Accessed 20 August 2018).

Pandie, M, van der Plas, H, Martens, G. and Mendelson, M., 2012: The role of the Infectious Disease Unit at Groote Schuur Hospital in addressing South Africa's greatest burden of disease in *South African Medical Journal*, 102(6), 528 - 531.

Park, Y.M. and Kim, Y., 2014: A spatially filtered multilevel model to account for spatial dependency: application of self-rated health status in South Korea, *International Journal of Health Geographics*, 13(6), 1-10.

Peltzer, K, Ramlagan, S. and Satekge, M., 2012: Alcohol abuse, problem drinking and health risk factors among South Africa youths in *Journal of Psychology in Africa*. Available from: www.hsrc.ac.za (Accessed 03 September 2018).

Peltzer, K., 2013: Hypertension and associated factors in older adults in South Africa in *Cardiovascular Journal of Africa*, 24(3), 66-72.

Pickering, T.G., 1997: The effects of environmental and lifestyle factors on blood pressure and the intermediary role of the sympathetic nervous system, in *Journal of Human Hypertension*, 1, 9-18.

Pillinger, R., 2019: A transcript of random intercept models presentation, Random intercept models. Available from www.bristol.ac.uk (Accessed 11 August 2019).

Ressing, M, Blettner, M. and Klug, S.J., 2010: Data analysis of epidemiological studies in *Deutsches Arzteblatt International,* 107(11), 187-192.

Rice, N. and Jones, A., 1997: Multilevel models and health economics, published by Technical Paper Series, University of York.

Robert Wood Johnson Foundation., 2011: Mental disorders and Medical comorbidity, The Synthesis Project. Available from: http://www.rwjf.org (Accessed 12 June 2017).

Roomaney, R.A, Pillay-van-Wyk, V, Awotiwon, O.F, Dhansay, A, Grownwald, P, Jouber, J.D, Nglazi, M.D, Nicol, E. and Bradshaw, D., 2016: Epidemiology f lower respiratory infection and pneumonia in South Africa (1997-2015): a systematic review protocol in *BMJ Open,*6(9), 1-6.

Roundy, L., 2018: Cross-sectional research: definitions and examples, Educational Psychology: help and Review. Available from: www.study.com (Accessed 12 August 2019).

Rousseau, P., 2016: HIV hotspot mapping: evidence informed HIV response, EE Publishers. Available from: www.ee.co.za (Accessed 22 March 2019).

Ruankaew, N., 2005: GIS and epidemiology in *Journal Medical Association of Thailand*, 88 (11), 1738-8.

Rytkönen, M.J.P., 2004: Not all maps are equal: GIS and spatial analysis in epidemiology, *International Journal of Circumpolar Health*, 63(1), 9-24

SADHS., 2016: South African Demographic and Health Survey in the *Key Indicator Report of the National Department of Health,* published by Statistics South Africa, Pretoria.

Sahai, H. and Kurshid. A., 1993: Confidence interval for the meanof a poisson distribution: a review in *Biomedical Journal, 35(7), 857-867.*

Schwandt, H.M, Coresh, J. and Hindin, M.J., 2016: Marital Status, Hypertension, Coronary Heart Disease, Diabetes, and Death Among African American Women and Men: Incidence and Prevalence in the Atherosclerosis Risk in Communities (ARIC) Study Participants in *Journal of Family Issues*, 31(9), 1211 – 1229.

Sedgwick, P., 2014: Cross sectional studies: advantages and disadvantages in *Biomedical Journal*, 348.

Seedat, M, Van Niekerk, A, Jekwes, R, Suffla, S. and Ratele, K., 2009: Violence and injuries in South Africa: prioritizing an agenda for prevention in *Health in South Africa*, 374, 1011-1022.

Shahid, R, Bertazzon, S, Knudtson, M.L. and Ghali, W.A., 2009: Comparison of distance measures in spatial analytical modelling for health service planning in *BMC Health Service Research,* 9(200), 1-14.

Sherwin, R.P., 1983: What is an adverse health effects?, Environmental health Perspective 52, 177-182.

Socioeconomic Review and Outlook., 2018: Socioeconomic review, chapter 4, Gauteng Provincial Government.Available from: www.gauteng.gov.za (Accessed 26 March 2019).

Soo, M, Robertson, L.M, Ali, T, Clark, L.E, Fluck, N, Johnson, M, Marks, A, Prescott, G.J, Smith, W.C.S. and Black, C., 2014: Approaches to ascertaining comorbidity information: validation of routine hospital episode data with clinician-based case note review in *BMC Ressearch Notes*, 7(253), 1-8.

Soul City Institute., 2016: A Review of Literature on drugs and substance abuse amongst youth and young women in South Africa.Available from: www.soulcity.org.za (Accessed 12 July 2018).

South Africa, 2015: *Mortaility and cause of death: findings from death notification*, Statistics South Africa, Report No. P0309.3, ISIbalo House, Pretoria.

South African Health Review., 2017: South African Health Review: 20th edition, Health Systems Trust, Durban.

South African Government., 2019: Health, South African Government. Available from: www.govt.za/ health-sa/health/ (Accessed 27th March 2019).

Sowers, J.R, Epstein, M. and Frohlich, E.d., 2001: Diabetes, Hypertension, and Cardiovascular Disease in *Hypertension*, 37 (4), 1053-1059.

Stakhovych, S. and Bijmolt, T.H.A., 2009: Specification of spatial models: A simulation study on weights matrices in Regional Science, 88(2), 389-408.

StatsSA., 2011: Profile for older persons in South Africa in *Census 2011.* Available from: www.statssa.gov.za (Accessed 15 July 2018).

StatsSA., 2013: Census of Death 2013, Statistics South Africa. Available from: http://www.statssa.gov.za (Accessed 12 June 2017).

StatsSA.,2015: Stats in Brief, Statistics South Africa. Available from: http://www.statssa.gov.za (Accessed 12 June 2017).

StatsSA.,2017: Provincial Profile: Gauteng. Available from: www.statstsa.gov.za (Accessed 19 November 2018).

StatsSA., 2018: Crime against women in South Africa, Crime Statistics Series Volume V. Available from: www.statssa.gov.za (Accessed 26 November 2018).

StatsSA., 2018: Mid-year population estimates 2018.Available from: www.statssa.gov.za (Accessed 17 March 2019).

Strand, L.B, Tsai, M.K, Wen, C.P, Chang, S-S. and Brumpton, B.M., 2018: Is having asthma associated with an increased risk of dying from cardiovascular diseases? A Prospective Cohort Study of 446 346 Taiwanese adults in *BMJ Open*, 8 (10), 1-15.

Streiner, D.L., 2017: Parametric Statistics, Statistics in Public Health. Available from: www.oxfordbibliographies.com (Accessed 15 August 2018).

Stopka, T.J, Krawczyk, C, Gradziel, P. Geraghty, E.M., 2014: Use of spatial epidemiology and hot spot analysis to target women eligible for prenatal women, infants and children services in *American Journal of Public Health*, 104(1), 183 – 189.

Sudore, R.L, Yaffe, K, Satterfield, S, Harris, T.B, mehta, K.M, Simonsick, E.M, Newman, A.B, Rosano, C, Rooks, R, Rubin, S.M., Ayonayon, H.N. and Schillinger, D., 2006: Limited literacy and Mortality in the Elderly in *The Health, Aging, and Body Composition Study*, 21, 806-812.

Sundararajan, V, Henderson, T, Perry, C, Muggivan, A, Quan, H. and Ghail, W.A., 2004: New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality *in Journal of Clinical Epidemiology,* 57 (12), 1288-1294.

Sustain., 2019: What are the consequences of food poverty, Food Access Network. Available from:

www.sustainweb.org (Accessed 27 May 2019).

Tanser, F.C. and le Sueur, D., 2002: The application of geographical information systems to important public health problems in Africa, *International Journal of Health Geographics*, 1(4), 1-9.

The Mercury, 2018: SA's blood pressure rate is rising, IOL. Available from: www.iol.co.za (Accessed 15 March 2019).

Thelle, D.S. and Laake, P., 2015: Epidemiology in *Research in Medical and Biological Sciences (Second Edition),* chapter 9, 275-320.

Tuluri, F, Gorai, A.K. and James, A., 2007: Hot spot analysis for examining the associations between spatial air pollutants and asthma in New York State, USA: using Kernel Density Estimation (KDE). Available from: www. http://proceedings.esri.com/library/userconf/proc17/papers/2255_12.pdf (Accessed 22 March 2019).

Tunks, E.R, Crook, J. and Weir, R., 2008: Epidemiology of Chronic Pain with Psychological Comorbidity: Prevalence, Risk, Course, and Prognosis in La Revue canadienne de psychiatrie, 53(4), 224-234.

Ulugtekin, N., Alkoy, S., Seker, D.Z and Goksel, C., 2006: Use of GIS in Epidemiology: A case study in Istanbul, *Journal of Environmental Science and Health Part A,* 41, 2013-2026.

Valderas, J.M., Starfield, B., Sibbald, B., Salisbury, C. and Roland, M., 2007: Defining Comorbidity: Implications for understanding health and health services, *Annals of Family Medicine*, 7 (4), 357-363.

Van den Akker, M, Buntinx, F, Roos, S. and Knottnerus, J., 2001: Problems in determining occurrence rates of multimorbidity *in Journal of Clinical Epidemiology*, 54(7), 675-679.

Van Reenen, M. and Janssen, B., 2015: Scoring the EQ-5D-5L in *EQ-5D-5L User Guide*, version 2.1.

Vearey, J., 2011: Challenging urban health: towards an improved local government response to migration, informal settlements, and HIV in Johannesburg, South Africa in *Global Health Action*, 4, 1-9.

Vearey, J, Misago, J-P, Araia, T. and Monama, E., 2014: Migration, mobility and urban vulnerabilities: implications for urban governance in South Africa, ACMS Research Report, South African Cities Network (SACN).

Vearey, J, Thomson, K, Sommers, T. and Sprague, C., 2017: Analyzing local-level responses to migration and urban health in Hillbrow: the Johannesburg Migrant Health Forum in *BMC Public Health*, 17(3), 89-93.

Wang, W, Ying, Y, Wu, Q, Zhang, H, Ma, D. and Xiao, W., 2015: A GIS-based spatial correlation analysis for ambient air pollution and AECOPD hospitalizations in Jinan, China in *Respiratory Medicine*, 109(3), 372-378.

Warburton, D.E.R, Nicol, C.N. and Bredin, S.S.D., 2006: health benefits of physical activity: the evidence in Canadian Medical Association (CMAJ), 176(4), 801-809.

Weich, L. and Pienaar, W., 2009: Occurrence of comorbid substance use disorders among acute psychiatric inpatients at Stikland Hospital in the Western Cape, South Africa in *African Journal of Psychiatry*, 12, 213-217.

Weimann, A, Oni,T. and Dai, D., 2016: A cross-sectional and spatial analysis of the prevalence of multimorbidity and its association with socioeconomic disadvantage in South Africa: A comparison between 2008 and 2012 in *Social Science & Medicine,* 163(10), 144-156.

WHO., 2008: Our cities, our health, our future in *Acting on social determinants for health equity in urban settings*, Geneva, Commission on the Social Determinants of Health.

WHO., 2018: Ageing and Health in *Global Strategy and action plan on ageing and health.* Available from: www.who.int (Accessed 02 September 2018).

WHO.,2018: Country Health profile. Available from: www.who.int/gho/countries (Accessed 15 November 2018).

Wicklin, R., 2013: The role of statistics in the top public health achievements of the 20th century, The DO Loop. Available from: www.blogs.sas.com (Accessed 15 August 2018).

Wilkinson, P., Grundy, C., Landon, M. and Stevenson, S., 2003: GIS and Health, Gatrell, A.C. and Löytönen, M (eds), *GIS and Health*, Taylor & Francis, London, 179-190.

Worltey, R. and Mazerolle, L., 2008: Analysing crime patterns in *Environmental criminology and crime analysis*, Willan Publishing, New York.

Xi, B, Zhao, M, Veeranki, S.P. and Li, S., 2019: Beneficial associations of low and large dosage of leisure time physical activity with all-cause, cardiovascular disease and cancer mortality: a national cohort study of 88,140 US adults in *British Journal of Sports Medicine*, doi: 10.1136/bjsports-2018-099254.

Zulu, L.C, Kalipeni, E. and Johannes, E., 2014: Analyzing spatial clustering and the spatiotemporal nature and trends of HIV/AIDS prevalence using GIS: the case of Malawi, 1994-2010 in BMC Infectious Diseases, 14, 264-285.

Appendix

The table below highlights a list of questions used to address the analysis in the report. The table illustrates Likert-type questions found in the QoL Survey 2015.

QoL Survey 2015 question	Likert-type
To which population does the respondent belong?	Multiple choice
What is the sex of the respondent	Multiple choice
Would you say the water you receive is always clean, usually, sometimes, hardly ever, never?	Multiple choice
How is the refuse or rubbish of this household	Multiple choice
What energy source (e.g. candles, electricity) is most used for lighting in your household?	Multiple choice
Were you born in Gauteng or did you move into Gauteng from another province?	Multiple choice
Have you seen an improvement or deterioration in this country or suburb in the	Multiple choice

last 12 months?	
What is the biggest problem facing your community?	Multiple choice
What is the most important reason you live in your suburb?	Multiple choice
In the past 12 months, how often did any adult (18yrs and above) in this household go hungry because there wasn’t enough food?	Multiple choice
In the past 12 months, how often did any child (17yrs or younger) in this household go hungry because there wasn’t enough food?	Multiple choice
What is your employment status?	Multiple choice
Where do you usually go for health care?	Multiple choice
In the last year, have you or any other member of this household had any of the following conditions?	Multiple choice
Would you describe your health status in the past 4 weeks as excellent, good, poor, or very poor?	Multiple choice
How often do you exercise (e.g. walk, run, gym)?	Multiple choice
How worried are you?	Likert-scale
What is the highest level of education you have completed?	Multiple choice
What is your current age?	Multiple choice
How many people, including you, live in this household including under 18-year olds?	Multiple choice
Does anybody in this household receive social grant of any type, such as an old age pension, childcare or disability grant?	Close-ended
Can you tell me what is the total amount of money brought into the household per month by all household members? This is after deductions such as tax, medical aid and pension contributions	Multiple choice